Analyzing Software Patterns Network obtained from Portland Pattern Repository

Hironori Washizaki, Suthinan Thanintranon,
Masashi Kadoya and Yoshiaki Fukazawa
Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555 Japan
washizaki@waseda.jp
suthinan_mart@fuji.waseda.jp
kage_purinman@ruri.waseda.jp
fukazawa@waseda.jp

Takeshi Kawamura
USOL Tokyo Co., Ltd.
takeshi.kawamura@usol.co.jp

Joseph W. Yoder
The Refactory, Inc.
joe@refactory.com

***Abstract*— Software patterns are general reusable solutions to commonly occurring problems within a given context. Patterns usually form a network of relationships that support how to understand and utilize the patterns efficiently and effectively. However little is known about the nature of such pattern networks, such as the centrality of a pattern and its meaning. To clarify such characteristics of software patterns, we mine a network consisting 285 patterns from the current world-largest online pattern repository called Portland Pattern Repository (PPR). By applying network analysis techniques to the mined network and careful review of the result, we revealed several interesting characteristics of the pattern network and patterns in PPR, such as that the degree centrality could be useful to support developers and to more easily understand whether patterns under consideration are core patterns or peripheral ones in the entire pattern network. The "betweeness centrality" seems to be useful to support ways to identify those patterns that play a role of hub and grasp relationships among different pattern groups. [1] Moreover we believe these findings could contribute to further researches on pattern networks.**

Keywords—software patterns; network analysis; pattern repository

I. INTRODUCTION

A Software pattern is a general reusable solution to a commonly occurring problem within a given context, specifically for software development. Most patterns have relations to other patterns [2][20-21], such as inside usage (X uses Y in its solution), combinational usage (X can be combined with Y), similar patterns (X is similar to Y), and alternative patterns (X is an alternative to Y) [3]. Such relations are usually described in a "Related Patterns" section or other related sections of each pattern document. As a result, patterns usually form a network having relations among them to support users understand and utilize patterns efficiently and effectively.

However little is known about the nature of entire or partial pattern network, such as how are patterns connected with many patterns relatively important from the viewpoint of frequent applications. Knowing such characteristics could be beneficial for understanding, reusing and extending existing patterns and writing new ones.

To clarify such characteristics of pattern networks and organizational patterns, we mine a network consisting of 285 patterns (March 2013) from the online pattern repository called Portland Pattern Repository (PPR) [4] . We choose PPR as the target repository because, to the best of our knowledge, PPR is the oldest, largest and still active pattern repository so that its analysis should be beneficial to discover common, major and current characteristics in pattern networks containing various patterns. PPR is an origin of Wiki and being actively updated; each Wiki page in PPR describes a pattern or a document related to patterns. For example a page "Pattern Index" is an index for many patterns and was last edited on April 2013.

It should be noted that anyone can contribute to editing patterns and related documents in PPR. Because of this there are some peculiarities such as existence of non-pattern pages, the lack of a common format and no guarantee that all relationships have been captured. The first peculiarity has been solved by manual filtering described in the next section; however other two are not currently addressed so that these might affect the analysis result such as missing important relationships. In the future we will attempt to analyze other smaller, somewhat old but more refined repositories and catalogs such as the Pattern Almanac [16] and well-accepted books on patterns (such as [13]), and compare the results with the findings reported in this paper.

By applying network analysis techniques to the mined network and careful review of the result, we revealed several interesting characteristics of the pattern network and several patterns such as organizational patterns. For example, the degree centrality seems to reflect the commonness and generality of the corresponding pattern.

The remainder of this paper is organized as follows. First, we describe our analysis procedure in Section II. We then describe and discuss obtained results and findings in Section III. Finally, we conclude our work in Section IV.

[1] Preliminary result focusing on organizational patterns is presented at [1]. In this paper we show and discuss in detail about various pattern groups and their comparison.

II. Analysis Procedure

We analyzed the pattern network in PPR using the following steps.

(Step1) We collected 483 pattern names, incoming and outgoing relations, and belonging groups by crawling the PPR Wiki pages by using a Web crawler based on WebSPHINX [5]. We systematically dealt with any link described in Wiki pages as relations. Most of those links are about inside usages, combinational usages and similar patterns. PPR contains several non-pattern documents so that we regarded pages that are linked from the "Pattern Index" page and the "Category Pattern" page as patterns. We eliminated pages containing two or more patterns, in order to ensure that each page contains just one pattern. Moreover, we identified groups where patterns belong by checking links manually from pattern pages to category pages. Some pattern pages have links to multiple groups; in that case we choose one group manually by considering the major property of target patterns.

(Step2) We manually filtered out several non-pattern documents such as the pages whose names contain "guideline" and "discussion". Finally we obtained 285 patterns and 20 groups; each pattern belongs to just one group.

(Step3) We measured three major types of centrality (degree, closeness and "betweeness" defined below) [6] for each pattern by using a network analysis tool Pajek [7]. When measuring centrality, to make analysis simple, we did not distinguish the direction of relations; measuring centrality for each direction (i.e. incoming or outgoing) could be future work.

- The normalized degree centrality[2] (hereafter "degree centrality") C_d of a pattern *i* is defined by the following formula, where v_i is a node corresponding to the pattern *i*, $\deg(v_i)$ is the number of nodes connected to the node v_i, and *n* is the total number of the given network. The degree centrality can be seen as the simplest centrality. We assume that the degree centrality of each pattern quantifies the pattern's importance and the commonness of the context of the pattern in the belonging entire pattern network. If two patterns belong to a meaningful relationship such as the internal usage, combinatorial usage or similar patterns, their contexts should have some commonality; it results in a link between corresponding pattern pages if page authors are aware of the relationship.

$$C_d(v_i) = \frac{\deg(v_i)}{n-1}$$

- The closeness centrality C_c is defined by the following formula, where $s(v_i)$ is the sum of v_i's distances to all other nodes. We assume that the closeness centrality of each pattern reflects the pattern's importance within its corresponding pattern group since the closeness centrality in general tends to give high scores to nodes that are near the center of local clusters in an overall network [17].

$$C_c(v_i) = \frac{n-1}{s(v_i)}$$

- The betweeness centrality C_b is defined by the following formula, where $BC(v_i)$ is the betweenness of v_i, which reflects the number of shortest paths from all nodes to all others that pass through v_i. We assume that the betweeness centrality of each pattern quantifies the extent to which the pattern plays a role of hub connecting different pattern groups.

$$C_b(v_i) = \frac{2BC(v_i)}{(n-1)(n-2)}$$

[2] Instead of the degree centrality, we use the normalized degree centrality so that future work can compare different pattern repositories and catalogs.

III. Analysis Results

According to the above-mentioned procedure, we obtained the following results with corresponding findings: (1) number of patterns in each group, (2) centrality of pattern, and (3) Group comparison regarding degree centrality. Moreover by focusing one pattern group, organizational patterns, we discuss (4) implications of centrality of patterns.

(1) Number of patterns in each group

Table I shows the id, name and number of patterns for each group. In Table I, among 285 patterns, many patterns are about software product, such as Java Idioms (ID 11) and Software Design Patterns (17).

TABLE I. Groups of pattern in PPR

Group ID	Group name	N. patterns
1	(none)	2
2	Category Concurrency Patterns	9
3	Category Creational Patterns	1
4	Category Security Patterns	4
5	CategoryStructuralPatterns	6
6	Component Design Patterns	25
7	Derivations & extensions to MVC	8
8	Functional Pattern System For Object Oriented Design	4
9	Graphics Patterns	23
10	Individual patterns	6
11	Java Idioms	69
12	Not a pattern	2
13	Object Based Programming	10
14	Organizational Patterns	15
15	Patterns For Effective Meetings	1
16	Process Anti Patterns	1
17	Software Design Patterns	42
18	Testing Patterns	17
19	User Interface Patterns	3
20	Website Patterns	10

(2) Centrality of pattern

Fig 1, 2 and 3 show the histogram of degree, closeness and betweeness centrality of all 285 patterns, respectively. Moreover Fig 4 illustrates entire pattern network; in Fig 4, 10 highlighted nodes (listed in Table II) indicate patterns having high degree centrality (> 0.03). These highlighted patterns are mostly design patterns such as “Model View Controller” and “Adapter Pattern”; however there is one organizational pattern “Scape Goat” which has high degree centrality. We can see that most of those highlighted high-degree centrality patterns are located relatively center of the network.

In Fig 1, many patterns have a small degree centrality; which means that many patterns refer to a small number of patterns through a Wiki page link. However the distribution does not follow the well-known network property “power law distribution” so we cannot state that the pattern network obtained from PPR is a scale-free network. Though previous Wikipedia research reported that various quantities including the in-degree of links from other Wiki pages are distributed according to the power law [18]. The reason why the pattern network in PPR does not follow the power law distribution might be because of the smallness of the size or some nature of its evolution process; in the future we will investigate larger pattern networks and their evolution process in PPR.

In Fig 2, the distribution of patterns regarding closeness centrality seems to follow the normal distribution. In Fig 3, most of patterns have very low betweeness centrality; which means that in the pattern network, few patterns seem to play a role of hub that connects different pattern groups (like “Scape Goat” shown in Fig 10). However there might be a possibility that the evolution process of the pattern network affects the betweeness centrality; for example patterns that were written earlier in time may not refer to patterns that were written later, although it is always possible to update every Wiki page. Indeed, we confirmed that the last update dates of the target 285 pattern pages vary from 3 weeks ago to 16 years ago. In the future we will investigate their evolution process and its impact on the betweeness centrality.

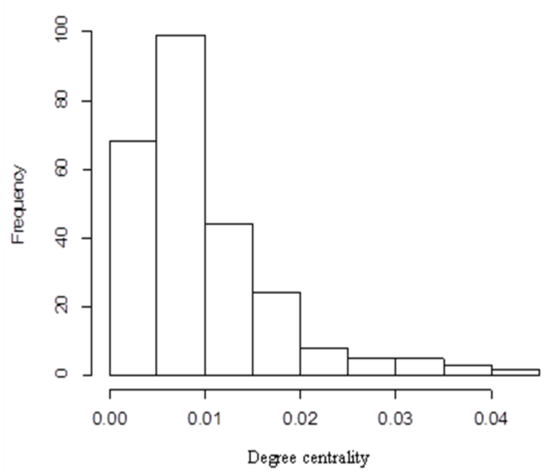

Fig. 1. Histogram of degree centrality of all patterns

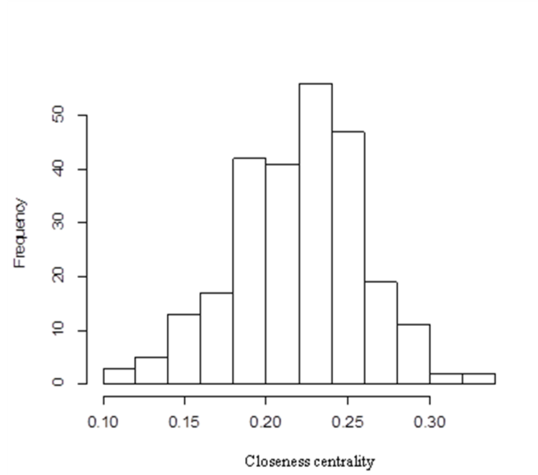

Fig. 2. Histogram of closeness centrality of all patterns

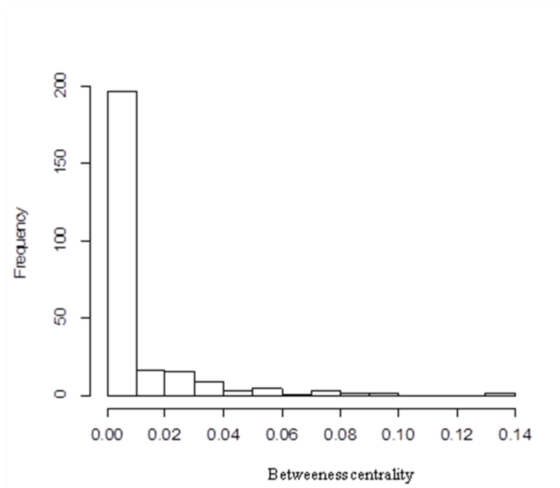

Fig. 3. Histogram of betweeness centrality of all patterns

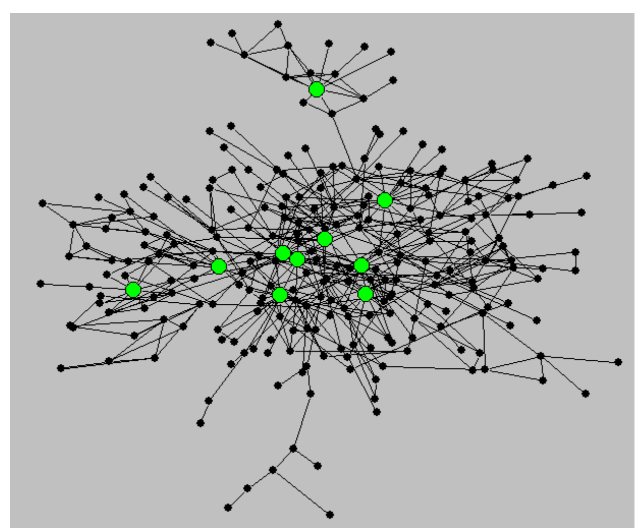

Fig. 4. Entire pattern network in PPR (each node indicates a pattern; highlighted large nodes indicate patterns having high degree centrality (> 0.03); each link indicates a relation between two patterns.)

TABLE II. MEASUREMENT RESULTS OF TOP-10 PATTERNS

Pattern name	N. patterns referred by the pattern	N. patterns referring to the pattern	Degree centrality	Closeness centrality	Betweeness centrality
ModelViewController	11	12	0.044747	0.295742	0.081803
AdapterPattern	6	15	0.040856	0.289089	0.092927
HandleBodyPattern	9	10	0.036965	0.256743	0.030538
ynchronizationStrategie	9	9	0.035019	0.203968	0.034224
VisitorPattern	7	11	0.035019	0.328225	0.131212
SceneGraph	6	11	0.033074	0.266598	0.079642
ValueObject	3	14	0.033074	0.279348	0.041879
ScapeGoat	6	10	0.031128	0.176875	0.013398
CompositePattern	4	12	0.031128	0.317676	0.131146
StrategyPattern	5	11	0.031128	0.325316	0.089017

(3) Group comparison regarding degree centrality

Fig 5 shows the box plot chart of degree centrality for each group. In Fig 5, each centrality is calculated against the whole pattern network (i.e. n=285 in its formula). Groups tend to have different distribution of degree centrality. Moreover the correlation coefficient between the group size (i.e. the number of patterns) and the arithmetic mean of degree centrality of each group is 0.15, so we do not confirm correlation between them.

Among 20 groups, there are four major groups having obviously wider range of degree centrality compared with other groups: Component Design Patterns (ID 6), Graphics Patterns (9), Object Based Programming (13) and Software Design Patterns (17). This could be because there are many patterns belonging to the same group 6, 9 and 17 so that they tend to have tight relationships among them.

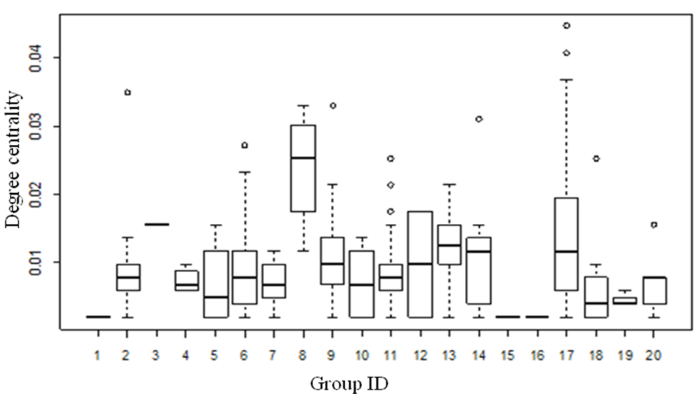

Fig. 5. Box plot of degree centrality for each group

(4) Implications of centrality of patterns

We try to figure out implications of centrality of patterns by focusing on one pattern group, organizational patterns. We chose that group because organizational patterns have a long history and are known as the basis for Agile software development movement [8-15], which is now widely accepted.

In Table I, there are only 15 organizational patterns (ID 14) including positive patterns and anti-patterns.

In Fig 5, organizational patterns (ID 14) tend to have a somewhat wide range of degree centrality compared with other groups such as groups 1, 2, 3, 4, 7, 15, 16, 18, 19, and 20. It is mainly because organizational patterns tend to have many relationships with each other closed inside the same group. Fig 6 shows how most of relationships are closed within the group. Among them, Fig 7 shows details of selected four patterns having high centrality.

Table III shows the number of related patterns (i.e. patterns referred by the pattern, and the number of patterns referring to the pattern), degree, closeness and betweeness centrality of each organizational pattern.

In Table III, many organizational patterns have several related patterns. Especially, there are three patterns having high degree centrality (i.e. having many related patterns) : “Scape Goat”, “Peace Maker” and “Train Hard Fight Easy”. “Scape Goat”, as an anti-pattern, gives other normal organizational patterns a common problematic context so that it is connected to other organizational patterns including the above-mentioned two patterns. In contrast, organizational patterns having low degree centrality such as “Slow Poison” and “Change Of Setting” are specific to their own contexts. It means that the degree centrality seems to somewhat reflect the commonness and generality of the corresponding pattern. In other word, the degree centrality could be useful to support developers to easily understand whether patterns under consideration are core patterns or peripheral patterns in the entire pattern network.

Regarding the betweeness centrality, “Scape Goat” seems to play the roles of hub. Actually they connect organizational patterns to other groups: Process Anti Patterns (ID: 16) and Patterns for Effective Meetings (ID: 15).

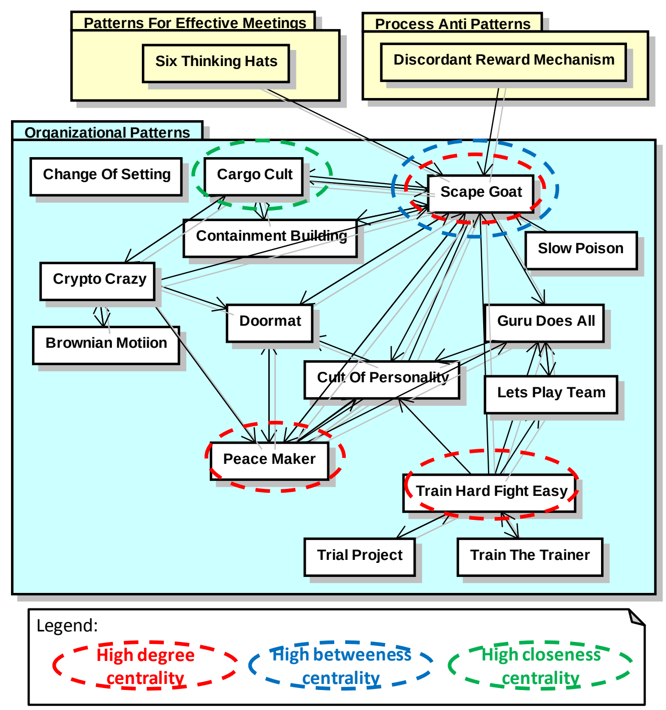

Fig. 6. Relationships among organizational patterns and related patterns belonging to other groups in UML class diagram (each class box indicates a pattern; each package box indicates a pattern group; due to space limitation not all of patterns are shown for non-organizatoinal pattern groups.)

Regarding the closeness centrality, "Cargo Cult" has the highest centrality. It might be seen as a relatively important pattern within the group according to our assumption described in section I[3].

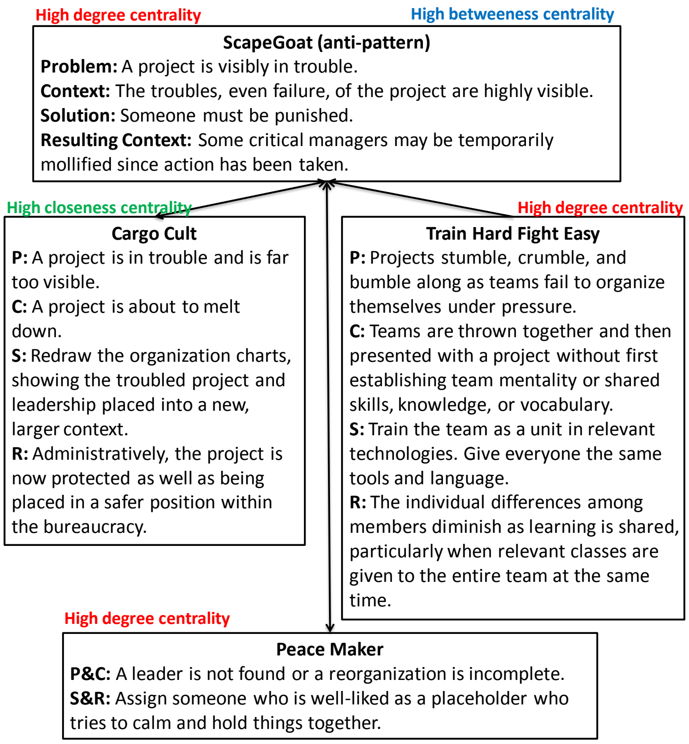

Fig. 7. Partial relationships and details of selected patterns having high centrality ('P', 'C', 'S' and 'R' denote Problem, Context, Solution and Resulting Context, respectively.)

TABLE III. MEASUREMENT RESULTS OF ORGANIZATIONAL PATTERNS

Pattern name	N. patterns referred by the pattern	N. patterns referring to the pattern	Degree centrality	Closeness centrality	Betweeness centrality
Scape Goat	6	10	0.031128	0.176875	0.013398
Peace Maker	4	4	0.015564	0.15091	0.000084
Train Hard Fight Easy	6	2	0.015564	0.151088	0.000334
Cargo Cult	3	4	0.013619	0.210311	0.027906
Crypto Cracy	5	2	0.013619	0.175787	0.00646
Cult Of Personality	3	4	0.013619	0.150999	0.000258
Door Mat	2	4	0.011673	0.150733	0
Guru Does All	2	4	0.011673	0.150999	0.003344
Containment Building	2	2	0.007782	0.175307	0
Lets Play Team	1	2	0.005837	0.131458	0
Brownian Motion	1	1	0.003891	0.149593	0
Train The Trainer	1	1	0.003891	0.131323	0
Trial Project	1	1	0.003891	0.131323	0
Change Of Setting	1	0	0.001946	0.205272	0
Slow Poison	1	0	0.001946	0.15038	0

[3] There is a possibility that a group Not A Pattern (ID: 12) affects the closeness centrality because it is connected to organizational patterns via "Cargo Cult" and "Change Of Setting". We will investigate its effect in the future.

IV. CONCLUSION AND FUTURE WORK

In summary, we mined the PPR patterns network consisting of 285 patterns including 15 organizational patterns, analyzed the pattern network, and found the following interesting characteristics of the pattern network.

- Although the pattern network cannot be seen as a scale-free network, it is still true that many patterns refer to a small number of patterns. In the pattern network, few patterns play a role of hub that connects different pattern groups.
- Four major groups have obviously wider range of degree centrality compared with other groups: Component Design Patterns, Graphics Patterns, Object Based Programming and Software Design Patterns. This could be because there are many patterns belonging to the same group. Organizational patterns have a somewhat wide range of degree centrality compared with other groups except for a few major groups. This is because organizational patterns tend to have many relationships with each other.
- The degree centrality seems to somewhat reflect the commonness and generality of the corresponding pattern. The degree centrality could be useful to support developers to more easily understand whether patterns under consideration are core patterns or peripheral ones in the entire pattern network.
- The betweeness centrality seems to be useful to support users identify patterns playing a role of hub and grasp relationships among different pattern groups.
- The closeness centrality might reflect the pattern's importance within its corresponding group; this assumption needs further investigation.

We believe that these findings could contribute to the software development community for understanding, reusing and extending existing patterns and for writing new patterns. For example, developers or managers who want to form agile teams could consider reusing those organization patterns in PPR starting by referring to ones that have a high degree centrality such as "Scape Goat". Or, if developers and managers want to grasp relationships among organizational patterns and other groups' patterns, "Scape Goat" should be considered first. In the future we plan to develop a pattern-browser that shows preferred sequences for learning about a system of patterns based on these findings. Moreover we believe these findings could contribute to further research on pattern networks.

In this analysis, we regarded Wiki page links as pattern relationships; however this is often based on page authors' awareness of other patterns so they might be different from pattern authors' intentions for the actual relationships. We will handle this threat to validity by referring to relationships specified in original pattern documents. Moreover, we did not deal with the meaning (i.e. type) of relationships such as the internal usage and combinational usage; we will try to discriminate these in future work as well. Other relationships of patterns could also be considered, such as commonality and relationships among authors, and actual combinational usages in actual software and organizations; analysis on those multiple and overlay networks in patterns could discover characteristics of patterns from the viewpoint of relationships in detail. It might lead to detecting various community structures in the network [19] .

In the future, we will investigate how are these findings related to actual agile or non-agile software development adapting organizational patterns and product ones. Such investigation will include further analysis for the impact of directions and types (such as inside usage and combinational usage [3]) of relations on characteristics of pattern networks and patterns.

ACKNOWLEDGEMENT

Our thanks go to our shepherd Michael Weiss for his valuable comments and guides to improve our paper significantly during its shepherding process.

REFERENCES

[1] Hironori Washizaki, Masashi Kadoya, Yoshiaki Fukazawa and Takeshi Kawamura, “Network Analysis for Software Patterns including Organizational Patterns in Portland Pattern Repository,” Agile 2014 Conference, Research Lightning Talk, Orlando, July 28-August 1, 2014.

[2] Atsuto Kubo, Hironori Washizaki, Atsuhiro Takasu and Yoshiaki Fukazawa, “Extracting Relations among Embedded Software Design Patterns,” Integrated Design & Process Science, vol.9, no.3, 2005.

[3] Walter Zimmer, “Relationships between design patterns,” In Pattern Languages of Program Design, Vol.1, Addison-Weslay, 1995.

[4] Cunningham & Cunningham, Inc., “Portland Pattern Repository,” http://c2.com/ppr/

[5] Rob Miller, “WebSPHINX: A Personal, Customizable Web Crawler,” http://www.cs.cmu.edu/~rcm/websphinx/

[6] Stephen Borgatti and Marting Everett, “A Graph-Theoretic Perspective on Centrality,” Social Networks, Vol.28, No.4, Elsevier, 2005.

[7] Vladimir Batagelj and Andrej Mrvar, “Pajek - Program for Large Network Analysis,” http://pajek.imfm.si/

[8] “Organizational patterns” from Wikipedia, http://en.wikipedia.org/wiki/Organizational_patterns

[9] Ward Cunningham, “Episodes: a pattern language of competitive development,” In Pattern Languages of Program Design, Vol.2, Addison-Wesley, 1996.

[10] Scrum Pattern Community and ScrumPLoP, “Scrum Patterns,” https://sites.google.com/a/scrumplop.org/published-patterns/home

[11] James O. Coplien and Jeff Sutherland, “Software Scrum Patterns,” 2008, https://sites.google.com/a/scrumorgpatterns.com/www/scrumpatternssummary

[12] Mike Beedle, Martine Devos, Yonat Sharon, Ken Schwaber and Jeff Sutherland, “SCRUM: An extension pattern language for hyperproductive software development,” Addison-Wesley Sofware Patterns Series, 1999.

[13] James O. Coplien and Neil B. Harrison, “Organizational Patterns of Agile Software Development,” Prentice Hall, 2004.

[14] James O. Coplien, “Organizational Patterns,” in Pattern Languages of Program Design, Addison-Wesley, 1995.

[15] Neil B. Harrison and James O. Coplien, “Patterns of productive software organizations,” Bell Labs Technical Journal, Vol.1, No.1, 1996.

[16] Linda Rising, “The Pattern Almanac,” Addison Wesley, 2000.

[17] Activate Networks, “Who is central to a social network? It depends on your centrality measure,” http://www.activatenetworks.net/who-is-central-to-a-social-network-it-depends-on-your-centrality-measure/

[18] Jeffrey Stuckman and James Purtilo, “Analyzing the Wikisphere: Methodology and Data to Support Quantitative Wiki Research,” Journal of the American Society for Information Science and Technology, Vol.62, No.8, 2011.

[19] M. E. J. Newman, "Detecting community structure in networks," The European Physical Journal B - Condensed Matter and Complex Systems, Vol.38, No.2, 2004.

[20] Atsuto Kubo, Hironori Washizaki, Atsuhiro Takasu, Yoshiaki Fukazawa, “Analyzing relations among software patterns based on document similarity,” International Conference on Information Technology Coding and Computing (ITCC 2005), Las Vegas, April 4-6, 2005.

[21] Hironori Washizaki, Atsuto Kubo, Atsuhiro Takasu, Yoshiaki Fukazawa, “Relation analysis among patterns on software development process,” 6th International Conference on Product Focused Software Process Improvement (PROFES 2005), Oulu, July 13-15, 2005.