Assimilation Processes of Newcomers in Online Communities

Abstract

Online communities (OCs), computer-mediated virtual space, have become one of the crucial parts of our lives. For example, users of Reddit.com visited 1.36 billion times in July 2017 that is comparable to Facebook or YouTube. While several studies explain the underlying mechanisms of members' participation and revisit behaviours, we still do not understand the evolutionary processes of members who progress from newcomers to core members. The gap mainly stems from the dominance of variance studies that deal with covariance between variables that capture different aspects of a given phenomenon in OC literature. This study takes the process-oriented approach and aims to reveal the evolutionary socialisation process of OC members by applying a novel inter-disciplinary computational qualitative method based on critical realism. Critical realism is employed for the study to help explaining complex structures of the online community and interactions between its users. Primarily, we analyse two-years event data obtained from the database of an OC. The events allow us investigate how newcomers assimilate in OCs. We find a common sequence of events created by newcomers of OCs to identify demi-regularities through clustering and visualisation. Specifically, we find that a group of assimilated users show a demi-regularity of "Like" events for other's postings and comments , which is consistent with previous studies (e.g., Morrison (1993)). Our results can show how newcomers are assimilated into OCs visually .

Keywords: Online community, digital trace data, assimilation process, sub-sequence mining, process study.

1.0 Introduction

Due to the voluntary behaviour of OC members, one of the critical success factors of OCs is retention of their members. In particular, understanding assimilation processes of OC members takes the utmost importance as assimilated members tend to remain longer and contribute more to OCs. While literature is abundant with studies revealing the factors that affect the participation of OC members, nature of members, and outcomes of OC activities, we still do not understand the details of the assimilation processes of OC members. This paper fills the research gap through an innovative process oriented analytical method.

Online communities have become one of the most substantial parts of our lives. For example, in Reddit.com, one of the most massive OC globally and ranked 8th in July 2017. The monthly unique users are 1.36 billion times in July 2017, which is comparable to 2 billion times of Facebook or 1.5 billion times on YouTube . A user reads 10.56 pages and stays 16 minutes on average per day. Reddit retains 36 million accounts as of June 2015. Based on the above values, Reddit members use more than 12.09 million hours daily, which is equivalent to 1,380 years. Globally, there are massive OCs in each country: Sina Weibo and WeChat in China, DCInside and

Ppomppu in South Korea, and Mixi and 2ch in Japan. OC activities contribute to the welfare of people by using OCs in the 21c digital era.

Understanding assimilation processes of OC members takes utmost importance for the success of OCs. Like any other organizations, one of the critical success factors of OCs is the retention of members. In particular, due to the voluntary participation thus increased fluidity of OCs (Faraj et al. 2011) there is almost no control on members leaving the communities. The fate of OCs is dependent on the voluntary behaviour of their members who decide to join and leave the communities anytime they want to. As assimilated members tend to remain longer and contribute more to the OCs, understanding the assimilation processes of OC members takes utmost importance for OC managers.

However, the existing knowledge on assimilation processes of OC members is partial and limited to variance perspectives. A group of scholars have attention on OC user's voluntary participation to explain who contributes more and why devote their times and efforts to generate contents in the OCs (Bateman et al. 2011; Hippel and Krogh 2003; Ray et al. 2014). The voluntary participation focuses on the actor-side stories of OCs. We, however, still do not understand how OC members progress from newcomers to core members of the communities. It is necessary to track a settled member's behavioural change from the moment she joined the OC to understand the evolution of membership. However, there is very little studies no research that clearly explains how a voluntary newcomer evolves into a core user in "fluid" organisational context through what behavioural changes.

Existing studies lack micro-level behavioural analysis of new members during their assimilation processes. New members need to learn the culture of target OCs and behave in a way accepted by others according to the norm of the OCs. As the behaviour of members in OCs are visible only through postings, it is natural to investigate how the postings of new members are accepted by existing OC members. We investigate how the posting behaviour or new members evolves according to comments received by other members with regards to the breach of OC norms.

In this study, we use sequence analysis to identify common sub-sequences that enable a newcomer assimilated in the OC. In a process-oriented approach, sequence data is essential to understand how members' events (footages) unfold and change as time goes.

We utilise the newcomers' assimilation process that is widely explored in the organisational research (e.g., Morrison (1993) or Jones (1986)). Specifically, in this study, we do a case study using detailed data sets from an online celebrity community to answer the research questions: How do newcomers evolve into a core member in an OC? What are the generative mechanisms of newcomers' assimilation on the OC? We answer the questions by analysing the evolutionary change of newcomer's participation pattern by using internal databases that contain exposed activities such as posts and comments as well as unexposed footages of members such as voting for postings or comments. By utilising the holistic data sets stored in the internal database of the OC, we can extend our understanding of OCs behind the scene. Our study extends the assimilation process study including OCs. We also adopt computational processing mining techniques to identify common sub-sequences that represent demi-regularities of assimilation process (Leenders et al. 2016). It also helps practitioners who run an OC as business as well as adopt online communities on their businesses to maintain high retention rate.

2.0 Conceptual Background

2.1 Newcomers Assimilation in Traditional Organizations

The newcomers' assimilation process has been of interest to many scholars for a long time as organisations can expect the longer retention of members when they are fully assimilated (Kramer and Miller 1999; Waldeck and Myers 2007). Since prompt assimilation of organizational members has strong impacts on organisational performance, researchers have been making large efforts to broaden our understanding of assimilation in organisations (Morrison 1993; Myers and Oetzel 2003). Ziller (1964) interprets assimilation processes from two perspectives. On individual perspective, intrinsically, "the socialisation (assimilation) process involves conflict between satisfaction of individuals and group needs" due to the conflict between the dependence and independence of an individual to the organisation. Moreover, an organisation grows, the problem gets harder to resolve. From a group perspective, newcomers' assimilation into a group is a course of actions to learn behaviours and attitudes necessary for a role in the organisation (Morrison 1993).

Generally, newcomers can suffer from the cultural and behavioural differences from other person's (olbies) interpretations and responses to actions or events of

organisational members that do not comply with the newcomers' pre-existing knowledge. As a result, newcomers should re-evaluate their knowledge about the organisation and try to find information about olby's behavioural styles and work specification to modify his/her future actions. Many organisations adopt socialisation tactics to minimise this kind of newcomer's difficulty (Jones 1986). Jones categorises six socialisation tactics and tests them empirically. He identifies that different socialisation tactics impact different organisational dimensions of assimilation. Particularly, socialisation tactics of an organisation and following processes impact significantly on role orientations and subsequent adjustments on organisations. He also reveals that information about role requirements and future organisational progress unfolds on reducing the uncertainty of newcomer assimilation process.

Newcomers' proactive activities during the socialisation process is reported to be an important factor. Morrison (1993) extends the proposed assimilation model from the perspective of organisational tactics to an interactive model between organisation and newcomers. She addresses the crucial factor of newcomers' activities for organisational assimilation is a newcomer's proactive information seeking. She reports that newcomers' proactive information seeking activities influences socialisation perspectives of newcomers positively. Kramer and Miller (1999) use a similar approach to Morrison (1993)'s. They emphasise newcomers' proactiveness using role-adopting and organisational assimilation in their study. At the end of the study, they call for further studies on the behaviours and processes related to organisational entry by combining multiple approaches better to explain the complicated process during organisational assimilation.

Levine and Moreland (1994) suggest group assimilation (socialisation) framework over time based on the psychological process. The authors consider the reciprocal influence between individuals and organisations as well as temporal perspective during an individual's assimilation process for an organisation. They assume three psychological processes underlying newcomer's assimilation: evaluation, commitment and role-transition process and interactions between each other. When an individual joins an organisation, she evaluates her role, responsibility, and rewards in the organisation. If these factors meet her decision criteria, the individual increases her commitment level. As time passes and her degree of commitment goes beyond the scope of her role, the organisation proposes her to change the role to a higher level, and she re-evaluates her position in the organisation. Organisational response to an

individual is not a decision of the organisation as an entity but a consensual response based on the shared views of the people who constitute the organisation.

2.2 Distinctive characteristics of online communities

There are many studies on online communities to reveal organisational characteristics different from traditional organisations (Arguello et al. 2006; Bateman et al. 2011; Oestreicher-Singer and Zalmanson 2013; Ray et al. 2014; Wang et al. 2012). Since online communities open their contents and knowledge to the public, everyone can read, participate, or use the information for free. In other words, OC is an environment that occurs the market failure of public goods due to free-riders (e.g., Fehr and Gächter (2000). Moreover, OC has very fluid structures that any members can join and leave at any time, and it is impossible to punish someone who is free-riding on the OC. Under this circumstances that there are many free-riders and contents on OCs are public goods, researchers try to answer why members continue to invest their times on OCs without "visible" benefits such as money, authority or power on OC and how OCs grow continuously. Specifically, scholars suggest some characteristics of OCs as crucial such as the classification of members of OCs (Sun et al. 2014; Velasquez et al. 2014), motivators of OC participation (Bateman et al. 2011; Ray et al. 2014), the design of OCs (Fiedler and Sarstedt 2014), and impact of members' sense of belonging by measuring the willingness to pay for business (Oestreicher-Singer and Zalmanson 2013).

Von Hippel and von Krogh introduce the concept of voluntary participation in their study of the open source software development community (Hippel and Krogh 2003), and it is widely accepted and studied in online community research. The concept of voluntary participation explains the success factor of OCs depending on member characteristics. Scholars deal with the voluntary participation as a core concept to explain members' voluntary activities in OCs (e.g., Bateman et al. (2011) or Ray et al. (2014)). Bateman et al. (2011) unravel why an OC member chooses to return repeatedly and engages in various behaviours such as writing posts and comments and moderating discussions by developing a theory of community commitment that explains various aspects of members who visit again and actively participate in OCs voluntarily. Yang et al. (2017) demonstrate that the different commitment levels (posters – high commitment members or lurkers – low commitment members) are induced by influences of perceived online community support and member relations.

The authors use the organisational support theory to justify the relationships between commitment level and organisational factors.

The concept of fluidity of an OC is suggested by Faraj et al. (2011). Authors urge that "fluidity is fundamental characteristics" and "fluidity engenders a dynamic flow of resources in and out of the community—resources such as passion, time, identity, the social disembodiment of ideas, socially ambiguous identities, and temporary convergence." Since everybody can join and leave at any time due to the fluidity of OCs, we imagine that the newcomers' assimilation processes in OCs will be profoundly different from those in traditional organisations (e.g., Jones (1986) or Morrison (1993)). For example, traditional organisations have official and compulsory training programs for newcomers to support their assimilation processes. The roles and responsibilities are clearly defined for newcomers and they receive regular feedbacks on their performances from managers. There are also unofficial channels to share information among organizational members (Morrison 1993).

Similarly, many OCs provide newcomers with formal or informal guidelines on the rules of the OCs through regular notice posts or informal comments by existing members for newcomers. However, it is unexplored yet that whether newcomers of OCs take the similar assimilation processes with those in formal organisations or completely different processes.

Existing studies on OCs are mainly concerned about issues with regards to organisational structure such as the impact of membership turnover (Qin al. 2014; Ransbotham and Kane 2011), herding behaviours on OCs (Oh and Jeon 2007), member attachment in OCs (Ren et al. 2012), and positive relationship between sense of belonging by adopting OCs in an online music platform and willingness to pay for the contents (Oestreicher-Singer and Zalmanson 2013), and we still do not understand how newcomers of OCs assimilate to the OCs.

3.0 Method

We adopt process mining as the principal methodology for the analysis of member's behavioural changes. Process mining is one of the computational qualitative methods frequently used in social science. Process studies generally aim to discover how and why things emerge, develop, grow, and terminate over time as distinct from variance questions dealing with covariance between dependent and independent variables

(Langley et al. 2013). Thus, the temporal order and sequence of events are crucial components in process studies in IS and computational science techniques like sequence analytics (Gaskin et al. 2014), which is widely used in Genetics discipline and used to identify patterns of sequential events that may provide clues in finding similarities of processes in different organizational contexts (and named as Organizational Genetics). Sequence analytics, one of the very few applications of computational science to process studies, proposes a robust alternative method by leveraging large volume of digital footprint (in our context, writing a post, commenting, messaging, recommending and complaining articles and so on). It allows the researchers to analyse micro-level individual activities based on the sequential pattern mining (Klarner and Raisch 2013). Sequence analytics can allow scholars to explore the unfolding sequence of events that are crucial in understanding the procedural dynamics (Abbott 1990; Gaskin et al. 2014). Sequence analytics in this study is used to reveal common patterns of event sequences created during assimilation processes of new members in OCs.

We collected data from a Korean celebrity fan online community (http://jungdonghagallery.com) which lasted between September 2012 and July 2014. The OC has ideal features to be used in this study. Firstly, the OC impose the cultural conflict between olbies (members who already settled in the community) and newbies (who joined the community but not settled yet). The OC was created by a sudden migration from a big online community hosting web site, called DC (http://dcinside.com), which is famous for its unique culture in terms of using jargons and community rules to ensure the higher level of freedom of expression that is distinguished from other OCs in S Korea. The OC members decided to keep the DC culture when they moved to the new web site. Therefore, new members who joined the new website had to assimilate themselves to the unique culture to settle down in the community. This the members who moved from the DC website were familiar with the DC culture while new members who joined after the OC moved had to assimilate themselves to the unique culture. The existence of dual types of members makes us comparing pattern differences between them under membership tenure.

Secondly, during the OC was active, the celebrity had many chances to broadcast and build up a reputation. So, there are many heterogeneous people joined the OC, and it makes our analysis not biased. Also, the OC went through an OC lifecycle stages including creation, growth, conflict, and decline during two years therefore ideal to

study the process aspects of OC lifecycle and how fluidity makes an impact to the lifecycle process.

Our dataset contains all events created by members from the beginning (Sep 2012) to the closure of the OC in July 2014. It consists of member's writings (postings, comments, and direct messages between members), member's activity history logs (votes, complaints, membership registration, writing histories of postings and comments, and last login date). By analysing membership registration, there are 539 registered members in the dataset. However, some members did not do anything after registration and others did a very small number of activities, we remove some members who have no or less than five events. After filtering out inactive members, we have a list of 399 members. Since it is common to keep a member's nickname constantly, we divide into classes using nicknames between olbies and newcomers. We retrieve olbies' nickname from the original online forum. As a result, olbies are constituted of 103 members who migrated from the DC web site. The rest of them are 296 newcomers who joined the community after the migration.

Figure 1 shows the daily statistics of total number of events created by members (NumEvent), the number of newly registered members (NumJoin), and the number of members who made the last login (NumLastLogin). We consider the members who made the last login as the exiting members on the day. Based on our data, 322 members left before the OC closed in July 2014. We observe that there are one new joiner and one leaver during most of the days.

The OC members created 56,777 posting and 428,769 comments during the active period. They made 149,425 and 365 votes for postings and comments respectively. They also issued complaints 13,862 and 4,248 times for 2,152 postings and 1,548 comments.

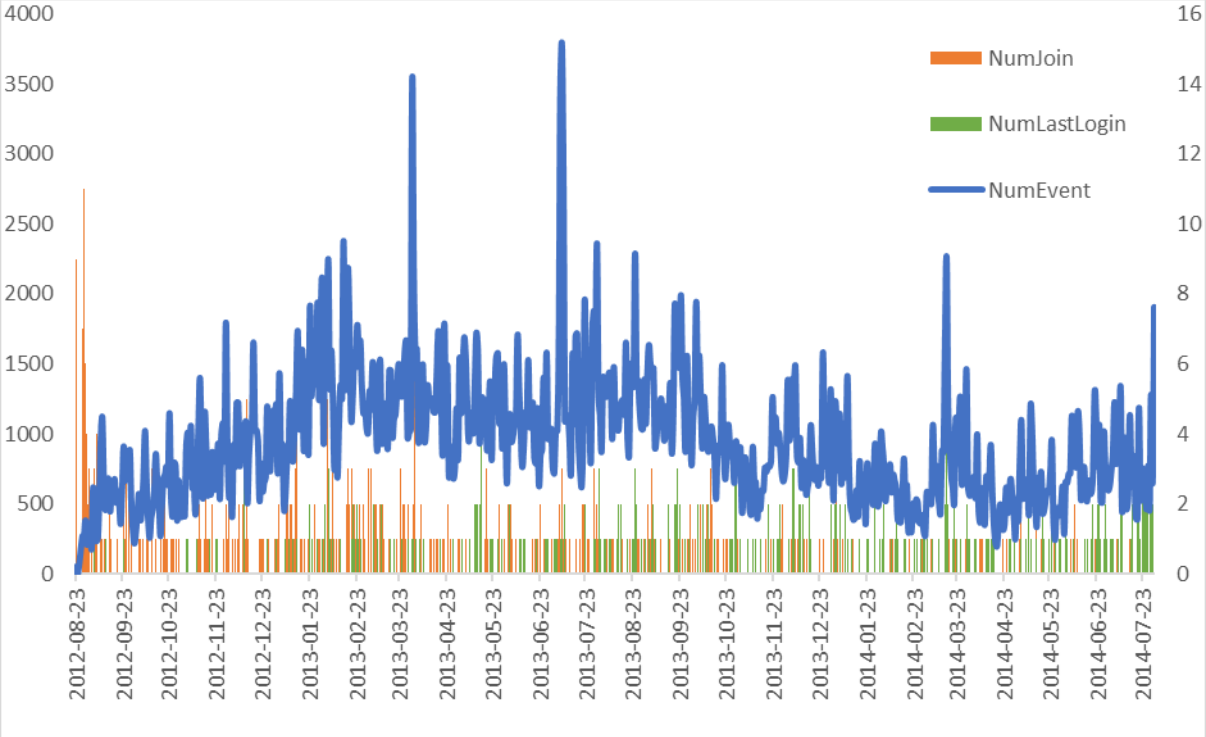


Figure 1. Average number of events, member registrations and leaving.

Registered members exchange messages 17,503 times among them. After removing anonymous member's events (the OC allows non-registered post, vote, and complain articles and comments but not messaging) and cleaning member messages, there are 39,026 postings, 289,917 comments, 29,388 votes for postings, 135 votes for comment, 352 complaints for postings, 98 complaints for comments, and 9678 message exchanges, respectively.

Classification	Members	No of postings	No of comments	Joining date	Last login date
Newcomers	296	46	461	2013/6/9	2014/2/5
Olbies	103	232	1,255	2012/11/17	2014/1/10
Total	399	94	666	2013/4/17	2014/1/29

Table 1. Descriptive statistics of each group.

Based on our membership classification, Table 1 shows descriptive statistics for registered member's events. 103 olbies write 232 postings and 1,255 comments on average during their tenure in the OC. On the other hand, 296 newcomers wrote an average of 46 posts and 461 comments. In short, olbies more contribute to the OC than newcomers. Olbies wrote postings five times more and made comments three

folds than newcomers. Date of joining and last login in Table 1 is the average value of each group. We can say that olbies joined earlier as well as stayed longer than newcomers based on these average joining date and last login date.

To identify different "assimilation process" sequences between newcomers who settled successfully and not, we choose some "actively participated" newcomers from the samples. We assume that the "actively participated" newcomers are evolved during their participation and interactions with other members. There are no well-known criteria to classify "evolving" newcomers from data, we inspected each newcomer's activities and tenure (staying period, calculated by last login date – join date) in the OC and chose 38 members out of 296 newcomers. Finally, there are 103 olbies, 38 selected newcomers, and 292 newcomers in our data set.

Category	**Type**	**Code**	**Description**	**Available events**
Actor	Creator	CR	The initiator of the conversation (postings or messages)	
	Non-Creator	NC	Respondent of conversation	
Action	Posting	P	Write a new posting (document)	CRP
	Comment	C	Write a new comment for a posting	CRC, NCC
	Vote	V	Vote a posting or a comment	CRV, NCV
	Complaint	D	Declare a post or a comment	NCD
	Message	M	Send a message to another member	CRM, NCM

Table 2. Data coding structure.

After classifying members as olbies, selected newcomers, and the others, we develop an ontology to describe possible events created by members in the OC. Each actor (member in the OC) can do actions of five types. In Table 2, we identify two actor types and five types of actions that are important and can make interactions between actors (OC members).

Technically, there can be ten types of events created by OC members. However, there are only eight types of events available. Firstly, there is no NCP type event in OC because every posting is a starting point of discussion. Secondly, there is no actor's self-complaint for her postings or comments since she is able to delete her

problematic postings or comments rather than using the complaint functionality. The type of CRC is made when a member post comments to her own articles. This action type is usually made when a member responds to comments posted by other members. We categorise the remaining comments as NCC. Therefore, most of the comments are classified as NCC. Commenting is an important action within OCs for interactions among members on the shared board of the communities. For votes, there are few cases in which members vote their own contents (postings or comments). This action type is categorised as CRV. This action type indicates a strong tendency of seeking attention from other members as an article that attract more than 10 votes gets a badge of "good article". Other votes are classified as NCV, which is the majority type of vote actions. On the OC, there is no metadata in the database to classify messages into either initial message or response messages. So, we used the title of the messages to classify any messages whose subjects start with "RE:" into respondent type (NCM) and all other messages into initiation type (CRM) which initiates a direct conversation between members.

Member ID	Datetime	Actor	Action	Event
12004	20/09/2012 14:12	CR	M	CRM
12004	20/09/2012 14:13	NC	C	NCC
3917	20/09/2012 14:34	NC	C	NCC
3917	20/09/2012 14:37	NC	C	NCC
8027	20/09/2012 15:14	CR	P	CRP

Table 3. Sample coding results.

Table 3 shows a sample of event sequences on a randomly chosen time period. We code all events following the coding scheme. Since our primary interests are concerned with generative mechanism of assimilation processes of newcomers, we focus on identifying patterns of those assimilation processes

Sequence analysis

We perform sequence analysis using TraMineR, which was implemented on R (Gabadinho et al. 2011). TraMineR is designed to analyse sequences in social science. This package allows various types of data format: state sequence demonstrates temporal duration of sequence on it, so it can be useful when we draw temporal changes of sequences. Event sequence exhibits transitions of events without considering event duration. Event sequence is useful to figure out which sequential

patterns are essential within the sequence because it shows sequential patterns of member's transitional tendency. This tendency can be a candidate of demi-regularity. The interpretation of the discovered demi-regularities is a way to find generative mechanisms.

4.0 Results

Before we start to analyse event sequences, we review structures of each group's sequences – olbies, selected newcomers, and the remaining newcomers by drawing sequence distribution diagram. Since our interests are to reveal a newcomer's assimilation process, we choose five hundred sequences from the beginning for each member and make state sequence using TraMineR. State sequence is generally used to treat temporal duration. We visualise sequence distribution diagram using the member's State sequence in Figure 2. The x-axis of Figure 2 shows the sequential order of each member. Seq000001 displays the first event for each member. Y-axis shows the portion of each event type by colour. Most of the events are NCC – respondent's comment on a posting. The remaining events are CRC, NCV, CRP, CRM, NCM, and NCD. The response to other messages (NCM) and complaint (NCD) rarely happens in the data.

We draw sequence distribution diagrams for each group in the below figure. Compared to olbies, two newcomer groups show different sequence distribution. Generally, newcomers reduce their efforts on their postings (CRC) over time in Figure 3 (b) and (c). On the other hands, olbies write postings more than newcomers (CRP). However, sequence distribution shows us event proportion of given data sets. It does not consider a number of the event for each sequential order. For example, olby's proportion of NCC (responding comments to other's posting) is smaller than newcomer's, but it does not mean the total number of olbies' NCC events is smaller than newcomers'. Now we know three groups having different patterns on their sequences based on sequence distribution diagrams.

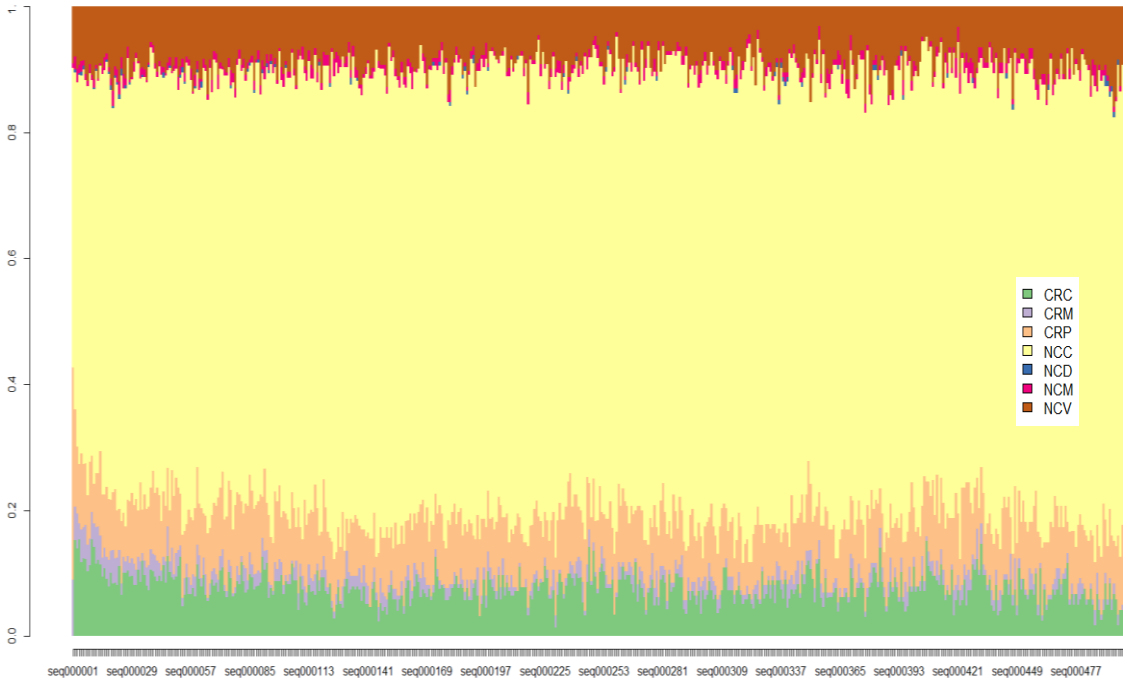
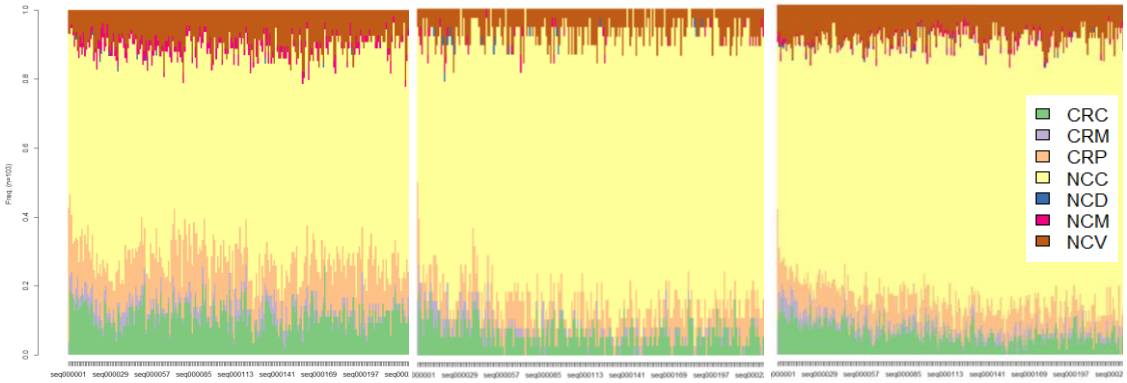


Figure 2. Sequence distribution diagram of all members



(a) Olbies (b) Selected newcomers (c) The other newcomers

Figure 3. Sequence distribution of each group

Our interest in sequence analysis is to identify demi-regularities of members' behaviour during assimilation processes. For this, we transform the state sequence as event sequence using TraMineR's *seqecreate* fucntion. Event sequence is a form of sequence that focuses on the transitions of events. Once the transformation is done we use *seqefsub*, a function of TraMineR (Gabadinho et al. 2011) to extract subsequence patterns from the data. The results show in Table 4. During finding sub-sequences, we use the first hundred state sequences only to focus on initial sequences of each member. If we use more extended sequences to find sub-sequences, it might contain the mid-to-last sub-sequences of each member. We search for sub-sequences using pmin.support value of more than 0.4. Pmin.support is used when we want to search sub-sequence that is seen at least pmin per cent of total sequences. For example,

pmin.support value of the first sub-sequence of Olbies is 0.85 that is seen 80 times out of 94 members' sequence. The count value is the corresponding value of pmin in the total sequences. In Table 4, we report the top 10 sub-sequences for each group. The whole list of sub-sequences is supplied as an appendix file.

Group	**Subsequence**	**Support**	**Count**
All members	(NCC>CRP)	0.73	254
	(NCC>NCV)	0.70	243
	(CRP>NCC)	0.68	238
	(NCV>NCC)	0.68	237
	(CRP>CRC)	0.68	236
	(CRC>NCC)	0.64	223
	(NCC>NCV)-(NCV>NCC)	0.64	221
	(NCC>CRP)-(CRP>NCC)	0.63	219
	(CRP>CRC)-(CRC>NCC)	0.62	216
	(NCC>CRP)-(CRP>CRC)	0.62	215
Olbies	(CRP>CRC)	0.83	78
	(NCC>CRP)	0.80	75
	(CRP>NCC)	0.76	71
	(NCC>CRP)-(CRP>CRC)	0.76	71
	(CRC>NCC)	0.74	70
	(CRP>CRC)-(CRC>NCC)	0.74	70
	(CRP>CRC)-(NCC>CRP)	0.69	65
	(CRP>NCC)-(NCC>CRP)	0.69	65
	(NCC>CRP)-(CRC>NCC)	0.69	65
	(NCC>CRP)-(CRP>NCC)	0.83	78
Selected newcomers	(NCC>NCV)	0.89	34
	(CRP>NCC)	0.87	33
	(NCC>CRP)	0.84	32
	(NCC>NCV)-(NCV>NCC)	0.84	32
	(NCV>NCC)	0.84	32
	(CRC>NCC)	0.82	31
	(NCC>CRP)-(CRP>NCC)	0.82	31
	(NCC>CRP)-(NCC>CRP)	0.82	31
	(CRP>CRC)	0.79	30
	(CRP>CRC)-(CRC>NCC)	0.79	30
Unselected newcomers	(NCC>CRP)	0.68	147
	(NCC>NCV)	0.68	146
	(NCV>NCC)	0.66	142

	(CRP>NCC)	0.62	134
	(NCC>NCV)-(NCV>NCC)	0.61	131
	(CRP>CRC)	0.59	128
	(NCC>CRP)-(CRP>NCC)	0.57	123
	(CRC>NCC)	0.56	122
	(CRP>CRC)-(CRC>NCC)	0.54	116
	(NCC>CRP)-(CRP>CRC)	0.53	114

Table 4. Top 10 sub-sequences for each group from the initial 200 sequences

The group of selected newcomers, who survived until the end of the OC, shows highly similar sub-sequence patterns compared to other groups during the initial stage. 32 ~ 35 members out of 38 show same sub-sequences. On the other hands, 56 ~ 71 per cent of the other newcomers have common sub-sequence patterns in other groups. It means that newcomers are a very similar pattern to each other from the perspective of activities. Based on the results, we can find that there are differences between groups. However, it is hard to find exact differences or demi-regularities that can explain the sequential differences. We use clustering analysis to refine the results.

Earlier, we defined ontologies based on possible actions and actors in an OC. The sub-sequence patterns in Table 4 indicates some differences between groups, but it is very difficult to interpret the results directly. To understand the generative mechanisms of assimilation process in an OC, we cluster sub-sequences using distances between sub-sequences calculated by *seqdist* in TraMineR. We use optimal matching (OM) distances that generate 'edit distances'. Edit distances are the minimal manipulation cost of transforming one sequence to another and is popularised by Abbott (Gabadinho et al. 2011). We use insertion/deletion parameter as one (default value) and substitution-cost matrix (submat) as "CONSTANT". There are two substitution-cost matrix calculation methods – "CONSTANT" and "TRATE". "CONSTANT" method calculates all substitution costs as equal. On the other hand, "TRATE" use the transition rate between states observed in the sequence data. However, transition rate can be changed due to the stage of member's tenure (i.e., investigation, socialisation, or maintenance stage) or member groups (olbies, assimilating newcomers, or read-only newcomers). Since the transition rate can change dynamically due to the actor's status and group, we use the "CONSTANT" method when we calculate (sub) sequence distance.

In the next step, we conduct a clustering analysis by using sequence distances. We use hierarchical clustering with Ward's linkage rather than k-means clustering method

since we do not know how many clusters exist before the analysis. Based on the hierarchical clustering result in Figure 4, we can classify the sub-sequences into either two clusters or four clusters. We check the analysed results through manual inspections. If we follow the two clusters model, the sub-sequences are divided by actors – creator and non-creator. However, we cannot find the demi-regularity differences between newcomers who assimilated and the others under the two clusters model. We inspect the four clusters model: we can classify these four clusters as 1) non-creator's comment dominant sequences (cluster 1 – NCC dominant), 2) creator's postings and response dominant sequences (cluster 2 - CRP-CRC dominant), 3) creator's activity dominant but no-immediate response (cluster 3 – CRP or CRC dominant), and 4) non-creator's vote dominant (cluster 4 – NCV dominant). In other words, these four patterns designate demi-regularity of 1) "everyday participation" - responses to others (cluster 1, NCC dominant), 2) "a responsive writer" – writing a posting and responsive comments to it (cluster 2, CRP-CRC dominant), 3) "dangling efforts" – writing a posting alone or not immediate comments for her postings (cluster 3, CRP or CRC dominant but changed to/from NC events), and 4) "Likes!!" – voting/liking a post/comment (cluster 4, NCV).

In cluster 1, most sub-sequences are started or ended with NCC – non-creator's comment. As we have seen, NCC is the most of the total events. There are 1739 out of 5281 sub-sequences in cluster 1. These sub-sequences show the most common member's events in an OC. We can expect that sub-sequence pattern in cluster 1 is likely to occur mostly, but this pattern is found less likely from olbies compared with newcomers.

Cluster 2 is one of the most important patterns. In cluster 2, sub-sequences include (CRP-CRV) at least once. (CRP-CRV) is composed of writing a new posting and (immediate) successive comments for the posting[1]. This pattern has more importance than other patterns because a member generates a new posting and makes efforts on her new posting persistently – it can increase other members' attention and participation. Also, everyone including a posting writer involved here will be creating or enhancing social networks on the OC, and it will make participants engage more in the future. We can easily expect that sub-sequence patterns in cluster 2 can be higher

[1] It can be possible to write comments for writer's old postings, but it is likely to happen to write comments for the posting that she wrote a minutes ago.

in not only olbies but also the newcomers who assimilated than newcomers who didn't.

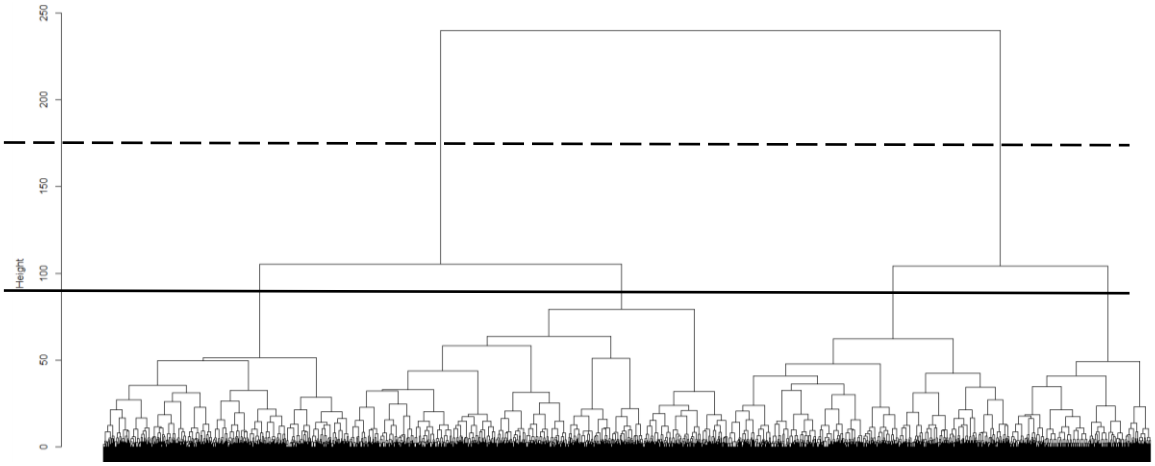


Figure 4. Hierarchical clustering dendrogram of sub-sequences

Cluster 3 also includes creator dominant patterns – CRP or CRC. However, patterns in cluster 3 do not include immediate responses to others. For example, sub-sequences in cluster 3 contain transitions from creator events (CRP or CRC) to non-creator events (NCC, NCV, or NCD) or vice versa. This sub-sequence pattern shows us that a member's content accumulation and responses to others in the OC, but it does not include "immediate" response to other members.

Cluster 4 also contains unique sub-sequence patterns that are not revealed yet. Sub-sequence patterns in cluster 4 a can be interpreted as of a thorough reading of a posting and a reaction of liking it. It is also dealt with as information seeking and acquisition process that is mentioned in prior studies (Morrison 1993). If information seeking is also useful in an online voluntary organisation, these cluster 4 patterns will be higher in selected newcomers rather than unselected newcomers.

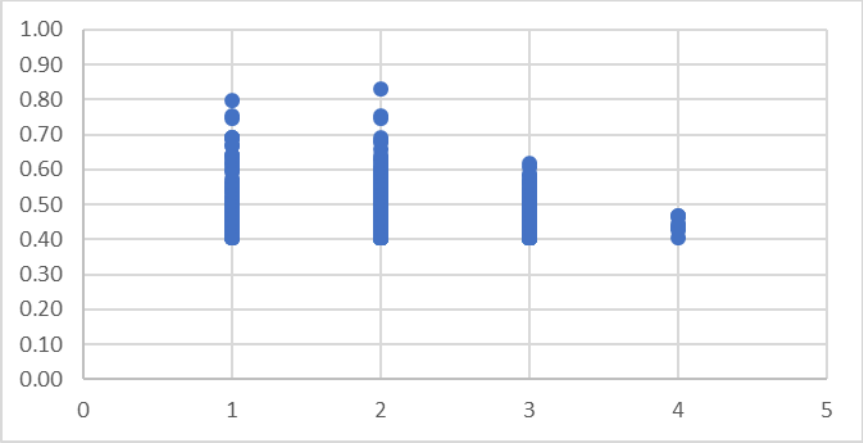


Figure 5. Olbies' sub-sequence distribution for each cluster

Figures 5 to 7 depict sub-sequence distribution for each cluster. The X-axis shows the cluster number. The Y-axis describes the pmin.support percentage of each sub-sequence. We extracted sub-sequences whose pmin.support value is more than 0.4.

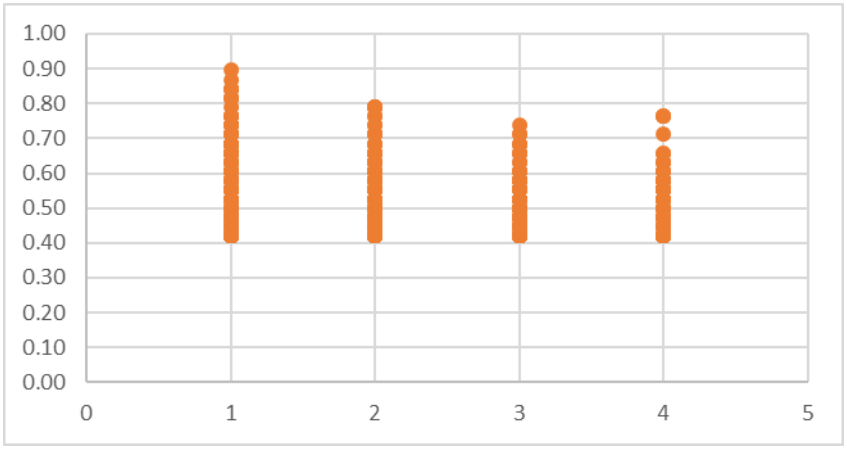


Figure 6. Selected newcomers' sub-sequence distribution for each cluster

Olbies show cluster 2 patterns mostly in Figure 5. They write comments for other's postings, but they write their postings frequently and make immediate responses to their postings. On the other hand, they show low demi-regularities of creator's events that do not include immediate responses to their postings. It can be interpreted that the most writers of olbies group follow cluster 2 patterns that contain immediate responses, so there are less dangling creator's efforts found.

The selected newcomers' sub-sequence patterns are different from olbies as well as unselected newcomers. Selected newcomers show a similar level of participation by writing comments to others (cluster 1, NCC dominant patterns) and writing postings and making immediate responses to them (cluster 2, CRP-CRC dominant patterns). At least selected newcomers (assimilated newcomers) are trying to write a posting and vote a posting or a comment more often than other newcomers. Specifically, selected newcomers to show not only high regular participation (NCC dominant patterns, cluster 1) but also to try to create and interact with other members responsively by their postings (CRP-CRC dominant patterns, cluster 2). They also seek information proactively and leave a sign to read (NCV dominant patterns, cluster 4).

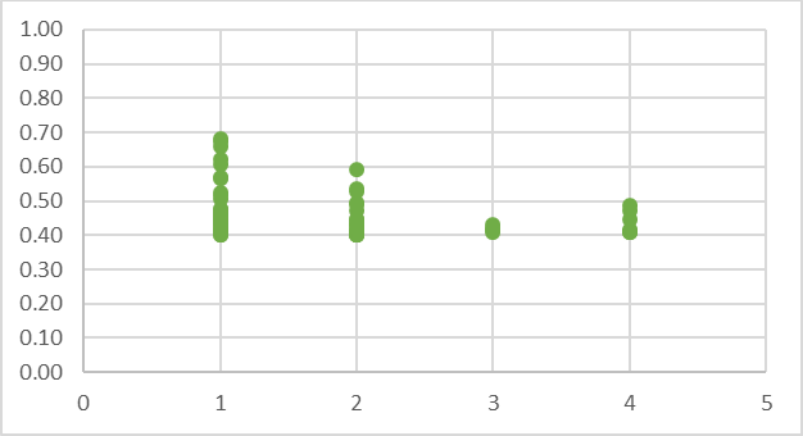


Figure 7. Unselected newcomers' sub-sequence distribution for each cluster

5.0 Conclusion

In this study, we apply a process mining technique, sub-sequence analytics, against hidden footages from a real OC database to understand newcomers' assimilation process. We use a database of two years of data from a Korean celebrity online community. Since this OC maintained only for two years and was moved from another OC, we can identify assimilated newcomers and compare their characteristics and demi-regularities directly with olbies and other newcomers who failed to assimilate. We build sequence data based on the identified ontologies and search common sub-sequences. After searching sub-sequences, we use clustering analysis to identify demi-regularities for OC members. We identify four types of demi-regularities. Based on visualisations, we found that assimilated newcomers show higher demi-regularities on "a responsive writer" (cluster 2) and "Likes!!" (cluster 4). During our data cleansing process, we find that most votes (likes) are done by anonymous users. However, the assimilated newcomers maintain their login status and participate OC with their registered identity. The "Likes!!" is also consistent with the traditional organisation assimilation process. Votes (likes) are given after reading a posting or a comment. In other words, votes are a sign of proactive information seeking (Morrison 1993). Based on the sub-sequence extraction, the assimilated newcomers are also a prolific writer. They are responsive to their postings (CRP-CRC), but they also create the most "dangling" postings. However, they learn by writing something frequently how they receive attention from others. The direct messaging functionalities have no effect or no role in the assimilation process.

We expect this study to expand our understanding of the newcomer's assimilation process in organisations. Our results can be utilised in various area. OC managers can use our results directly to maintain their OCs. Specifically, it is crucial for newcomers to maintain login status during their OC usage. It is also used in another area where membership retention is essential. For example, a practitioner can use our process mining technique for a user of her App to check the user's assimilation status.

References

Abbott, A. (1990) *A Primer on Sequence Methods,* Organization Science (1:4), pp. 375–392.

Arguello, J., Butler, B., Joyce, E., Kraut, R., Ling, K. S., and Wang, X. (2006) *Talk to Me: Foundations for Successful Individual-Group Interactions in Online Communities*, in SIGCHI, pp. 959–968.

Bateman, P. J., Gray, P. H., and Butler, B. S. (2011) *The Impact of Community Commitment on Participation in Online Communities,* Information Systems Research (22:4), pp. 841–854.

Faraj, S., Jarvenpaa, S. L., and Majchrzak, A. (2011) *Knowledge Collaboration in Online Communities,* Organization Science (22:5), pp. 1224–1239.

Faraj, S., and Johnson, S. L. (2011) *Network Exchange Patterns in Online Communities,* Organization Science (22:6), pp. 1464–1480.

Fehr, E., and Gächter, S. (2000) *Cooperation and Punishment in Public Goods Experiments*, American Economic Review (90:4), pp. 980–994.

Fiedler, M., and Sarstedt, M. (2014) *Influence of Community Design on User Behaviors in Online Communities*, Journal of Business Research (67:11), Elsevier B.V., pp. 2258–2268.

Gabadinho, A., Ritschard, G., Müller, N. S., and Studer, M. (2011) *Analyzing and Visualizing State Sequences in R with TraMineR,* Journal of Statistical Software (40:4).

Gaskin, J., Berente, N., Lyytinen, K., and Yoo, Y. (2014) *Toward Generalizable Sociomaterial Inquiry: A Computational Approach for Zooming In and Out of Sociomaterial Routines*, MIS Quarterly (38:3), pp. 849–871.

Hippel, E. von, and Krogh, G. von. (2003) *Open Source Software and the 'Private-Collective' Innovation Model: Issues for Organization Science,* Organization Science (14:2), pp. 209–223.

Jones, G. R. (1986) *Socialization Tactics, Self-Efficacy, and Newcomers' Adjustments to Organizations,* Academy of Management Journal (29:2), pp. 262–279.

Klarner, P., and Raisch, S. (2013) *Move to the Beat--Rhythms of Change and Firm Performance,* Academy of Management Journal (56:1), pp. 160–184.

Kramer, M. W., and Miller, V. D. (1999) *A Response to Criticisms of Organizational Socialization Research: In Support of Contemporary Conceptualizations of Organizational Assimilation,* Communication Monographs (66:4), pp. 358–367.

Langley, A., Smallman, C., Tsoukas, H., and Van de Ven, A. H. (2013) *Process Studies of Change in Organization and Management: Unveiling Temporality, Activity, and Flow*, Academy of Management Journal (56:1), pp. 1–13.

Leenders, R. T. A. J., Contractor, N. S., and DeChurch, L. A. (2016) *Once upon a Time*, Organizational Psychology Review (6:1), pp. 92–115.
Levine, J. M., and Moreland, R. L. (1994) *Group Socialization: Theory and Research*, European Review of Social Psychology (5:1), pp. 305–336.
Morrison, E. W. (1993) *Longitudinal Study of the Effects of Information Seeking on Newcomer Socialization*, Journal of Applied Psychology (78:2), pp. 173–183.
Myers, K. K., and Oetzel, J. G. (2003) *Exploring the Dimensions of Organizational Assimilation: Creating and Validating a Measure*, Communication Quarterly (51:4), pp. 438–457.
Oestreicher-Singer, G., and Zalmanson, L. (2013) *Content or Community? A Digital Business Strategy for Content Providers in the Social Age*, MIS Quarterly (37:2), pp. 591–616.
Oh, W., and Jeon, S. (2007) *Membership Herding and Network Stability in the Open Source Community: The Ising Perspective*, Management Science (53:7), pp. 1086–1101.
Qin, X., Salter-Townshend, M., and Cunningham, P. (2014) *Exploring the Relationship between Membership Turnover and Productivity in Online Communities*, in AAAI, , January 30, pp. 406–415.
Ransbotham, S., and Kane, G. C. (2011) *Membership Turnover and Collaboration Success in Online Communities: Explaining Rises and Falls from Grace in Wikipedia*, MIS Quarterly (35:3), p. 613.
Ray, S., Kim, S. S., and Morris, J. G. (2014) *The Central Role of Engagement in Online Communities*, Information Systems Research (25:3), pp. 528–5