

Peripheral members create and core members integrate: cooperation pattern of members in networks during co-creations

Kunhao Yang (✉ yang-kunhao@g.ecc.u-tokyo.ac.jp)

University of Tokyo

Itsuki Fujisaki

University of Tokyo

Kazuhiro Ueda

University of Tokyo

Research Article

Keywords: core members, peripheral members, co-creations, networks

Posted Date: February 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-250228/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Main Manuscript for

Peripheral members create and core members integrate: cooperation pattern of members in networks during co-creations

Authors: Kunhao Yang^{1*}, Itsuki Fujisaki^{1,2}, Kazuhiro Ueda^{1*}

¹ Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan (〒153-8902).

² Research Fellowship for Young Scientists (DC2), Japan Society for the Promotion of Science (JSPS), Tokyo, Japan (〒102-0083).

* Kunhao Yang, Kazuhiro Ueda

Email: yangkunhao@g.ecc.u-tokyo.ac.jp; ueda@gregorio.c.u-tokyo.ac.jp

Abstract

Cooperation has become the principal way in which creative activities are carried out in modern society. Previous research has shown that in the cooperation of creative activities (i.e., co-creation), different participants can play two completely different roles: some participants are the *originators* who generate initial contents, and others are the *revisors* who provide revisions or coordination. In this research, we investigated the different participants' roles (i.e., the originator vs the revisor) in co-creations and how these roles affected the final cooperation-group outcome. By using cooperation networks to represent the cooperative relationships among participants, we found that peripheral members (i.e., those in the periphery of the co-operation networks) and core members (i.e., those in the centre of the co-operation networks) played the roles of originators and revisors, respectively, mainly affecting the quantity versus the quality of creative outcomes. These results were robust across three different datasets and three different indicators defining core and peripheral members. Our results uncover the cooperation pattern in co-creations and advance the understanding of how co-creations occur in networks.

Main Text

Introduction

Creative activity is regarded as one of the key factors in the development of modern society¹. Since modern society has become increasingly complex, many previous studies²⁻⁷ have found that compared to individual efforts, cooperation has become the main way of carrying out creative activities. In this regard, an important issue is understanding how cooperation occurs among participants during creative activities. Many previous studies⁸⁻¹³ have discussed this issue by comparing the importance of different participants in cooperation during creative activities (i.e., in co-creations). In these studies, co-creation was considered as comprising co-submitting behaviours of original contents (co-submissions of new ideas, software codes, scientific papers, etc.); they then used networks to represent the co-submission relationships (e.g., the co-author relationship in research-paper-writing activities) among participants. Based on these cooperation networks, these studies divided the participants into two types: *core* members and *peripheral* members. The core members were defined as those members occupying the well-connected nodes (participants) in the network, who also made up the information hub of the network. By contrast, peripheral members were defined as those in nodes with few connections far from the information hub⁸⁻¹⁴. Based on these two definitions, these previous studies found that contrary to intuition, the peripheral members were more important than core members in co-creations: the peripheral members submitted more original content than the core members.

However, co-creation is more complex than original content co-submissions. Previous studies^{15,16} found co-creation to be a cyclic (i.e., repeating) process of revisions: after the initial contents were submitted by some participants, these contents always needed several revisions by other participants to become the final outcome (see the illustration in Fig. 1). As a result, different participants could contribute to co-creations through two completely different roles¹⁷⁻²¹: some

participants were the *originators* who generated initial content, while others were the *revisors* who provided revisions or coordination. Therefore, in order to understand co-creations among different participants, it is not enough to focus only on content submission behaviours. The more important issues are what roles the different participants play in co-creations and how these roles affect the final outcome of the cooperation group or community.

-----Fig. 1 about here-----

Based on the above discussions, this research revisits the issue of co-creation and focuses on the different roles of participants in co-creation. Specifically, based on previous studies⁸⁻¹³, we also focus on the core and peripheral members in cooperation networks, and compare their different roles. In the following sections, we empirically analyse three different co-creation activities represented by three different datasets. In particular, we seek to answer two important questions: 1) what different roles (i.e., originator and revisor) the different participants (i.e., the core members and peripheral members) play; and 2) how these roles affect the final outcome of the co-creations.

In discussions of co-creation, previous studies⁸⁻¹³ only focused on the content submission behaviours, ignoring that revisions are also an essential part of co-creation. In this study, we focus on both initial content submission behaviours and revision behaviours. As a result, we found that these two different behaviours were conducted by participants with different network positions. Overall, by analysing a large body of data, our research will provide a deeper understanding of the cooperation pattern among participants with diverse network positions.

Results

1. Overview of the data and measurement of the core-peripheral position

To investigate the participants' different roles in co-creations, we used three datasets collected from websites: 1) SCP-Wiki, representing story-writing activities²², 2) Github, representing software development activities²³, and 3) Idea

Storm, representing idea generation activities²⁴ (see details in the ‘Methods’ section). In all the datasets, the final outcomes (i.e., SCP-stories in SCP-Wiki, source codes in Github, and ideas in Idea Storm) were generated by a cooperation chain of participants: an originator submitted the initial content and the revisors provided revisions and generated the next versions of the content; the final outcome was the final version of the content (see Fig. 1).

To define the core-periphery positions, cooperation networks need to be constructed first. Based on previous studies⁸⁻²¹, we constructed cooperation networks through the cooperation relationships in our three datasets. The participants who worked on the same contents (i.e., the same SCP-stories, source codes, or ideas) were assumed to be connected in the respective networks (i.e., undirected weighted networks; the weight of the connections indicates the times of cooperation between the two members). Based on previous research^{8-14,25-30}, we used three metrics (i.e., core-periphery metric hereafter) to measure the core-periphery positions of participants: *degree*, *k-core*, and *eigenvector centrality* (see the details of computations in the ‘Materials and Methods’ section). Degree⁸⁻¹² indicates whether the focused node (i.e., participants) is well-connected. A larger degree value indicates that the focused node has more connections with other nodes in the network. K-core²⁸⁻³⁰ indicates whether the node is in a cohesive subgroup (i.e., a subgroup with many connections among its members). The large value of k-core reflects a large number of connections among the focused node and its neighbours (i.e., nodes with connections with the focused node). Eigenvector centrality^{13-15,27} indicates whether the focused node is the information hub of the network. A large value of eigenvector centrality reflects that the focused node is in the centre of the information flow of the network. In this research, nodes with larger values of degree, k-core, or eigenvector centrality were regarded as closer to the core of the network. In the following analysis, we show that our results are robust across these three metrics. Additionally, in the following analyses, we will compare the values of our three core-periphery metrics at different time points to identify the core (peripheral) members at different time points. However, the features (e.g., the density) of the networks at different time

points will also affect the core-periphery metrics' values and may make them incomparable. Therefore, we have normalised our three metrics over time to ensure that their values can be compared between different time points (see details in the 'Materials and Methods' section).

2. Role division of core and peripheral members

First, we investigated who played the role of the originator (revisor) in co-creations. The red lines in Fig. 2 show the percentages of the total initial content submitted by participants with different values of core-periphery metrics. The results were consistent across all three datasets regardless of which metric out of the three we used: participants with smaller values of core-periphery metrics (i.e., peripheral members) submitted a larger proportion of initial content, while participants with larger values of core-periphery metrics (i.e., core members) contributed less initial content. Therefore, compared to the core members, the peripheral members seemingly played a more important role in the initial content submissions. However, these results can be easily explained by the fact that there are far more peripheral members than core members in the networks. Therefore, we then examined the relationship between the percentiles of the core-periphery metrics at time point t and the likelihood of submitting initial content at time point $t+1$. This relationship reflects the *possibility* of one core (peripheral) member submitting initial contents at the next time point. The blue lines in Fig. 2 show that in all three datasets, the participants with a smaller value of core-periphery metrics at time point t had a significantly larger likelihood of submitting initial contents at time point $t+1$. In other words, when a participant was a peripheral member, she/he was more likely to submit initial contents, while when she/he became a core member, she/he came to be less likely to submit initial contents. Nevertheless, there is still an alternative explanation for the above results: the peripheral members were more likely to submit initial contents because they were newcomers. In other words, most of the peripheral members are newcomers, and if a newcomer does not get a chance to create initial content, she/he may simply choose to leave the community; therefore, the relationship between the core-periphery positions and the possibility of initial

content submission can only be explained by the time factor. In order to exclude this alternative explanation, we used logistic models to control the number of days that a participant had spent in the community. In other words, the logistic models employed the likelihood of initial content submissions as the dependent variable, the values of the core-periphery metrics as the independent variable, and the number of days that a participant has spent in the community as the control variable. The insets in Fig. 2 show the predicted possibilities of initial content submissions for the participants who had different values of core-periphery metrics but shared the same number of days in the communities (which equals the average number of days for which all participants spent in the communities). The results indicate that even after controlling for the time factor, the peripheral members still have a significantly larger possibility of submitting initial contents at the next time point than the core members. These results were consistent across all datasets using all the different metrics.

-----Fig. 2 about here-----

To summarise, the above results showed that in co-creations, the peripheral members mainly played the role of originator and the core members mainly played the role of revisor. However, these results were only based on the quantity (i.e., possibility) of initial content submissions. Another question then arises of content quality: how did the fact of being a core (or peripheral) member affect the content quality by role? In particular, previous research¹⁹⁻²¹ found that core members often have more corresponding professional knowledge on creative activities than peripheral members. Therefore, if the core members play the role of originator (or revisor) in the co-creations, can they significantly benefit the quality of the final outcome? To address this issue, we analysed the quality of the final outcomes for our three datasets in the next section.

3. Quality of original content submitted by peripheral and core members

We employed participant-generated quality metrics, *rate* and *star*, to measure the quality of the final outcomes (i.e., SCP-stories and codes) in the SCP-Wiki and

Github data, respectively. Both indicate how many other participants in the communities thought that the SCP-stories or codes were good (see details in the ‘Materials and Methods’ section). In Idea Storm, we used a specialist-generated metric to measure the quality of ideas, namely, whether or not the team of specialists at Dell (i.e., Dell is the owner of the Idea Storm website) thought that the submitted ideas were valuable (also see details in the ‘Materials and Methods’ section). We then examined the relationships between the percentiles of the core-periphery metrics of originators at t and the quality of the final outcomes that they submitted at $t+1$. The quality of final outcomes was measured by (a) the average rate of the stories in SCP-Wiki, (b) the average number of stars of the code repositories in Github, and (c) the proportion of valuable ideas in Idea Storm.

The green lines in Fig. 3 show that the value of quality metrics slightly increased when the values of the core-periphery metrics increased. These results seemingly indicate that the final outcomes originating from the core members had a higher quality. However, we considered that the higher quality could also be caused by subsequent revisions, because previous research¹⁹⁻²¹ indicates that the contents that had originated from the core members easily received more revisions by other core members.

-----Fig. 3 about here-----

To control for the influence of revisors and other related factors, we then used regression models considering five important control variables¹⁹⁻²¹. We controlled 1) the average value of the core-periphery metrics of the revisors, 2) the number of content revisions, 3) the time point of the initial content publication, 4) the number of days from the originator’s first participation (i.e., submission or revision) until now (which indicated the days for the originator being observed), and 5) the originator’s number of previous participations. Control variables 4) and 5), which indicated the previous experience of the originators on the websites, were added because previous research¹⁹⁻²¹ found that the quality of outcomes from very experienced participants (i.e., originators) in the communities (i.e., the websites) tended to be overestimated.

In contrast, the quality of outcomes from inexperienced participants in the communities tended to be underestimated. By controlling all the related factors, the results of regression models (in Tables 1–3) showed that the values of the core-periphery metrics of the originators did not have a significant positive effect on content quality (i.e., the three core-periphery metrics did not have significant positive coefficients in any of the regression models; see the specific coefficients in the fourth row of Tables 1–3). These regression results indicate that regardless of their greater corresponding professional knowledge, core members having higher values of the core-periphery metrics did not provide benefit to the quality of the final outcome through the role of **originators** in co-creations.

Additionally, a more important point in the regression results is that the average value of the core-periphery metrics of the revisors had a significant positive effect on the final outcome quality. This result indicates that compared to those with small values of core-periphery metrics (i.e., revisors who were peripheral members), the revisors with large values of core-periphery metrics (i.e., revisors who were core members) were those who mainly improved the final outcome quality. In other words, this result indicated that the core members benefited from the quality of the final outcome through the role of **revisors** in co-creations. Based on these results, an important question about what the revisions from the core members changed was resolved. In the next section, we investigate this question.

-----Table 1 about here-----

-----Table 2 about here-----

-----Table 3 about here-----

4. Core members' revisions contribute to the integration of writing style of different stories

To examine the effect of revisions on the changes in the content, we only focused on the SCP-Wiki data. Since the SCP-Wiki provides a unique *revising history* function, we can compare the revised parts between different versions of SCP-stories. Using this function, we gathered what parts of the SCP-stories were specifically changed during every revision. To measure the change between two versions of SCP-stories, we used a document-embedding model³¹ based on a powerful representation model for language, BERT³², to project every version of the SCP-stories onto a 768-length vector. The relationships among these vectors can reflect the relationships among the corresponding versions of SCP-stories³¹⁻³³: a larger cosine distance of the vectors represents a smaller similarity between the versions (see details in the ‘Materials and Methods’ section). We then computed the average cosine distances between the focal version and all other versions of all SCP-stories in the SCP-Wiki. This average cosine distance indicates the focal version’s distance to the centre of all stories in the SCP-Wiki community. We then traced the change in cosine distance caused by revisions. We computed the change in the cosine distance between the k times and the $k-1$ times versions of the same story:

$$\Delta \text{CosineDistance}_{t_k}^{\theta} = \text{CosineDistance}_{t_k}^{\theta} - \text{CosineDistance}_{t_{k-1}}^{\gamma}$$

The superscript and subscript represent the author and the submission time point of the drafts, respectively.

A negative value of $\Delta \text{CosineDistance}_{t_k}^{\theta}$ indicates that the revision made the focused version closer to other stories in the SCP-Wiki community. Conversely, a positive value indicates that the revision made the focused version further from other stories in the SCP-Wiki community.

Using a linear regression model (i.e., OLS), we then examined the relationship between the changes in the average cosine distances and the values of the core-periphery metrics of the revisor. In the regression, we controlled the following values: 1) the average cosine distance of the $k-1$ times version (i.e., $\text{CosineDistance}_{t_{k-1}}^{\gamma}$); 2) the number of content revisions; 3) the time point of the publication of the focal version; 4)

the number of days from the revisor's first participation (i.e., submission or revision) until now (which also represented the number of days for the revisor being observed); 5) the revisor's number of previous participations; and 6) the SCP story ID of the corresponding revision (as dummy variables that only affected the intercept). Variables 4) and 5), which indicated the previous experience of the revisors, were added because previous studies¹⁹⁻²¹ found that participants with more experience always tended to show a more conservative attitude towards the contents with larger average cosine distances. All regression results (in Table 4) showed significant negative correlations between the values of the core-periphery metrics of the revisor and the change in the average cosine distance (for degree: *coefficient* = -0.01, *p-value* < 0.01; for k-core: *coefficient* = -0.03, *p-value* < 0.01; for eigenvector centrality: *coefficient* = -0.02, *p-value* < 0.01). Based on previous studies³³⁻³⁵, we consider that the decrease in the average cosine distance among the versions of stories can be explained as reflecting the integration of the writing style among SCP-stories (see the illustration in Fig. 4): through revisions by the core members, different SCP-stories began to share the same writing style; as a result, the average distance between them became smaller; additionally, through this integration of writing style, the quality of the final version of the SCP-stories was also improved.

-----Table 4 about here-----

-----Fig. 4 about here-----

Discussion

Based on the above results, this study found the cooperation pattern among different participants in co-creation: the peripheral members create and the core members integrate. Specifically, the peripheral members generated most of the initial content submissions in co-creations. Based on these initial contents, core members provided their revisions and integrations, which improved the quality of the final outcomes in the co-creations.

Many previous studies⁸⁻¹³ have discussed the importance of peripheral members in co-creations. However, these studies only focused on the content submission behaviours and ignored the possibility that revisions were also an important part of co-creation. In this research, we focused on both initial content submission behaviours and revision behaviours. In this way, we found that the core members and the peripheral members contributed to co-creations from different sides through their roles: the peripheral members mainly contributed to the quantity of initial content through the role of originator, while the core members mainly contributed to the quality of the final outcomes through the role of revisor. Therefore, based on the co-creations among the peripheral members and the core members, the community can achieve a large number of creative outcomes with high quality. In this regard, this insight also contributes to creative organisation design in the real world. To make an organisation more creative, managers should not only depend on the core members of the R&D team but should also exploit the members outside the R&D team to increase the participants' diversity of network position in co-creations.

Materials and Methods

1. Description of three datasets: SCP-Wiki, Github, and Idea Storm

In this research, we elicited data from SCP-Wiki (<http://www.scp-wiki.net/>), Github (<https://github.com/>), and Dell's Idea Storm (<http://www.ideastorm.com/>) to investigate the different roles of participants in co-creations. The SCP-Wiki is a website for participants to submit and discuss their original science fiction stories²³. Its data records the author, content, and time point for every SCP-story submission and revision. Github is one of the largest software development platforms in the world²⁴. The Github data records the author and the time point of every source code upload and revision. Idea Storm is a website designed by Dell to collect interesting ideas from users²⁵. The Idea Storm data records the author and the time point for every idea submission and discussion.

Using the APIs provided by these three websites, we finally gathered a) 92,477 drafts (i.e., different revised versions) of 4,653 SCP-stories that were submitted by 3,405

participants in the SCP-Wiki from January 2009 to December 2018; b) 99,230 code repertories from January 2010 to April 2020 that were programmed in *Python* and received at least one *pull request* (that is, a revision from other participants; see details in the next section); and c) 837 ideas that had at least one *comment* from other participants (see the details about the comments in the next section) from 6,333 members from February 2007 to October 2018.

2. Construction of cooperation networks

Concerning our three datasets, we constructed the respective cooperation networks based on the cooperation relationships between participants. In general, all three networks were undirected weighted networks. The connections (i.e., edges) indicate that the two participants have cooperation on the same content, and the weights of the connections indicate the times of cooperation between the two participants. Specifically, in SCP-Wiki, if two participants had worked on the same SCP story either as an originator or a revisor, they were connected in the cooperation network. In Github, the owners of the code repertories (i.e., originators) and the participants who submitted the *pull request* for the code repertories (i.e., the revisors) were connected. A pull request is a revision sent by a revisor to request that the owner of a code repertory update the source code. In Idea Storm, the originators of the ideas and the participants who sent *comments* on the ideas (i.e., the revisors) were connected. The comment is a function in Idea Storm for members to discuss how to refine each other's ideas. In this way, we constructed three cooperation networks for our three datasets.

3. Computation of core-periphery metrics and their normalization

Previous research^{8-14,25-30} has indicated two ways to identify the core members in networks: 1) based on the connections among nodes (i.e., participants); and 2) based on the information flow among nodes. In this research, we used both methods.

Based on the connections among nodes, we employed two indicators: the *degree* and the *k-core*. The value of degree indicates the number of other nodes connected to the focal node. Therefore, a larger degree value indicates that the focal node has more connections with other nodes in the network. Additionally, previous research⁸⁻¹² found

that core members not only have a high number of connections but also have a high density of connections around them. In other words, the core members are always in the cohesive subgroups of the network. In order to consider both the number of connections of the focal node and the cohesiveness of its subgroup, we also employed *k-core* (also called *k-plex*)²⁸⁻³⁰. A node with the value k , according to the *k-core*, is in a subgraph where every node is connected to k other nodes in this subgroup. In other words, a larger *k-core* value reflects that the focal node has more connections with other nodes and is in a more cohesive subgroup.

Based on the information flow among nodes, we also used *eigenvector centrality* as an indicator. Eigenvalue centrality identifies the information hubs of the network^{13-15,27}. Intuitively, the computation of eigenvector centrality can be considered as a voting process in the network; at first, every node in the network is assigned the same number of votes. Then, in every iteration of voting, the focal node deploys its votes to its neighbours (i.e., other nodes have connections with the focal node) based on the weights of the connections. Meanwhile, the focal node also receives votes from its neighbours and uses these votes to vote in the next iteration. After a sufficiently large number of iterations, every node's votes will converge, and this number is taken as the given node's eigenvector centrality²⁷. In mathematical terms, the numbers of votes will converge to the components of the first eigenvector (i.e., the eigenvector with the largest eigenvalue) of the adjacency matrix of the network (i.e., a matrix using zero and positive numbers, respectively, to represent whether a connection exists between two nodes)²⁷. Therefore, a larger eigenvector centrality indicates that the focal node receives a larger number of votes from its neighbours. If we consider the voting process as an information diffusion process, the nodes with large eigenvector centrality can be considered information hubs in the network.

To summarise, we used degree, *k-core*, and eigenvector centrality to identify the core members in networks. For all three indicators, a larger value of the metrics indicates that the focal node is closer to the core of the networks. Additionally, in this research, we compared the values of the above three metrics between different time points to identify

the core members of the networks at different time points. However, since the networks also changed with time, the values of the metrics at different time points were not directly comparable. To solve this problem, we normalised our metrics over time to ensure that their values indicate the same core-periphery position even at different time points. For degree and k-core, we used the number of connections in the whole network at a certain time point to normalise their values (i.e., $\hat{X} = \frac{X}{|E|}$, where \hat{X} is the normalised value of the metric and $|E|$ is the number of connections in the whole network at a certain time point). For eigenvector centrality, we normalised it by the norm of the first eigenvector at a certain time point. In this way, regardless of the difference in time, a large value of one of our metrics indicates that the focal node was closer to the core of the networks.

4. Quality metrics of the content and the regression models

To measure the quality of content in our datasets, we used three different metrics. In SCP-Wiki, we used the *rate* to measure the quality of the story²². The rate is a participant-generated indicator. For every registered participant in the SCP-Wiki, she/he can either vote ‘1’ for an SCP-story, which indicates that she/he thought that the content of the story was good, or vote ‘-1’ for the story, which indicates that she/he thought the content was not good. The final value of *rate* for an SCP-story is the sum of all ‘1’ and ‘-1’ ratings. For instance, 201 participants voted for scp-001 (i.e., the first story in the SCP-website); 191 of them voted ‘1’ and 10 of them voted ‘-1’. In this case, the scp-001 rating was 181. In the Github data, we employed the number of *stars* for every code repertory to measure its quality²³. Stars simply indicated how many registered participants in Github thought that the source code in the repertory was good. Finally, in Idea Storm, according to previous research^{24,36}, we used the *idea status* to measure idea quality. The idea status was a five-class metric generated by a specialist team at Dell. The idea status shows the evaluation of the submitted ideas. It includes ‘Submitted’, ‘Archived’, ‘Acknowledged’, ‘Partly Implemented’, and ‘Implemented’. According to previous research^{24,36}, ideas with a ‘Partly Implemented’ or ‘Implemented’ status were defined as valuable ideas. Therefore, in Idea Storm, the quality of ideas is a binary variable with a value of 0 (i.e., *not valuable*) and 1 (i.e., *valuable*).

In order to examine the relationship between the originator's value of the core-periphery metrics and the quality of the content, we conducted three regression models. For the SCP-Wiki data and the Github data, we used linear regression models (i.e., OLS). For the Idea Storm data, we used a binary logistic regression model because the quality of ideas was a binary variable. The formula for the regressions is as follows:

For the SCP-Wiki data:

$$\begin{aligned}
 & \textit{Rate on logarithmic scale} \\
 & = \alpha_1 \textit{Core Periphery Metric of the originator on logarithmic scale} \\
 & + \alpha_2 \textit{Average K – core of revisors on logarithmic scale} \\
 & + \alpha_3 \textit{Number of content revisions on logarithmic scale} \\
 & + \alpha_4 \textit{Time point of publication} + \alpha_5 \textit{Days being observed of the originator} \\
 & + \alpha_6 \textit{The originator's number of previous participations} + \alpha_0
 \end{aligned}$$

For the Github data:

$$\begin{aligned}
 & \textit{Number of Stars on logarithmic scale} \\
 & = \alpha_1 \textit{Core Periphery Metric of the originator on logarithmic scale} \\
 & + \alpha_2 \textit{Average K – core of revisors on logarithmic scale} \\
 & + \alpha_3 \textit{Number of content revisions on logarithmic scale} \\
 & + \alpha_4 \textit{Time point of publication} + \alpha_5 \textit{Days being observed of the originator} \\
 & + \alpha_6 \textit{The originator's number of previous participations} + \alpha_0
 \end{aligned}$$

For the Idea Storm data,

$$\begin{aligned}
 & \textit{Odds Ratio(the idea is valuable or not)} \\
 & = \Lambda(\alpha_1 \textit{Core Periphery Metric of the originator on logarithmic scale} \\
 & + \alpha_2 \textit{Average K – core of revisors on logarithmic scale} \\
 & + \alpha_3 \textit{Number of content revisions on logarithmic scale} \\
 & + \alpha_4 \textit{Time point of publication} + \alpha_5 \textit{Days being observed of the originator} \\
 & + \alpha_6 \textit{The originator's number of previous participations} + \alpha_0)
 \end{aligned}$$

Here, $\Lambda(\vec{v})$ is the logistic function, $(1 + e^{-\vec{v}})^{-1}$, and the ‘core periphery metric’ is the value of one of the three indicators, that is, degree, k-core, and eigenvector centrality, for the originator.

Since the quality of one content was affected both by the features of the originator and by the features of the revisors¹⁹⁻²¹, in all models we added 1) the average value of the core-periphery metrics of all corresponding revisors, 2) the number of content revisions, 3) the time point of the content publication, 4) the number of days from the originator’s first participation (i.e., submission or revision) until now, and 5) the originator’s number of previous participations as control variables. As we explained in the ‘Results’ section, control variables 4) and 5), which indicated the previous experience of the originator, were added because previous research¹⁹⁻²¹ found that contents from experienced participants (i.e., originators) in the communities (i.e., the websites) tended to be overestimated. In contrast, contents from inexperienced participants in the communities tended to be underestimated.

The statistics of all variables in the three regressions can be found in the ‘Supplementary Information’.

5. Document-embedding and revision behaviour analysis in SCP-Wiki

In order to examine how the revision behaviours changed the content, we used SCP-Wiki data. SCP-Wiki provides a *history* function that allows its users to see the previous drafts (i.e., previous versions) of stories and to compare the previous text with the revised text. In order to measure the change in content, we used sentence embedding based on the latest representation model for natural language: BERT³¹ (hereafter referred to as *sentence-BERT*). Since BERT uses contextual word-embedding³², the sentence-BERT can capture a trivial semantic change in the revisions³¹. According to a previous study³¹, this sentence embedding model outperformed other document-embedding models in text classification tasks. Additionally, the cosine distance between the embedding results of this model showed a very strong correlation for the semantic relationship between the embedded texts³¹⁻³⁵. As explained in the Results section, we first calculated the average

cosine distance between the focal version and the versions of all the SCP-stories. We then traced the change in average distances by every revision.

We built a multiple linear regression model (i.e., OLS) to examine the relationship between the values of the core-periphery metrics of the revisor and the change in the average distance of the focal version. It is worth noting that since not all revisions changed the text content of the story (e.g., some of them only added pictures, revised punctuation, or adjusted the display), we finally focused on the revisions that changed the text contents (in total, we analysed 43,693 revisions). The formula for the regression model is as follows:

$$\begin{aligned}
 & \textit{Change of average cosine distance} \\
 & = \alpha_1 \textit{Core Periphery Metric on logarithmic scale} \\
 & + \alpha_2 \textit{Average Cosine Distance of the previous version} \\
 & + \alpha_3 \textit{Number of content revisions on logarithmic scale} \\
 & + \alpha_4 \textit{Time point of publication} \\
 & + \alpha_5 \textit{Days being observed of the revisor} \\
 & + \alpha_6 \textit{The revisor's number of previous participations} \\
 & + \alpha_0 \textit{ID of SCP Story}
 \end{aligned}$$

Here, the ‘core periphery metric’ is the value of one of the three indicators, that is, degree, k-core, and eigenvector centrality, for the focal node.

In the regression, we controlled the following values: 1) the average cosine distance of the $k-1$ -times version (i.e., $\textit{CosineDistance}_{t_{k-1}}^y$); 2) the number of content revisions; 3) the time point of the story publication; 4) the number of days from the revisor’s first participation (i.e., submission or revision) until now (which also represented the days for the revisor being observed); 5) the revisor’s number of previous participations; and 6) the SCP story ID of the corresponding revision (as dummy variables that only affected the intercept). Control variables 4) and 5), which indicate the previous experience of the revisor, were added because previous studies¹⁹⁻²¹ suggested that the experienced participants of communities (i.e., the website) were largely constrained

by their previous experience and tended to be more conservative towards versions with a large average distance. Therefore, experienced participants tended to reduce the average distance in the focal versions¹⁹⁻²¹.

Additionally, the $\alpha_0|ID \text{ of SCP Story}$ in the formula indicated that the unique ID of every SCP-story, which was added as a fixed effect in the model, only affected the constants³⁷. Thus, the model estimated 4,653 different constants for the drafts under 4,653 different stories.

The statistics of every variable in the model are shown in the ‘Supplementary Information’.

Code Availability and Data Availability

The R-code during the current study and the three datasets analysed during the current study (including data for making Figures) are available in the Mendeley Data:

<http://dx.doi.org/10.17632/t8nxnf472r.1>

Acknowledgments

This study was supported by JSPS KAKENHI Grant Number JP16H01725 and JST CREST Grant Number JPMJCR19A1 for the third author.

Author Contributions

K.Y., I. F., and K. U. designed the research. K.Y. gathered and analysed the data. K.Y., I. F., and K. U. wrote the paper.

Competing Interest Statement

The authors declare no competing interests.

References

1. Schumpeter, J. A. *The Nature and Essence of Economic Theory* (Routledge, 2017).
2. Pentland, A. *Social Physics: How Social Networks Can Make Us Smarter* (Penguin, 2015).
3. Afuah, A. & Tucci, C. L. Crowdsourcing as a solution to distant search. *Acad. Manage. Rev.* **37**, 355-375 (2012).
4. Azoulay, P., Graff Zivin, J. S. & Wang, J. Superstar extinction. *Q. J. Econ.* **125**, 549-589 (2010).
5. Guimera, R., Uzzi, B., Spiro, J. & Amaral, L. A. N. Team assembly mechanisms determine collaboration network structure and team performance. *Science* **308**, 697-702 (2005).
6. Singh, J. & Fleming, L. Lone inventors as sources of breakthroughs: Myth or reality? *Manag. Sci.* **56**, 41-56 (2010).
7. Wuchty, S., Jones, B. F. & Uzzi, B. The increasing dominance of teams in production of knowledge. *Science* **316**, 1036-1039 (2007).
8. Lakhani, K. R. *The Core and the Periphery in Distributed and Self-Organizing Innovation Systems*. (Massachusetts Institute of Technology, 2006).
9. Painter, D. T., Daniels, B. C. & Laubichler, M. D. Innovations are disproportionately likely in the periphery of a scientific network. *Theory Biosci* 1-10 (2019).
10. Petersen, A. M. Quantifying the impact of weak, strong, and super ties in scientific careers. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E4671-E4680 (2015).
11. Csermely, P. The wisdom of networks: a general adaptation and learning mechanism of complex systems: the network core triggers fast responses to known stimuli; innovations require the slow network periphery and are encoded

- by core-remodeling. *Bioessays* **40**, 1700150-1700161 (2018).
12. Barcomb, A., Kaufmann, A., Riehle, D., Stol, K. J. & Fitzgerald, B. Uncovering the periphery: A qualitative survey of episodic volunteering in free/libre and open source software communities. *IEEE Trans. Softw. Eng.* **40**, 962-980 (2018).
 13. Cattani, G. & Ferriani, S. A core/periphery perspective on individual creative performance: social networks and cinematic achievements in the Hollywood film industry. *Organ. Sci.* **19**, 824-844 (2008).
 14. Wang, X., Lan, Y. & Xiao, J. Anomalous structure and dynamics in news diffusion among heterogeneous individuals. *Nat. Hum. Behav.* **3**, 709-718 (2019).
 15. Basadur, M., Runco, M. A. & VEGAxY, L. A. Understanding how creative thinking skills, attitudes and behaviors work together: a causal process model. *J. Creat. Beh.* **34**, 77-100 (2000).
 16. Horng, J. S. & Hu, M. L. The mystery in the kitchen: culinary creativity. *Creat. Res. J.* **20**, 221-230 (2008).
 17. Wenger, E. *Communities of Practice: Learning, Meaning, and Identity* (Cambridge Univ. Press, 1999).
 18. Butler, B., Sproull, L., Kiesler, S. & Kraut, R. Community effort in online groups: who does the work and why? In *Leadership at a Distance: Research in Technologically Supported Work* (ed. Weisband, S. P.) 171-194 (Psychology Press, 2002).
 19. Lave, J. & Wenger, E. *Situated Learning: Legitimate Peripheral Participation* (Cambridge Univ. Press, 1991).
 20. Arazy, O., Nov, O., Patterson, R. & Yeo, L. Information quality in Wikipedia: the effects of group composition and task conflict. *J. Manag. Inf. Syst.* **27**, 71-98 (2011).
 21. Arazy, O., Ortega, F., Nov, O., Yeo, L. & Balila, A. Functional roles and career paths in Wikipedia. In *Proc. 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* 1092-1105 (ACM, 2015).
 22. Pallante, M. E. *Secure, Contain, Protect: Building a Digital Folklore Mythos through Collaborative Legend Creation* (University of North Carolina, 2017).

23. Wu, L., Wang, D. & Evans, J. A. Large teams develop and small teams disrupt science and technology. *Nature* **566**, 378-382 (2019).
24. Yang, K., Fujisaki, I. & Ueda, K. Interplay of network structure and neighbour performance in user innovation. *Hum. Soc. Sci. Commun.* **6**, 1-8 (2020).
25. Bakshy, E., Karrer, B. & Adamic, L. A. Social influence and the diffusion of user-created content. In *Proc. 10th ACM Conference on Electronic Commerce* 325-334 (ACM, 2009).
26. Steinert-Threlkeld, Z. C. Spontaneous collective action: peripheral mobilization during the Arab Spring. *Am. Political Sci. Rev.* **111**, 379-403 (2017).
27. Bonacich, P. Power and centrality: a family of measures. *Am. J. Sociol.* **92**, 1170-1182 (1987).
28. Balasundaram, B., Butenko, S. & Trukhanov, S. Novel approaches for analyzing biological networks. *J. Comb. Optim.* **10**, 23-39 (2005).
29. Seidman, S. B. & Foster, B. L. A graphtheoretic generalization of the clique concept. *J. Math. Sociol.* **6**, 139-154 (1978).
30. Balasundaram, B., Butenko, S. & Hicks, I. V. Clique relaxations in social network analysis: the maximum k-plex problem. *Oper. Res.* **59**, 133-142 (2011).
31. Reimers, N. & Gurevych, I. Sentence-bert: sentence embeddings using siamese bert-networks. Preprint at <https://arxiv.org/abs/1908.10084> (2019).
32. Devlin, J., Chang, M. W., Lee, K. & Toutanova, K. Bert: pre-training of deep bidirectional transformers for language understanding. Preprint at <https://arxiv.org/abs/1810.04805> (2018).
33. Vajjala, S. & Banerjee, S. A study of N-gram and embedding representations for native language identification. In *Proc. 12th Workshop on Innovative Use of NLP for Building Educational Applications* 240-248 (ACM, 2017).
34. Kozłowski, A. C., Taddy, M. & Evans, J. A. The geometry of culture: analyzing the meanings of class through word embeddings. *Am. Sociol. Rev.* **84**, 905-949 (2019).
35. Fu, Z., Tan, X., Peng, N., Zhao, D. & Yan, R. Style transfer in text: exploration and evaluation. In *Proc. 32th AAAI Conference on Artificial Intelligence* 663-670

- (AAAI, 2018).
36. Bayus, B. L. Crowdsourcing new product ideas over time: an analysis of the Dell IdeaStorm community. *Manag. Sci.* **59**, 226-244 (2013).
 37. Allison, P. D. *Fixed Effects Regression Methods for Longitudinal Data Using SAS* (SAS Institute, 2005).

Figures and Tables

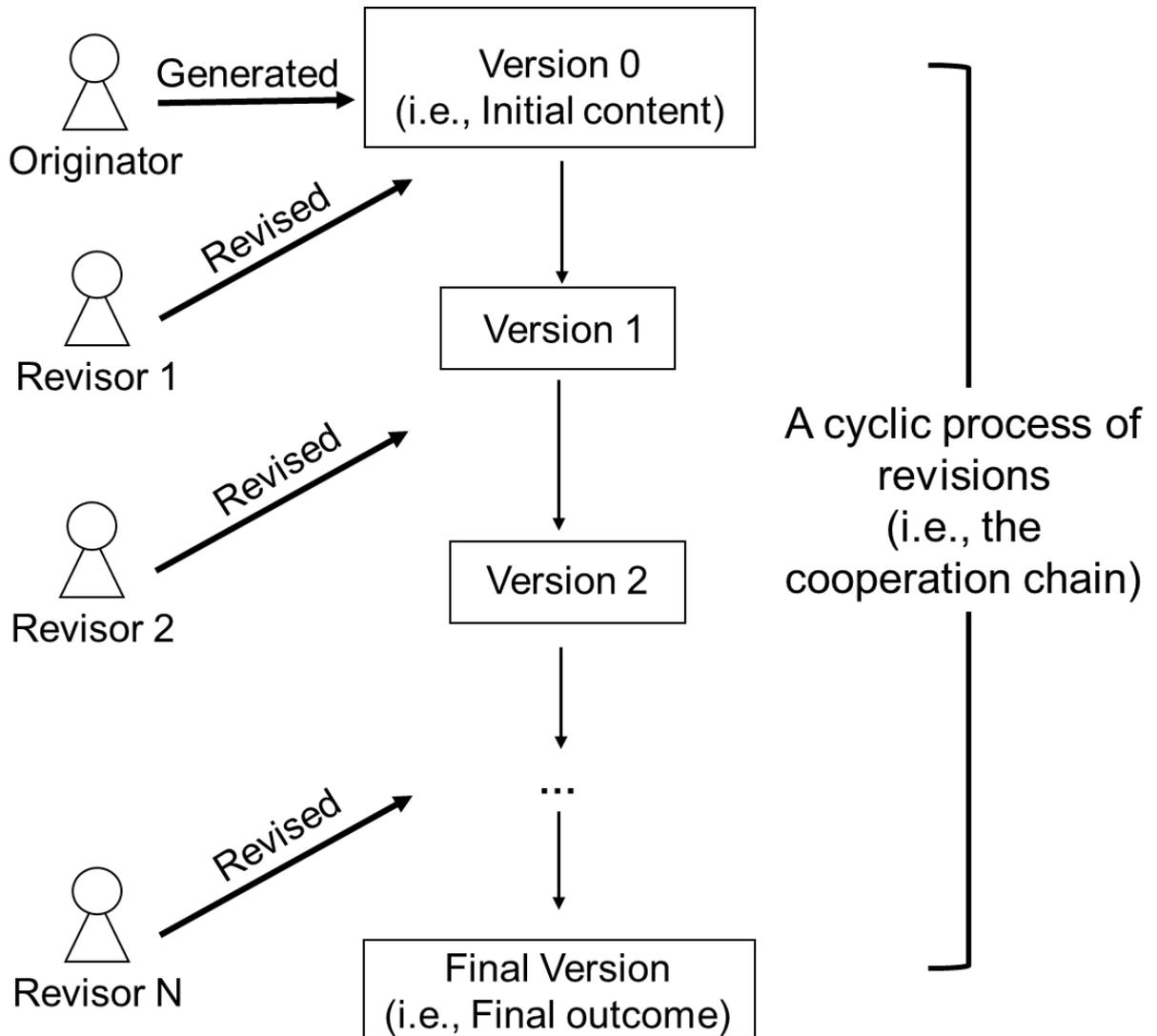
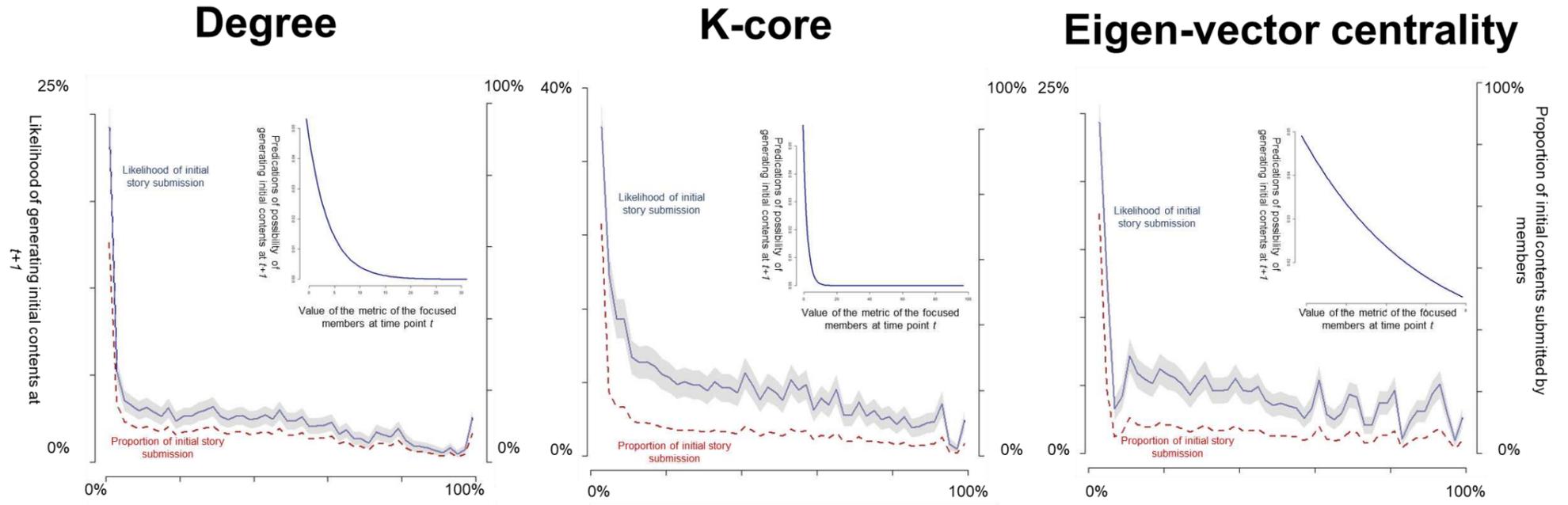
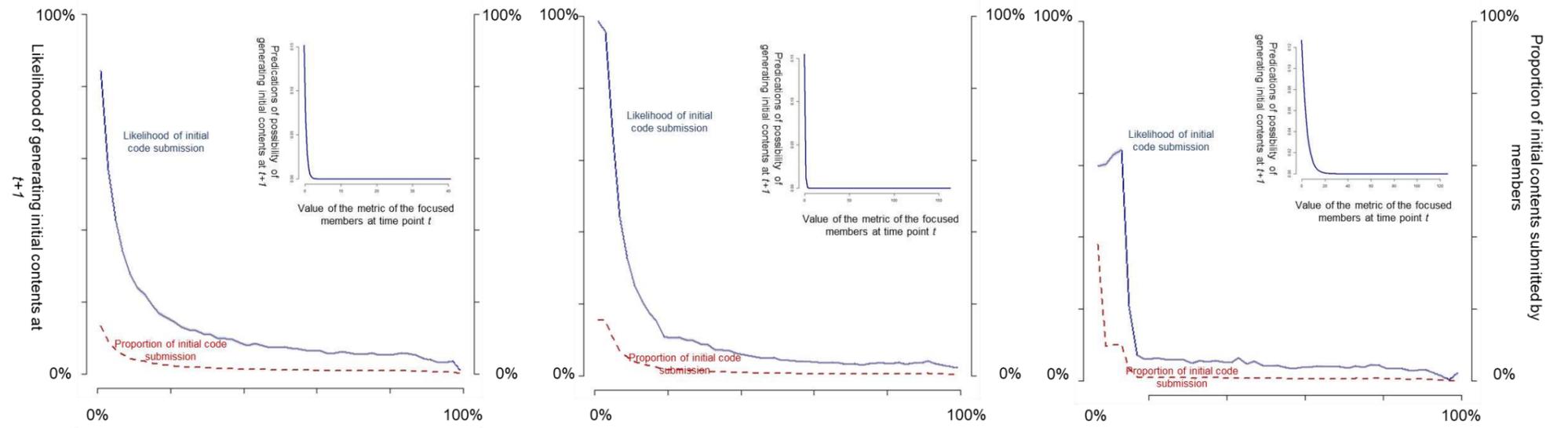


Fig. 1. Illustration of the different roles in co-creation. The cooperation process is a cyclic process of revisions: An originator submits the initial content while other participants (i.e., the revisors) provide revisions and generate new versions of the content in turn. The final output of the cooperation chain is the final version of the content.

SCP data



Github data



Idea Storm data

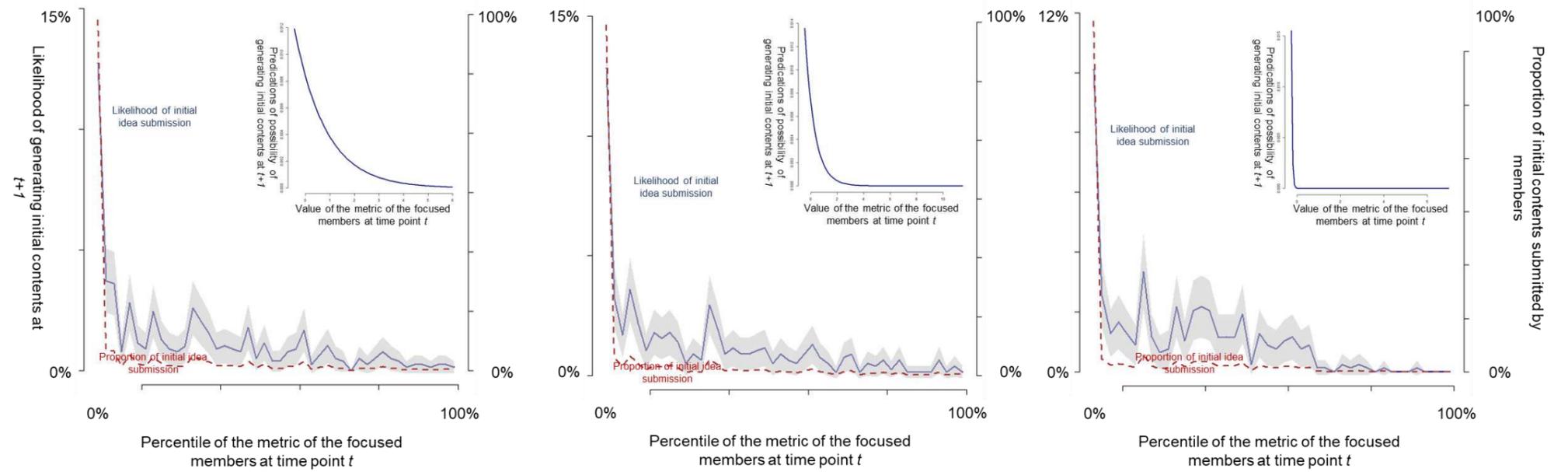
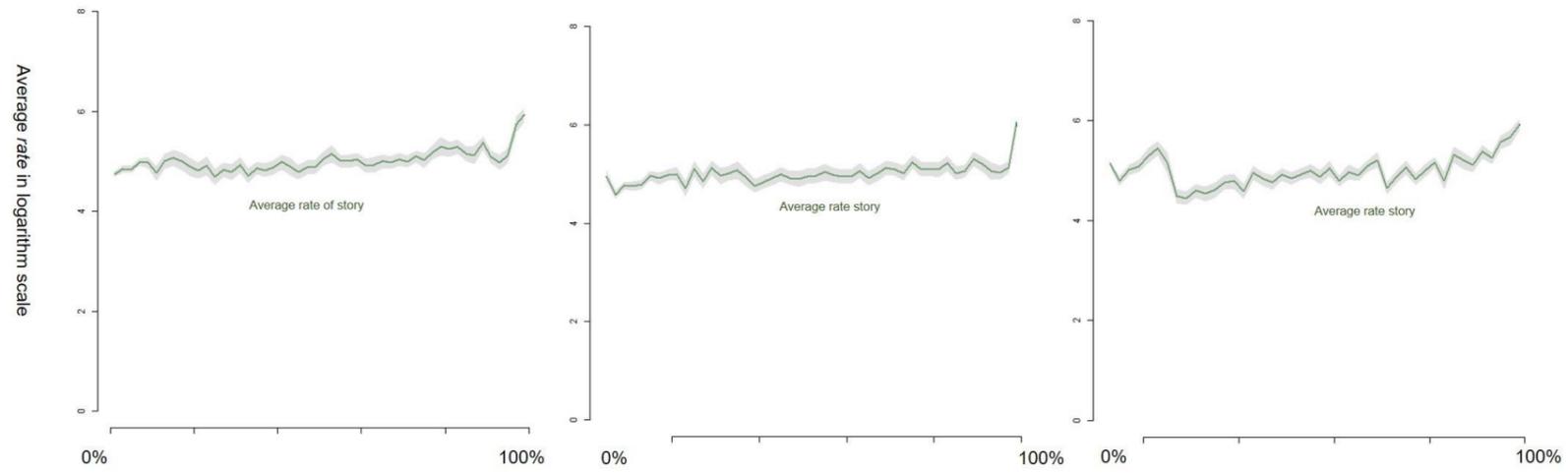
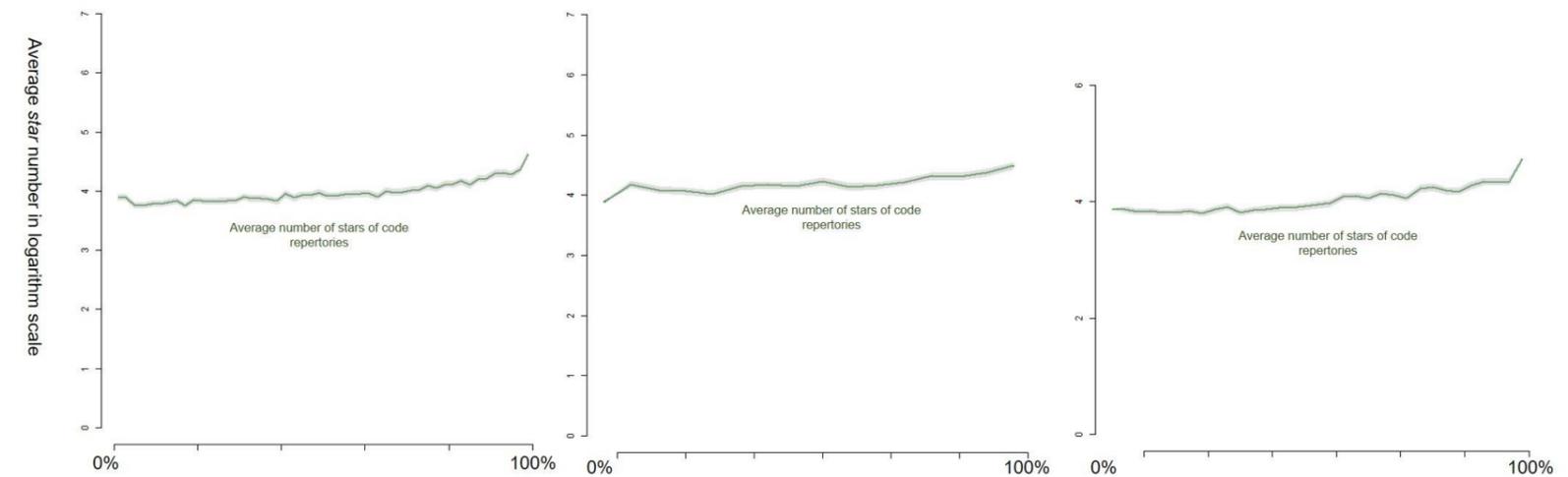


Fig. 2. Illustration of the relationship between the values of the core-periphery metrics and the initial content submissions; In Fig. 2, the relationships between the different values of the core-periphery metrics and the initial content submission in SCP-Wiki, Github, and Idea Storm are shown for each. The panels in the same row show the results for the same dataset but using different metrics (i.e., degree, k-core, eigen-vector centrality). The panels in the same column show the results using the same metric but for different datasets. For every panel, the x -axis indicates the percentile of values of the core-periphery metrics at time point t . The left y -axis indicates participants' likelihood of submitting an initial content at time point $t+1$. The right y -axis indicates the proportion of initial contents submitted by the participants to all the initial contents in the communities. The red lines corresponding to the right y -axis show what percentages of initial contents were submitted by participants with different values of the core-periphery metrics. The blue lines corresponding to the left y -axis show what percentages of participants with different values of core-periphery metrics would submit an original content at the next time point. The grey area shows the 95% confidence intervals of the blue lines, generated by two-tailed t -tests (note that the confidence intervals for the Github data are too narrow to be seen). The insets show the predicted possibilities of initial content submissions by the participants who had different values of core-periphery metrics but shared the same number of days in the communities (which equals the average number days that all participants spent in the communities). The predicted possibilities of initial content submissions were generated by a logistic model with the likelihood of initial content submissions as the dependent variable, the values of the core-periphery metrics as the independent variable, and the number of days that a participant spent in the communities as the control variable.

SCP data **Degree** **K-core** **Eigen-vector centrality**



Github data



Idea Storm data

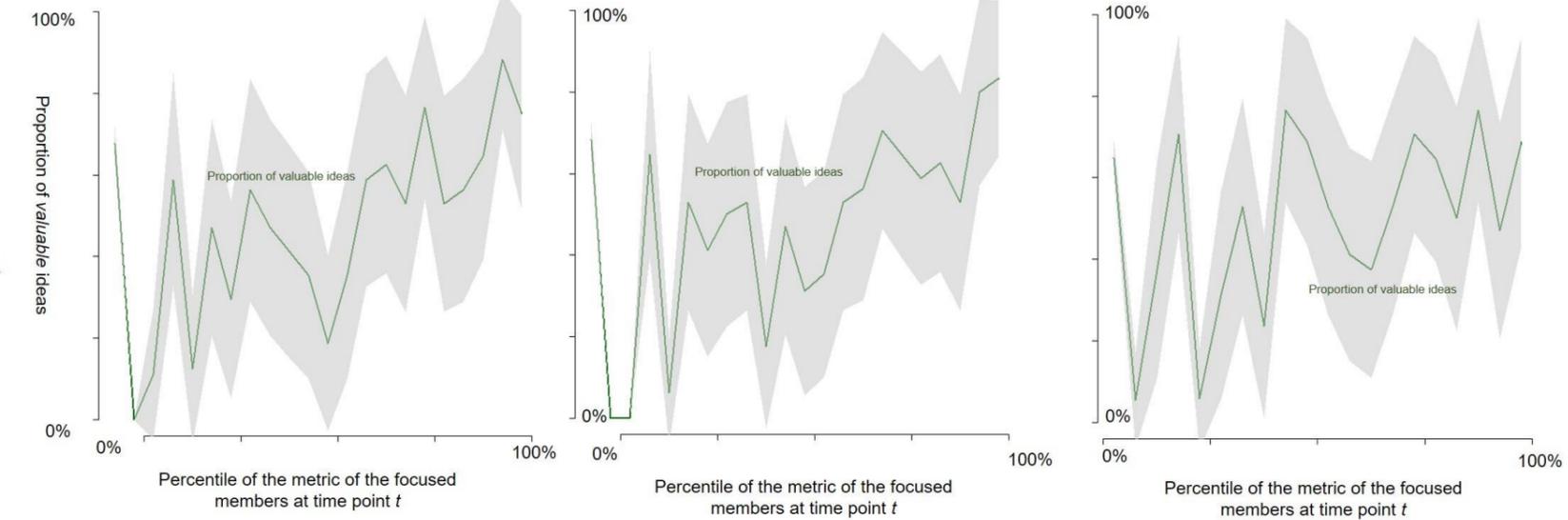


Fig. 3. Illustration of the relationship between the values of the core-periphery metrics and the quality of final outcomes; in Fig. 3, the relationships between the different values of the core-periphery metrics and the values of the content-quality metrics in SCP-Wiki, Github, and Idea Storm are shown for each. The panels in the same row show the results for the same dataset but using different metrics (i.e., degree, k-core, eigenvector centrality). The panels in the same column show the results using the same metric but for different datasets. For every panel, the x -axis indicates the percentile of values of the core-periphery metrics at time point t . The y -axis indicates the value of each quality metric of the content; in SCP-Wiki data, it represents the average rate on a logarithmic scale; in Github data, it represents the average star number on a logarithmic scale; in Idea Storm data, it represents the average proportion of valuable ideas. The green lines show the relationship between originators' values of the core-periphery metrics and the average quality of the final outcomes. The grey area shows the confidence intervals of the average value generated by the two-tailed t -test (note that the confidential intervals in Github data were too narrow to be seen). Note that, in the Idea Storm data, the proportion of valuable ideas fluctuated. The reason for this is considered as follows: Most participants only submitted a small number of ideas (e.g., one or two ideas); in case even only one idea was evaluated as valuable. Therefore, it inevitably generated a large fluctuation (e.g., from 0 % to 50 %) in the proportion of valuable ideas.

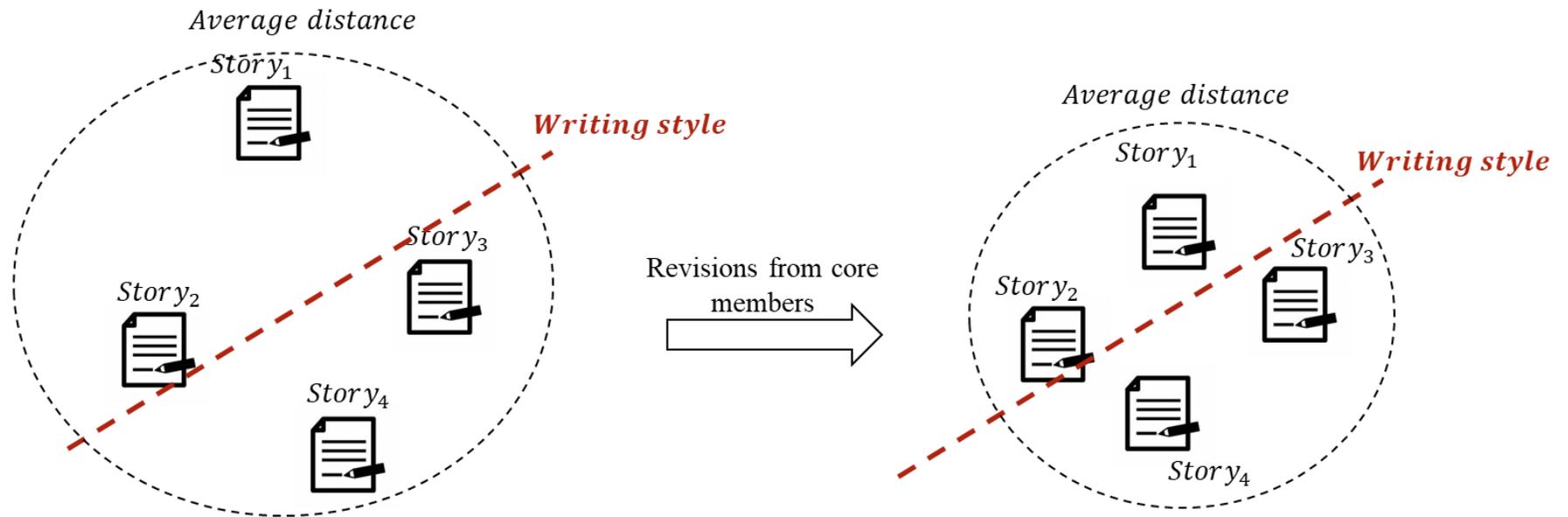


Fig. 4. The illustration of the integration role of core members. The decreases in the average cosine distance among the versions of stories can be explained as an integration of the writing style among SCP-stories: by revisions of the core members, different SCP-stories began to share the same writing style; as a result, the average distance between them became smaller.

Table 1. The results of the regression of content quality for SCP-Wiki data; the coefficients of the core-periphery metrics (i.e., the independent variable) are shown in the fourth row.

Dependent Variable: Rate in logarithm scale in SCP-Wiki					
Core-periphery metric: degree		Core-periphery metric: k-core		Core-periphery metric: eigen-vector centrality	
Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
Degree of originator in logarithm scale	0.027 (0.021)	K-core of originator in logarithm scale	0.008 (0.019)	Eigen-vector centrality of originator in logarithm scale	0.002 (0.018)
Average degree of revisors in logarithm scale	0.275*** (0.030)	Average k-core of revisors in logarithm scale	0.183*** (0.028)	Average eigen-vector centrality of revisors in logarithm scale	0.097*** (0.020)
Number of content revisions	0.293*** (0.014)	Number of content revisions	0.285*** (0.014)	Number of content revisions	0.266*** (0.014)
Time point of publication	0.027 (0.017)	Time point of publication	0.040** (0.017)	Time point of publication	0.057*** (0.016)
Days for the originator being observed	-0.062*** (0.016)	Days for the originator being observed	-0.035** (0.016)	Days for the originator being observed	-0.028* (0.015)
The originator's number of previous participations	-0.064** (0.029)	The originator's number of previous participations	-0.164*** (0.027)	The originator's number of previous participations	-0.286*** (0.016)
Constant	-0.000 (0.013)	Constant	-0.000 (0.013)	Constant	-0.000 (0.013)
R^2	0.247	R^2	0.239	R^2	0.237
Observations: 4,653					

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; Standard Error shown in (); since all coefficients in the table were estimated based on the standardized variables, the sizes of the coefficients are comparable.

Table 2. The results of the regression of content quality for Github data; the coefficients of the core-periphery metrics (i.e., the independent variable) are shown in the fourth rows.

Dependent Variable: Number of stars in logarithm scale in Github					
Core-periphery metric: degree		Core-periphery metric: k-core		Core-periphery metric: eigen-vector centrality	
Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
Degree of originator in logarithm scale	-0.008 (0.005)	K-core of originator in logarithm scale	-0.025*** (0.005)	Eigenvector centrality of originator in logarithm scale	-0.008** (0.004)
Average degree of revisors in logarithm scale	0.100*** (0.004)	Average k-core of revisors in logarithm scale	0.108*** (0.005)	Average eigen-vector centrality of revisors in logarithm scale	0.017*** (0.004)
Number of content revisions	0.474*** (0.003)	Number of content revisions	0.472*** (0.003)	Number of content revisions	0.486*** (0.003)
Time point of publication	-0.013*** (0.004)	Time point of publication	-0.008** (0.004)	Time point of publication	0.011*** (0.003)
Days for the originator being observed	-0.002 (0.003)	Days for the originator being observed	0.004 (0.003)	Days for the originator being observed	0.001 (0.003)
The originator's number of previous participations	0.059*** (0.004)	The originator's number of previous participations	0.068*** (0.005)	The originator's number of previous participations	0.011*** (0.003)

Constant	0.00000 (0.003)	Constant	0.00000 (0.003)	Constant	0.00000 (0.003)
R^2	0.242	R^2	0.240	R^2	0.236
Observations: 99,232					

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; Standard Error shown in (); since all coefficients in the table were estimated based on the standardized variables, the sizes of the coefficients are comparable.

Table 3. The results of the regression of content quality for Idea Storm data; the coefficients of the core-periphery metrics (i.e., the independent variable) are shown in the fourth rows.

Dependent Variable: Idea is valuable or not in in Idea Storm					
Core-periphery metric: degree		Core-periphery metric: k-core		Core-periphery metric: eigen-vector centrality	
Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
Degree of originator in logarithm scale	-0.549*** (0.154)	K-core of originator in logarithm scale	-0.845*** (0.156)	Eigenvector centrality of originator in logarithm scale	-0.029 (0.072)
Average degree of revisors in logarithm scale	5.631*** (0.425)	Average k-core of revisors in logarithm scale	5.436*** (0.397)	Average eigen-vector centrality of revisors in logarithm scale	1.569*** (0.157)
Number of content revisions	0.350* (0.186)	Number of content revisions	0.348** (0.171)	Number of content revisions	0.210** (0.102)
Time point of publication	-0.334* (0.184)	Time point of publication	-0.352* (0.198)	Time point of publication	-0.782*** (0.256)
Days for the originator being observed	-0.029 (0.120)	Days for the originator being observed	-0.024 (0.120)	Days for the originator being observed	-0.037 (0.091)
The originator's number of previous participations	0.198 (0.192)	The originator's number of previous participations	0.195 (0.186)	The originator's number of previous participations	-1.352*** (0.209)
Constant	0.405*** (0.128)	Constant	0.590*** (0.128)	Constant	0.209* (0.118)
AIC	462.734	AIC	473.470	AIC	783.515
Observations: 837					

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; Standard Error shown in (); since all coefficients in the table were estimated based on the standardized variables, the sizes of the coefficients are comparable.

Table 4. The results of the regression on the change of originality for SCP-Wiki data

Dependent Variable: Change of cosine distance among SCP-Wiki drafts					
Core-periphery metric: degree		Core-periphery metric: k-core		Core-periphery metric: eigenvector centrality	
Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
Degree of revisor in logarithm scale	-0.01** (0.005)	K-core of revisor in logarithm scale	-0.03*** (0.004)	Eigen-vector centrality of revisor in logarithm scale	-0.02*** (0.006)
Average cosine distance of the previous version	-1.88*** (0.01)	Average cosine distance of the previous version	-1.88*** (0.01)	Average cosine distance of the previous version	-1.88*** (0.01)
Number of content revisions	0.05*** (0.008)	Number of content revisions in logarithm scale	0.05*** (0.008)	Number of content revisions in logarithm scale	0.05*** (0.008)
Time point of publication	-0.02*** (0.008)	Time point of publication	-0.03*** (0.008)	Time point of publication	-0.02*** (0.008)
Days for the revisor being observed	-0.02 (0.004)	Days for the revisor being observed	-0.02*** (0.004)	Days for the revisor being observed	-0.01*** (0.004)
The revisor's number of previous participations	0.002 (0.005)	The revisor's number of previous participations	0.001 (0.004)	The revisor's number of previous participations	0.007 (0.005)
Average constant [#]	-0.093	Average constant [#]	-0.092	Average constant [#]	-0.092
R^2	0.440	R^2	0.441	R^2	0.440
Observations: 43,693					

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; Standard Error shown in (); all coefficients in the table were estimated based on the standardized variables. Thus, the sizes of the coefficients are comparable. Since

the regression estimated a different constant for every SCP-story, there were 4,653 different constants.
For simplification, we report the average constant here.

Figures

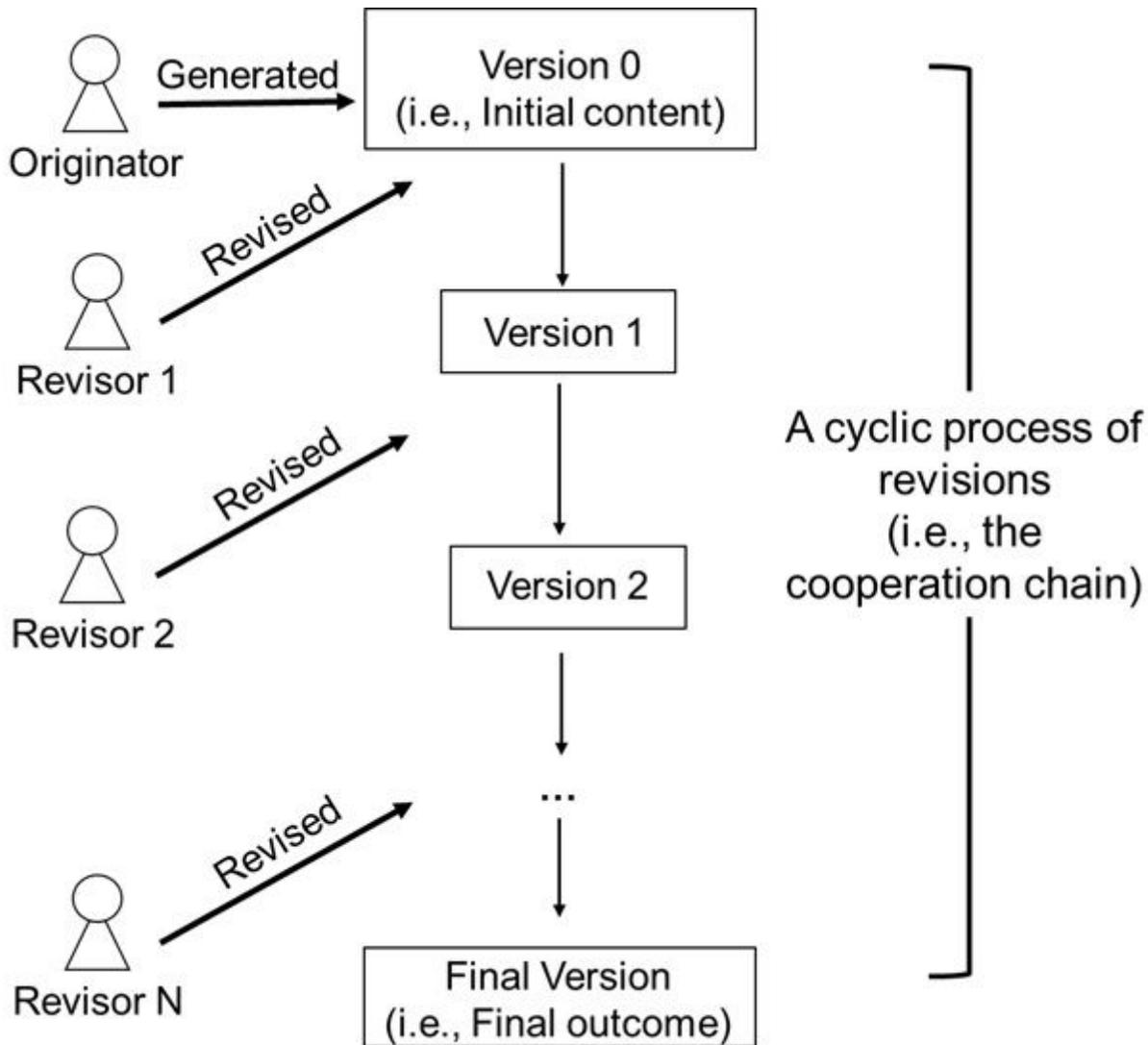


Figure 1

Illustration of the different roles in co-creation. The cooperation process is a cyclic process of revisions: An originator submits the initial content while other participants (i.e., the revisors) provide revisions and generate new versions of the content in turn. The final output of the cooperation chain is the final version of the content.

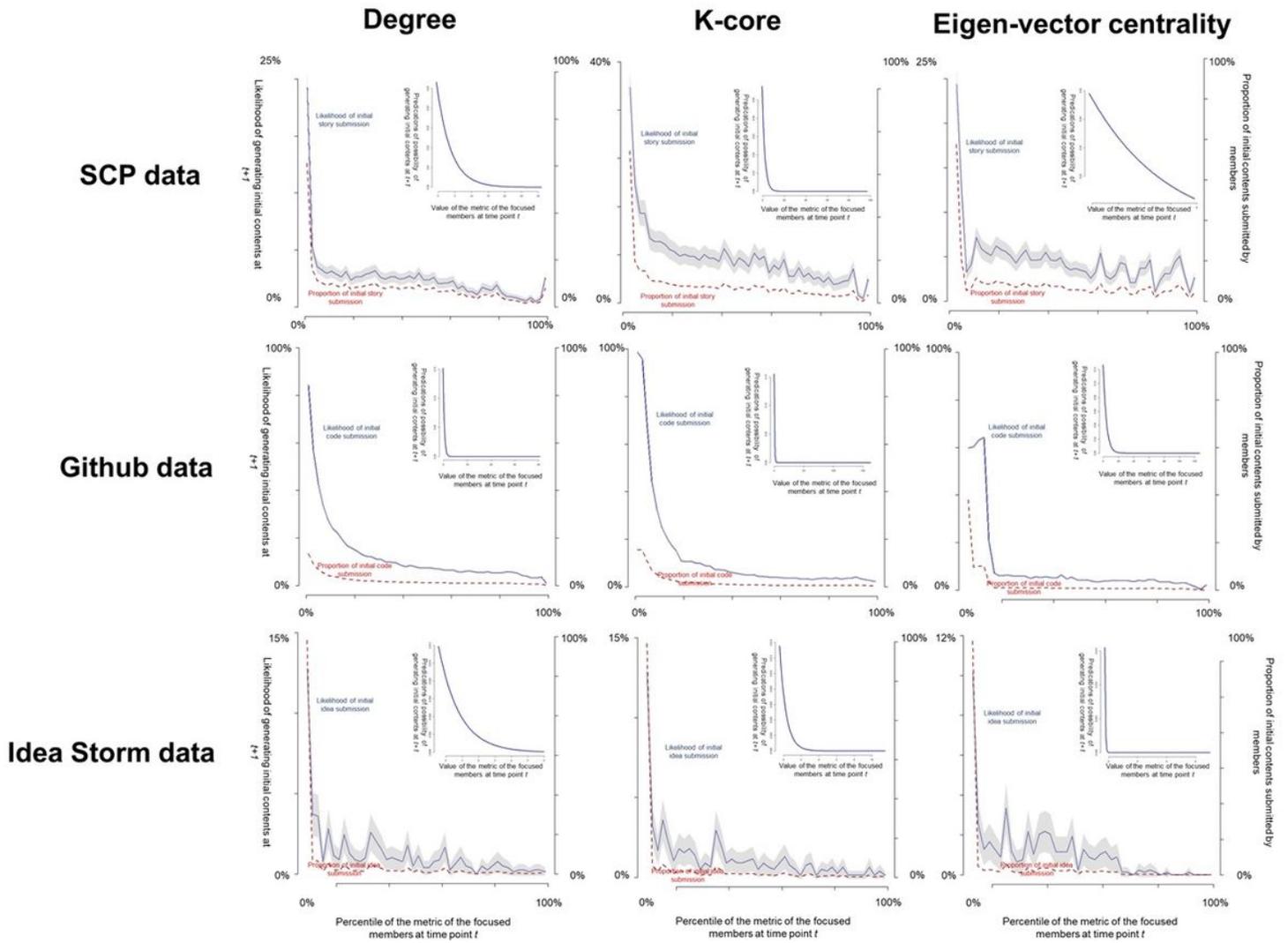


Figure 2

Illustration of the relationship between the values of the core-periphery metrics and the initial content submissions; In Fig. 2, the relationships between the different values of the core-periphery metrics and the initial content submission in SCP-Wiki, Github, and Idea Storm are shown for each. The panels in the same row show the results for the same dataset but using different metrics (i.e., degree, k-core, eigen-vector centrality). The panels in the same column show the results using the same metric but for different datasets. For every panel, the x-axis indicates the percentile of values of the core-periphery metrics at time point t . The left y-axis indicates participants' likelihood of submitting an initial content at time point $t+1$. The right y-axis indicates the proportion of initial contents submitted by the participants to all the initial contents in the communities. The red lines corresponding to the right y-axis show what percentages of initial contents were submitted by participants with different values of the core-periphery metrics. The blue lines corresponding to the left y-axis show what percentages of participants with different values of core-periphery metrics would submit an original content at the next time point. The grey area shows the 95% confidence intervals of the blue lines, generated by two-tailed t-tests (note that the confidence intervals for the Github data are too narrow to be seen). The insets show the predicted possibilities of

initial content submissions by the participants who had different values of core-periphery metrics but shared the same number of days in the communities (which equals the average number days that all participants spent in the communities). The predicted possibilities of initial content submissions were generated by a logistic model with the likelihood of initial content submissions as the dependent variable, the values of the core-periphery metrics as the independent variable, and the number of days that a participant spent in the communities as the control variable.

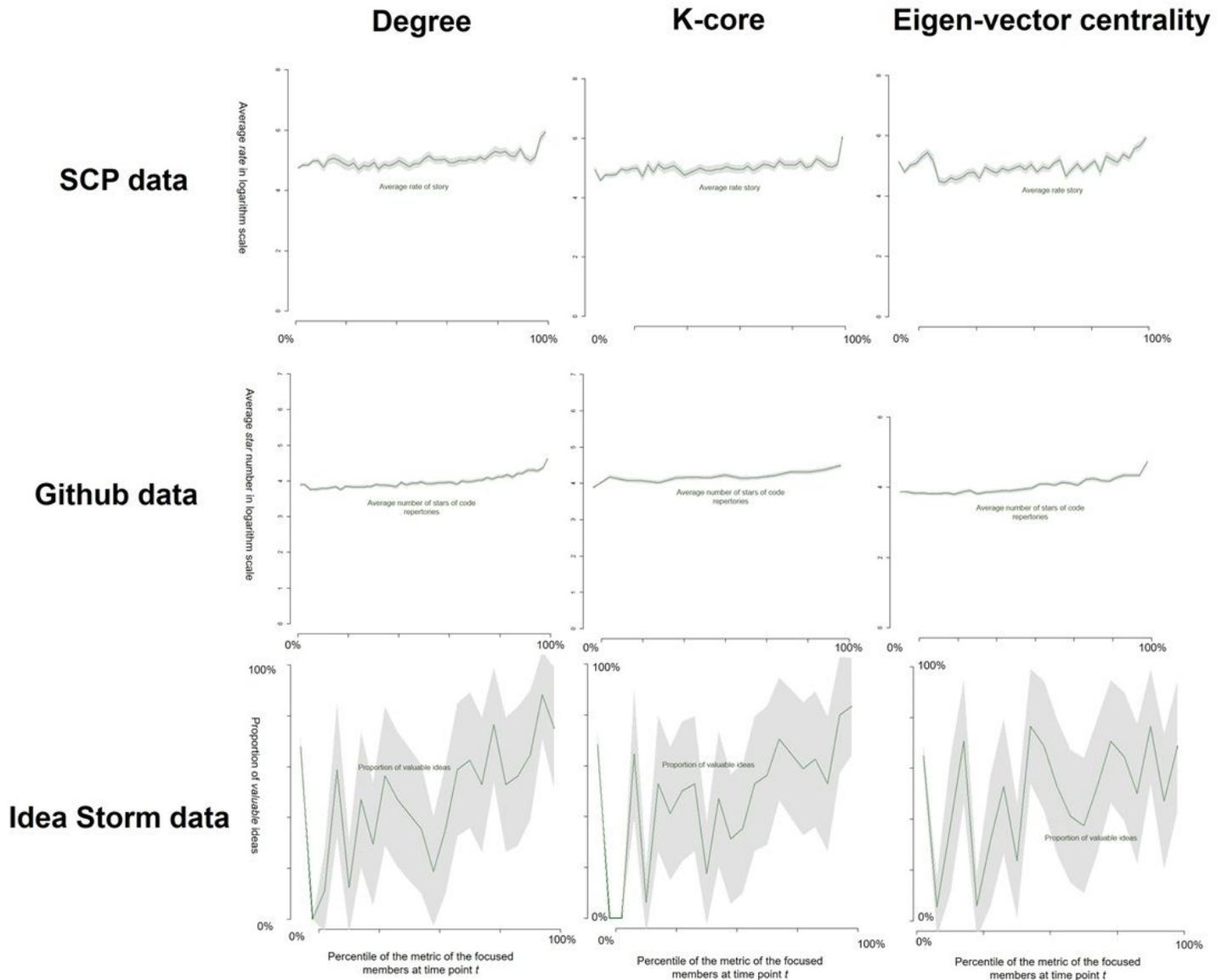


Figure 3

Illustration of the relationship between the values of the core-periphery metrics and the quality of final outcomes; in Fig. 3, the relationships between the different values of the core-periphery metrics and the values of the content-quality metrics in SCP-Wiki, Github, and Idea Storm are shown for each. The panels in the same row show the results for the same dataset but using different metrics (i.e., degree, k-core, eigenvector centrality). The panels in the same column show the results using the same metric but for different datasets. For every panel, the x-axis indicates the percentile of values of the core-periphery metrics at time point t . The y-axis indicates the value of each quality metric of the content; in SCP-Wiki

data, it represents the average rate on a logarithmic scale; in Github data, it represents the average star number on a logarithmic scale; in Idea Storm data, it represents the average proportion of valuable ideas. The green lines show the relationship between originators' values of the core-periphery metrics and the average quality of the final outcomes. The grey area shows the confidence intervals of the average value generated by the two-tailed t-test (note that the confidential intervals in Github data were too narrow to be seen). Note that, in the Idea Storm data, the proportion of valuable ideas fluctuated. The reason for this is considered as follows: Most participants only submitted a small number of ideas (e.g., one or two ideas); in case even only one idea was evaluated as valuable. Therefore, it inevitably generated a large fluctuation (e.g., from 0 % to 50 %) in the proportion of valuable ideas.

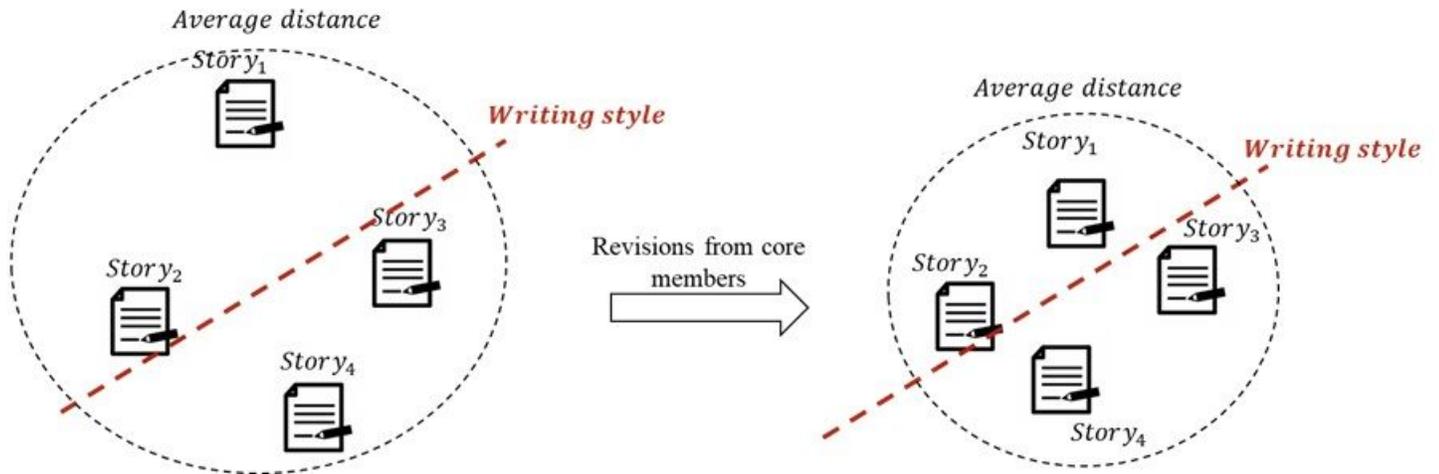


Figure 4

The illustration of the integration role of core members. The decreases in the average cosine distance among the versions of stories can be explained as an integration of the writing style among SCP-stories: by revisions of the core members, different SCP-stories began to share the same writing style; as a result, the average distance between them became smaller.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Sl.docx](#)
- [Sl.docx](#)