Visualizing Collective Idea Generation and Innovation Processes in Social Networks

Yiding Cao, Yingjun Dong, Minjun Kim, Neil G. MacLaren, Sriniwas Pandey, Shelley D. Dionne, Francis J. Yammarino, and Hiroki Sayama
Binghamton University, State University of New York, Binghamton, NY13902, USA
ycao20@binghamton.edu; Sayama@binghamotn.edu

Abstract

Collective idea generation and innovation processes are complex and dynamic, involving a large amount of qualitative narrative information that is difficult to monitor, analyze and visualize using traditional methods. In this study, we developed three new visualization methods for collective idea generation and innovation processes and applied them to data from online collaboration experiments. The first visualization is the *Idea Cloud*, which helps monitor collective idea posting activity and intuitively tracks idea clustering and transition. The second visualization is the *Idea Geography*, which helps understand how the idea space and its utility landscape are structured and how collaboration was performed in that space. The third visualization is the *Idea Network*, which connects idea dynamics with the social structure of the people who generated them, displaying how social influence among neighbors may have affected collaborative activities and where innovative ideas arose and spread in the social network.

Key words: Collective idea generation and innovation, social networks, collaboration, collective performance, idea generation, exploration and exploitation, idea embedding, Idea Cloud, Idea Geography, Idea Network.

1. Introduction

Organizations are increasingly relying on collectives: groups of individuals which share expertise and take collaborative actions [1-4]. In such collectives, multiple people work together interdependently to achieve a greater goal than would be possible for individuals to accomplish alone [5-8]. Sharing of expertise among collective members with diverse backgrounds and behaviors is recognized as an important factor in collective effectiveness [9-12]. Accordingly, there is an increasing need for researchers to better understand the role that collaboration plays in collective design outcomes [13, 14, 15]. However, it is difficult to investigate the collective collaboration process in depth for several reasons [16, 17]. First, a collective is a higher-level entity that can be composed of groups of groups, teams of teams, departments, or even organizations [18]: larger size and more complicated connections make investigation of a collective much harder than regular groups or teams [19]. Second, collective collaboration is generally a large-scale search and design processes typically involving heterogeneous individuals with diverse knowledge, expertise, and behaviors: a large number of variables could conceivably influence design outcomes [20]. Third, such collaborative tasks are usually knotty and open-ended with no simple solutions immediately available to anyone in the collective [19]. Consequently, the ideas produced by individuals in the collective are diverse and changing over time, making the collaboration process difficult to monitor. Furthermore, collective collaboration is a dynamic process that can take place over several days, weeks, months or even years [22-24].

Another difficulty in examining how collaboration impacts collective outcomes is in how to acquire measurements with which to evaluate performance at multiple levels. Earlier human-subject studies typically evaluated performance only at one level using straightforward metrics such as the number of ideas generated

[25], speaking time [26], performance scores [27], and win rates [28]. However, in most real-world scenarios, performance needs to be evaluated using multi-level measurements which examine the performance at both the individual and collective levels [29, 30, 31]. For example, when evaluating group work, people are not only concerned with what the whole group achieves but also with which group members made (or did not make) significant contributions [32].

When individuals of a collective come together to collaborate for problem solving, they often begin with a process of idea diversification, coming up with as many potential solutions as possible [33]. After that, the collective needs to narrow down the ideas to achieve convergence and obtain the best or most innovative solution [34, 35]. This process can be called collective idea generation and innovation. Many existing studies on collective idea generation and innovation attempt to detect idea divergence and convergence as a way of better understanding critical collective collaboration processes [36]; other studies focus on capturing the emergence of innovations [37]. Innovation is increasingly recognized as the leading goal of collective collaborative design, yet it is usually difficult to detect and track [38]. Furthermore, innovation is typically treated as a final achievement in a collective collaboration process, but it actually can occur at any time during the process [39]. Thus, it is essential to develop a way to detect novel ideas or solutions at any time during a collaboration and to identify which individual produced the innovative idea [37]. Finally, all of the divergence, convergence and innovation processes include generating and selecting ideas. One of the most difficult problems in studying collective idea generation and innovation is that the ideas acquired from the records of an experiment are mostly written in a natural language which is difficult for a computer to understand and analyze using traditional quantitative analysis methods [40, 41].

In sum, collective collaboration processes are complex and dynamic, involving a large amount of qualitative narrative information that is difficult to monitor, analyze and visualize using traditional methods. In this study, we developed three new visualization methods for collective idea generation and innovation processes and applied them to online collaboration experiment data. Our goal is to provide new visualization methods along with a high-level view of the issues involved in collective activity and performance visualization, including challenges and key issues to consider.

The rest of this paper is organized as follows. In Section 2 we describe our data source: an online social network experiment. Then, we report the data preprocessing and analysis methods on the data in Section 3. Section 4 describes the three visualization methods we have developed: Idea Cloud, Idea Geography, and Idea Network. The final section discusses contributions of these three new methods to collective idea generation and innovation studies.

2. Data Source

To investigate the mechanisms of collective idea generation and innovation process, we conducted an online social network experiment using an open-ended collaboration design and innovation task (task instructions shown in Table 1). In this task, all the ideas generated by the participants were open-ended free-form narratives. We recruited a multidisciplinary group of students at a mid-size US public university to participate in the experiments. Participants were assigned to collectives with 25-26 members each, connected in an experimentally configured social network. Each collective had diverse individuals with multiple academic backgrounds.

Table 1. Online experiment task instruction

Design Task: *Presented below are two examples of the developmental timeline of products:*
Example 1: Personal communication device *Telegraph → land line → mobile phone → smart phone → ???* *Example 2: Calculator* *Abacus → mechanical calculator → electronic calculator → ???* *In a similar way, you are to predict/design a future state for another product. Please describe* ***which product will undergo what kind of technological transition*** *(include only one transition in your idea).* *This is a group activity, so you should collaborate with others by liking and commenting on other ideas shown below.*

One online experimental session was two weeks long, during which time participants were requested to log in to an experimental platform, a social media-like platform we developed in-house and hosted on our laboratory's web server, using anonymized usernames. Participants were requested to spend at least 15 minutes each weekday working on the assigned collective design task through collaboration with their anonymous social neighbors on the online platform: posting novel or modified ideas to the platform and discussing by reading, liking and commenting on ideas posted by their neighbors. By utilizing their neighbors' ideas and responses/comments received from them, participants were expected to continuously elaborate and improve their idea quality over time. After the two-week session was over, participants received a final reminder email which asked them to complete an end-of-session survey form within one week. This survey asked participants to provide three final ideas they thought were the best among the posted ideas available on the platform. These final ideas were later evaluated by third-party experts who were not involved in the experiments. These evaluation results were used to quantitatively assess the quality of the final ideas developed by each collective.

This experiment aimed to simulate a collective idea generation and innovation process in a situation similar to real-world ones. A collective in our experiment was large and involved people with diverse knowledge, expertise, backgrounds, and behaviors. Like a typical real-world collective collaboration process, the duration of this experiment was relatively long. Additionally, the design task used in this experiment was open-ended and challenging for participants, stimulating them to display exploration and exploitation behaviors and making the collaboration process more complex. Participants' ideas changed over time, and the discussion record contained a large amount of qualitative narrative information which would be difficult to analyze using traditional methods.

The three new visualization methods proposed in this paper are developed to address all the difficulties mentioned above. These methods also allow for evaluating performance of the whole collective as well as that of each individual member.

3. Data Preprocessing

For our three visualization methods, some data preprocessing is needed. As mentioned earlier, the idea data acquired from the experiment were mostly in free-form text format. Therefore, we preprocess the text data using a semantic embedding algorithm to convert them into numerical representations. We call this step idea embedding. There are many semantic embedding algorithms, such as Bag of Words (BOW) [42], Latent Dirichlet Allocation (LDA) [43], word2vec [44], and doc2vec [45]; we used doc2vec in this study. Doc2vec, an adaptation of word2vec, is an unsupervised machine learning algorithm that can generate numerical vectors as a representation of sentences, paragraphs, or documents [45]. Compared to other algorithms, doc2vec can provide a better text representation with a lower prediction error rate because it can recognize the word ordering and semantics which are not addressed by other algorithms [46].

In our experiment, all daily ideas posted on the experimental platform and all final ideas submitted in the end-of-session form were converted to numerical vectors using a single doc2vec model. We set the dimension of output vectors to 400 in order to allow the numerical vectors to represent as much information from the original idea as possible. However, many of the 400 dimensions in the vectors obtained with doc2vec were undoubtedly correlated with each other, making the dataset highly redundant. Therefore, we applied Principal Component Analysis (PCA) to the set of idea vectors for each experimental session to reduce dimensionality [47]. We captured 58.52% and 30.21% of the data variance, respectively, with the first and the second principal components. Thus, the first two principal components, PC1 and PC2, were used to represent the idea vectors for the visualizations proposed in this paper.

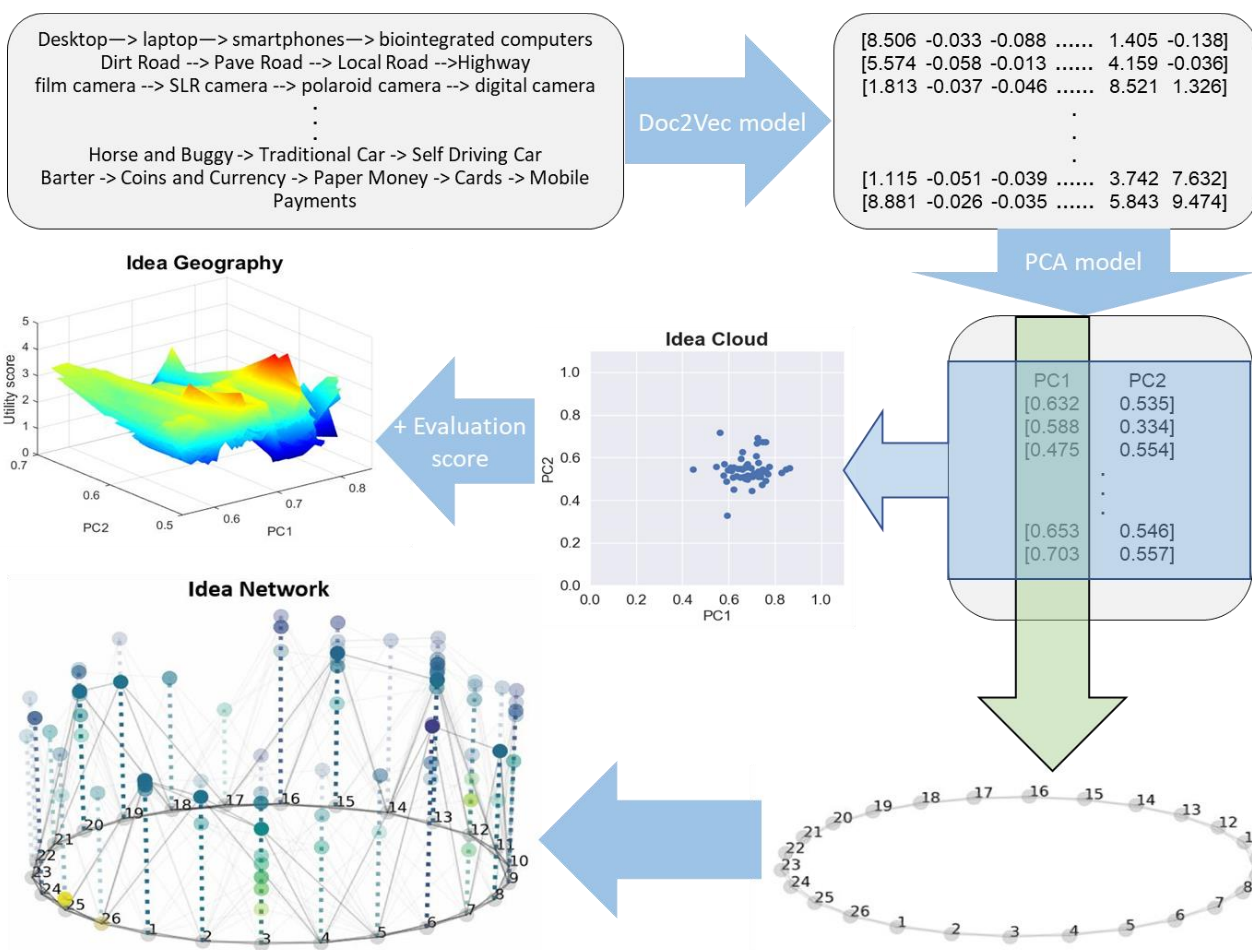

Figure 1 An overview of the workflow of proposed visualization methods. See text for details.

Figure 1 shows the workflow of how the proposed visualization methods were developed. First, the whole collection of daily generated and final designed ideas was used to build the doc2vec model that convert the ideas to 400-dimensional numerical vectors. PCA was performed on the output idea vectors to obtain PC1 and PC2 values, which were used to plot an Idea Cloud. Meanwhile, with the social network structure, the PC1 values were also used to build an Idea Network. The 2-dimensional problem space of an Idea Cloud provided the basis to construct an Idea Geography.

4. Proposed Visualization Methods

4.1 Idea Cloud

The first visualization method, which we call an "Idea Cloud," is a straightforward visualization of idea distributions as a point cloud in a 2D space made of the first two principal components (PC1 and PC2). The 2D principal component space offers an efficient, intuitive way to monitor the locations of ideas and thereby check which ideas are relatively similar to or different from each other.

The Idea Cloud can be used as a basic spatial map of ideas as shown in Figure 2A. Like a weather map showing storm tracks, the Idea Cloud can display the idea track in the idea space in a simple and visual way. For example, we can identify ideas which are located near, or far away from, the center of the idea cluster in the Idea Cloud, which may represent mainstream or unique ideas, respectively.

The Idea Cloud visualization method can also help us monitor how many key topics exist in the collective discussion. In the example shown in Figure 2B, a *k*-mean clustering algorithm was performed to the first two principal components, and the optimal number of clusters (*k*) was determined using the elbow method [48]. The idea points that belong to different clusters in Idea Cloud are visualized using different colors so one can see which ideas belong to which topic.

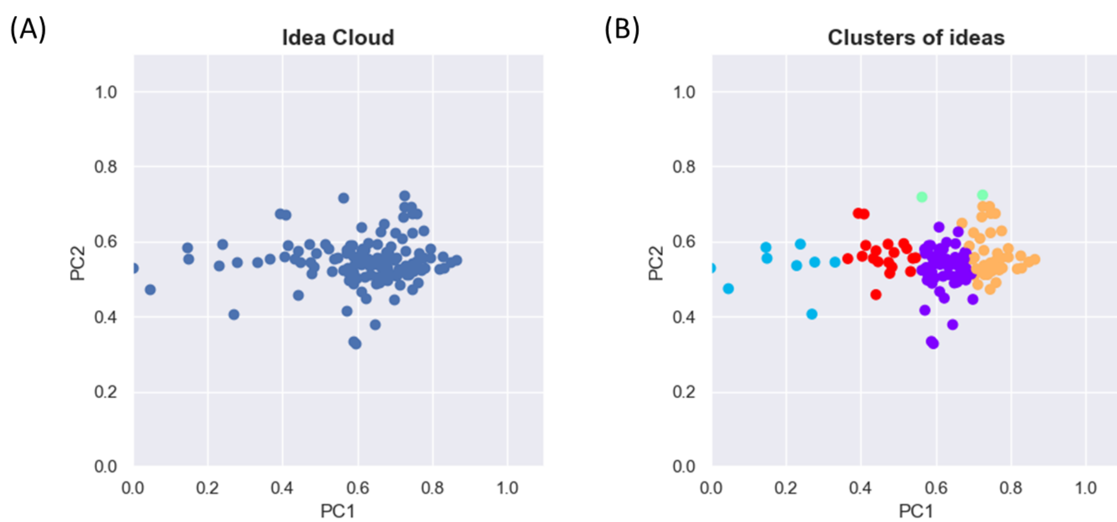

Figure 2 (A) Idea Cloud of whole collection of ideas (B) Clusters of ideas

The Idea Cloud can be used to visually detect unique, potentially innovative ideas. For example, Figure 3 compares the Idea Clouds of the same collective for two consecutive days, in which one can notice a new idea appeared on Day 2 (highlighted in red in Fig. 3B) in the area that was not explored on Day 1. This allows users to further inspect the original idea represented by this red dot to evaluate the innovativeness of this idea.

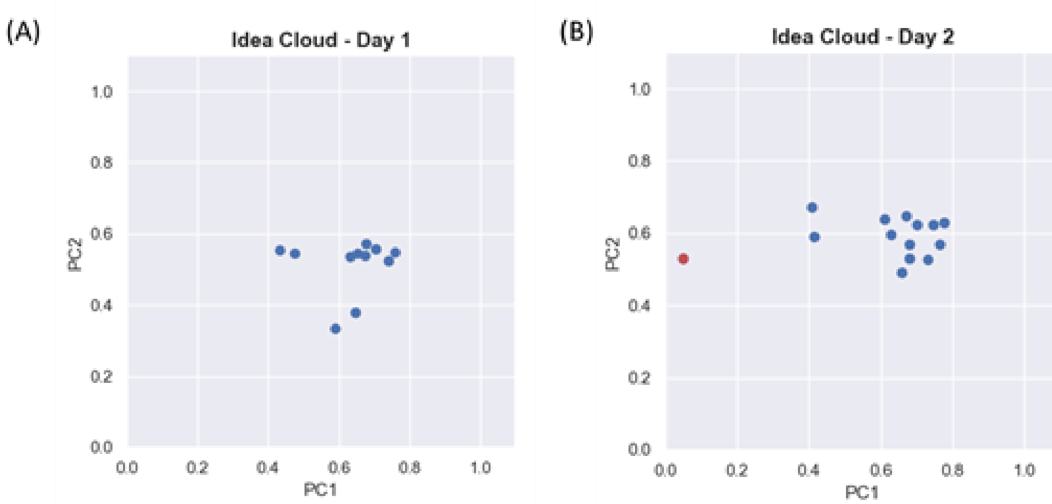

Figure 3 Idea Cloud of daily generated ideas (A) Idea Cloud of Day 1 (B) Idea Cloud of Day 2

Idea Clouds are also useful for comparison of collective performance across multiple collectives. For example, Figure 4 shows Idea Clouds of final designs submitted by two collectives under different experimental conditions. We can see that Collective 1 (Fig. 4A) generated a broader idea distribution than Collective 2 (Fig. 4B), which implies that Collective 1 may have produced more diverse ideas than Collective 2. Such a collective-level observation can help us better understand the relationships between conditions of collectives and their collective performance.

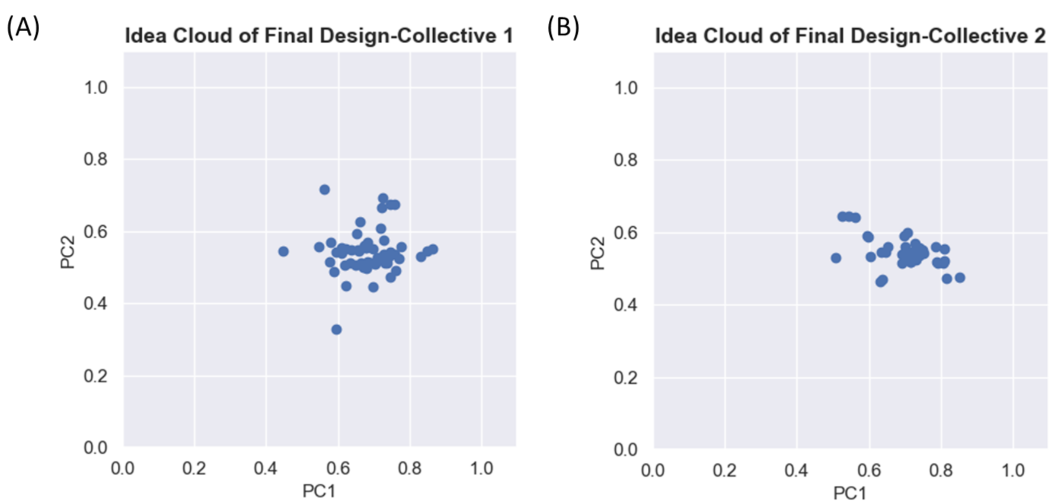

Figure 4 Idea Cloud of final designs (A) Idea Cloud of final designs of Collective 1 (B) Idea Cloud of final designs of Collective 2

Idea Cloud can also be used to track an individual's idea generation process over time, like GPS-based "run tracker" apps that are popular among runners who want to keep track of their running activities. Idea Cloud can track and visualize the sequence of ideas generated by one participant from the beginning to the end of an experimental session, with which one can evaluate the participant's exploration activities. For example, Figure 5 shows such idea tracking results for two individual participants, where we can see that Participant #3 (Fig. 5A) had a longer "running" distance than Participant #13 (Fig. 5B), indicating that Participant #3 was more likely to have had a greater ability to explore new ideas and produce diverse ideas.

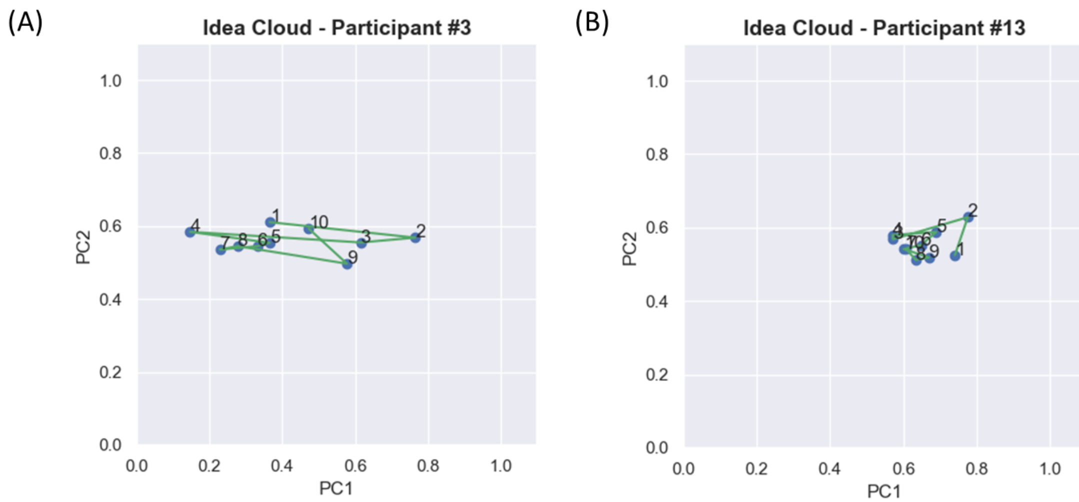

Figure 5 Idea Cloud of individual participants (A) Idea Cloud of Participant #3 (B) Idea Cloud of Participant #13

Moreover, by adding a third dimension for time, Idea Clouds can visualize temporal dynamics of collective idea generation. By combining multiple Idea Cloud plots along a time axis, we can intuitively trace how the idea distribution transitioned over time. Figure 6 shows an example in which the Idea Clouds from Day 1 to Day 10, as well as the Idea Cloud of the submitted final designs, are arranged in a chronological order from left to right in a 3D visualization space made of PC1, PC2 and time. In this example, we can see that the idea distribution in Day 1 had a broader distribution, which means on the first day, participants produced more diverse ideas at the beginning of the session. On Day 2, the idea distribution became a little concentrated, meaning that the participants began to show some idea convergence due to the one-day information sharing and learning. Day 3 and Day 4 showed similar divergence and convergence patterns, and after Day 4, the idea distribution kept concentrated near the center until the end of the session. This kind of temporal visualization of idea distributions helps the experiment designer or group manager to better understand collective dynamics and thus determine various collaboration parameters, such as the appropriate length of a session.

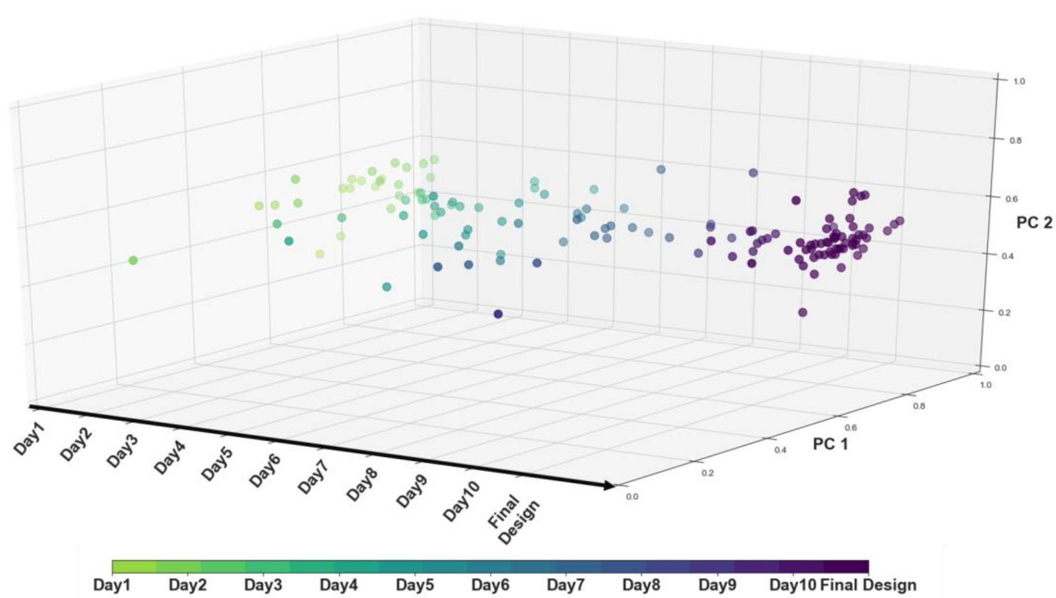

Figure 6 Time-series of Idea Clouds

4.2 Idea Geography

The second visualization method is what we call "Idea Geography." In an Idea Geography plot, the evaluation score of each final design idea is used as its elevation to construct utility terrains on the base of a 2D Idea Cloud. Figure 7 shows an example of Idea Geography, in which one can see mountains and valleys. The mountain areas represent regions populated by ideas with high evaluation scores, and on the contrary, the valley areas are populated by ideas with low evaluation scores.

Idea Geography can be useful to visualize and identify the location(s) of ideas which have the highest quality. For example, in Figure 7, the region with points "a" and "b" are easily identifiable as the "best idea" area. These two ideas described production timelines of headphones and earbuds, which were evaluated as the best final ideas of this collective.

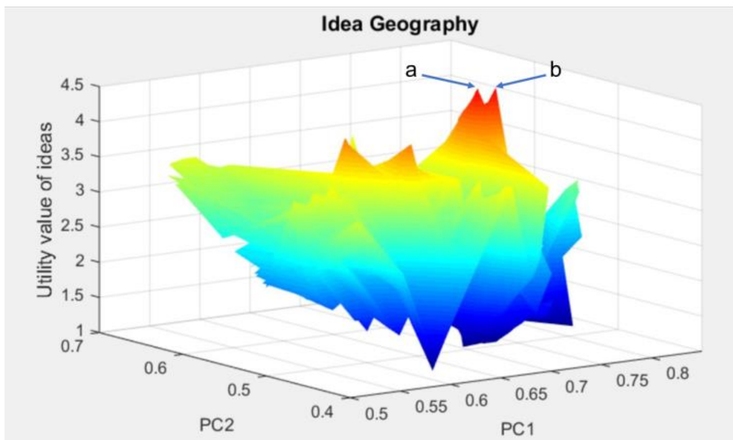

Figure 7 Idea Geography *Note: a and b represent two peaks in the Idea Geography*

Idea Geography can be used to compare the performance between multiple collectives. From the Idea Geography plots shown in Figure 8, we can see that in Collective 2 (Fig. 8B) there is a clearly identifiable utility mountain area (to the left in the plot) where most of the submitted high quality final designs were concentrated. In contrast, Collective 1 (Fig. 8A) has a very small mountain area (to the right in the plot). This observation suggests that Collective 2 may have had greater ability to find the high utility area through the two-week idea exploration and exploitation period than Collective 1.

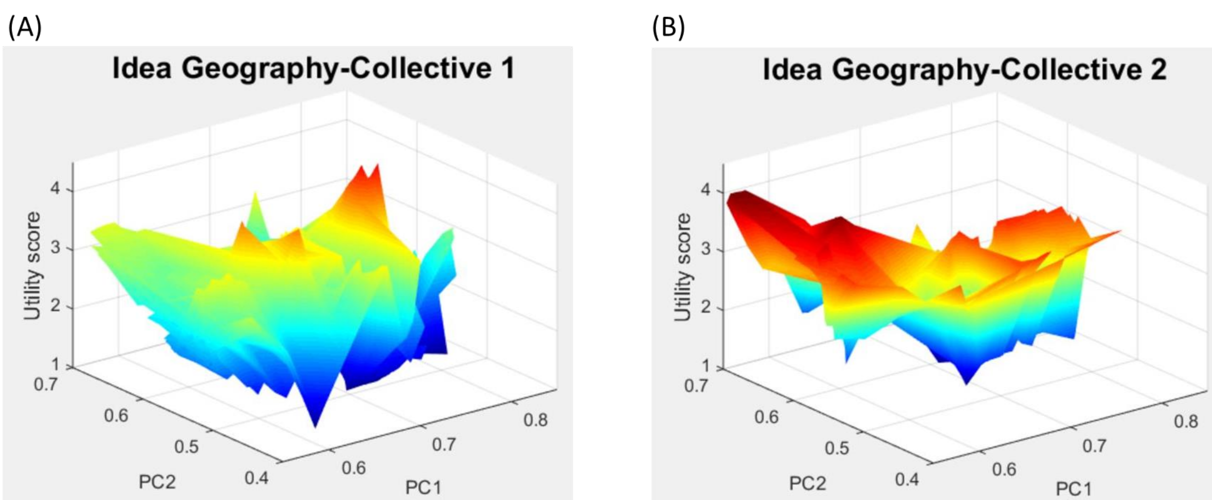

Figure 8 Idea Geography of Collectives (A) Idea Geography of Collective 1 (B) Idea Geography of Collective 2

Like Idea Cloud, Idea Geography can be used to evaluate not only the whole collective's performance but also an individual's performance as well. For example, in our experiment, participants were asked to submit three final designs. The idea points of final designs submitted by each participant can be marked in their collective's Idea Geography space with PC1, PC2, and utility value of these ideas. For example, in Figure 9, we can see that Participant #4 made one final design which had the highest score, but the scores of the other two ideas were low. In comparison, Participant #21 may have had a better idea selection ability because, even though he/she did not make the best idea, Participant #21's idea scores were consistently high. Thus, Idea Geography can help us evaluate an individual participant's overall performance in the idea space.

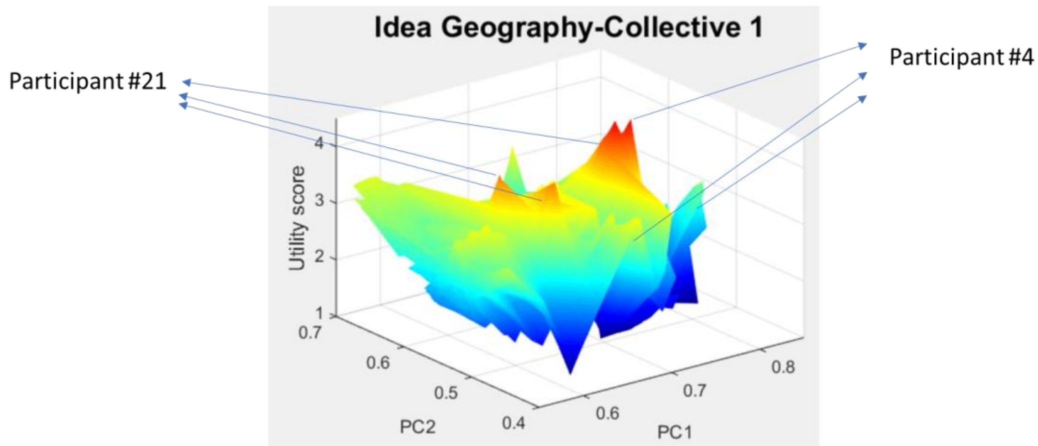

Figure 9 Idea Geography: Individual participants evaluation

The previous examples of Idea Geography used the evaluation scores of ideas for terrain elevation. We can also use elevation to represent other measurements according to what we need to examine. For example, Figure 10 used the length of an idea as the elevation to estimate which ideas described a greater number of steps of future states of technological products. Comparing this Idea Geography with previous one in Figure 7, we can find that the idea utility mountain area locates extremely close to the length of idea mountain area, which suggests that the longer, more elaborated ideas tended to receive higher evaluation scores.

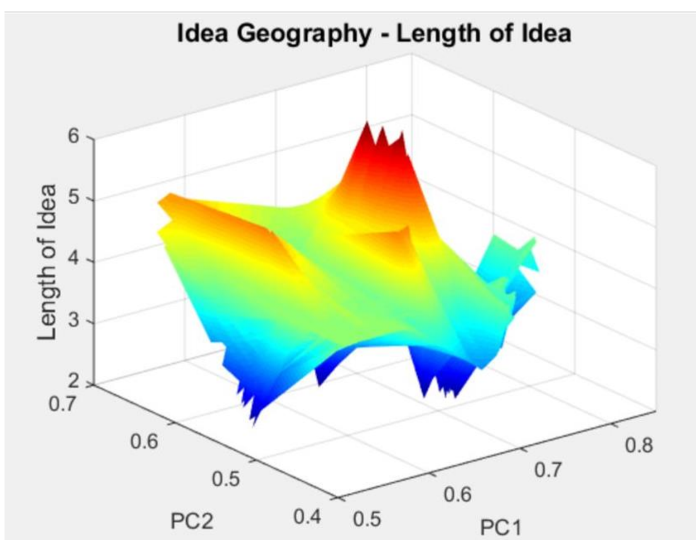

Figure 10 Idea Geography: lengths of ideas visualization

4.3 Idea Network

The third visualization method we propose is the "Idea Network." The Idea Network is a 3D visualization constructed on a network made of human nodes and social connection links among them. In our experiment, the underlying social network was a spatially-clustered regular network with node degree 4. More specifically, the participants were arranged in a shape of a ring, and each of them was connected to four nearest neighbors. Figure 11 shows an example of an Idea Network. If a participant (a human node) submits an idea, an idea node appears above the human node. The height of the placement of the newly generated idea node is determined by the value of its first principal component (PC1). The idea nodes are also connected to other idea nodes in the social neighborhood, following the topology of the social network. The shade of an idea node becomes less saturated as it becomes older to visualize the recency of ideas generated.

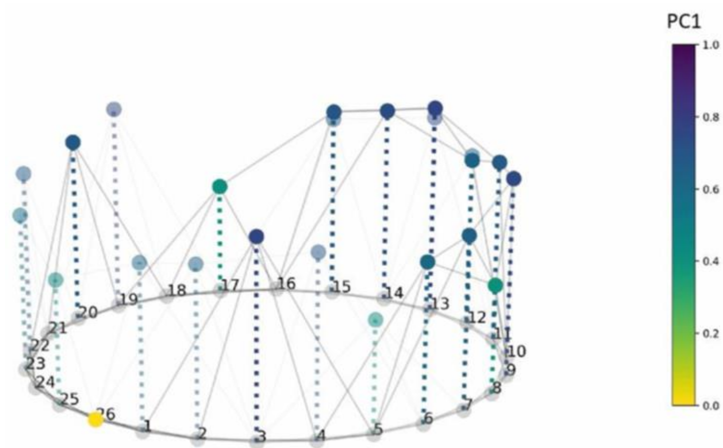

Figure 11 Idea Network

An Idea Network can be animated over time, connecting idea dynamics with the social structure of people who generated them and thus helping to observe how social influence among neighbors may have affected collaborative activities (an animation video can be found from link: http://shorturl.at/lBJ36). Such dynamic animations can offer an intuitive visualization of the idea spreading process in a social network.

Like the Idea Cloud that can detect the emergence of innovative ideas, the Idea Network can further show us where the innovative ideas arise. Figure 12 shows an example. The yellow point which represents an idea submitted by Participant #26 had its PC1 value very different from that of the other ideas, indicating that Participant #26 produced a highly unique, potentially innovative idea. This helps capture the location of the innovation and detect which people are more innovative during collective collaboration.

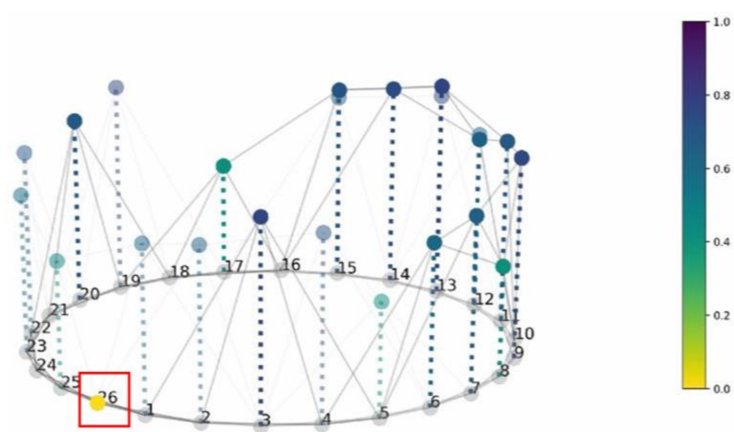

Figure 12 Idea Network: innovative ideas detection

The Idea Network can also evaluate an individual's idea exploration. We previously showed that the "run tracker" function of the Idea Cloud (Figure 5) captured that Participant #3 had a high ability to produce diverse ideas. The Idea Network can provide the same answer in a social context. Figure 13 shows that the ideas submitted by Participant #3 ranged widely in terms of the ideas' PC1 values, in stark contrast to the behaviors of her neighbors. This indicates that Participant #3's diverse ideas were not borrowed from his/her social neighborhood but were truly generated by Participant #3 as unique contributions to the collective collaboration.

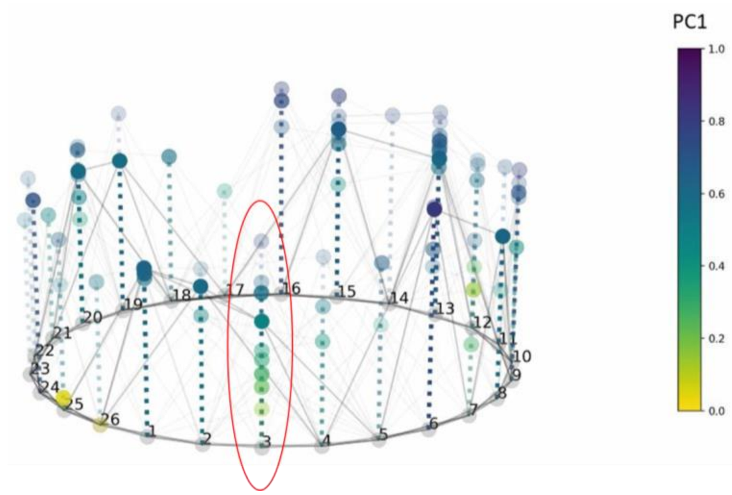

Figure 13 Idea Network: innovative individual detection

The last example of the use of Idea Network is evaluating collective-level efficiency. Figure 14 presents an example. If we simply count the total number of ideas generated, we can see that Collective 2 generated more ideas than Collective 1. However, Figure 14 also shows that, in Collective 2, there were several participants who made little to no contribution to idea generation (marked in red squares): most ideas were produced by a certain subset of people. Meanwhile, in Collective 1, almost every participant made contributions to the design and innovation process. This gives us an indication that Collective 2 might not be an efficient collective and it may be desirable to reorganize it or replace those non-contributing participants.

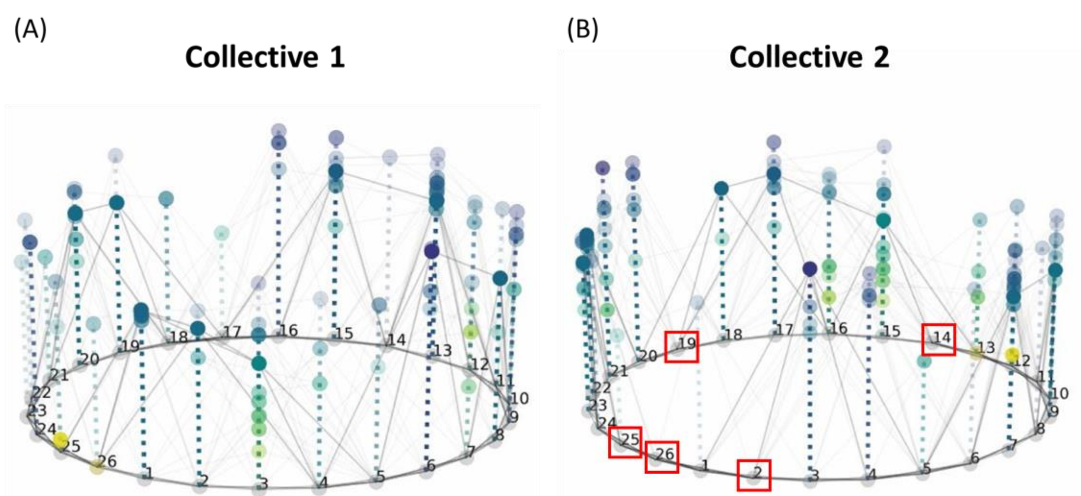

Figure 14 Idea Network of Groups (A) Idea Network of Collective 1 (B) Idea Network of Collective 2

5. Conclusion

In this paper, we proposed three methods for visualizing collective idea generation and innovation processes: Idea Cloud, Idea Geography, and Idea Network. The applications using our online experimental data represent a use case that exemplifies the benefits of these methods. In this use case, these visualizations provide easy, clear and efficient ways to monitor and analyze collective activities and performance in the idea generation and innovation process. We expect these methods to be useful for studying complex collective collaboration processes which involve large-size collectives, open-ended tasks and long durations.

The possible benefits of each of the proposed methods can be summarized as follows. The Idea Cloud can help an organization manager, or a collective member monitor collective activity and performance by tracking idea locations and the dynamic transitions of idea distributions over time during the whole collaboration process. Idea Clouds can also help detect clusters of key topics, detect the appearance of innovative ideas during discussion, and evaluate each individual participant's idea exploration ability. Idea Geography helps understand how the whole idea space is structured with regard to the ideas' utility values (or any other metrics of interest) and how the collective collaboration performed in the idea space; it can show the location(s) of the best ideas and help evaluate collective and individual participants' performance. Idea Networks are useful in visualizing idea dynamics in a social context, showing how social influence affects collaboration activity, and revealing where innovative ideas arise and how they spread in the network. Additionally, Idea Networks can provide information about each individual participant's contributions. These visualization methods offer many possible applications to researchers or human factors engineers in assessing collective performance over time, training interventions, and design features.

The three visualization methods excel at processing larger amount of ideas: in our case, hundreds to thousands of ideas generated by a collective in one experimental session. However, for a smaller number of ideas or small group idea generation analyses, limitations exist due to smaller sample size. To maintain the accuracy of idea vector prediction, the doc2vec algorithm typically needs a considerable amount of data. Additionally, all three visualization methods require enough data points to present a pattern of idea distribution or transition. A visualization method that can work well with both small and large data sets would be an important future development. Another limitation of the proposed visualization methods is only when an experimental session is completed can we apply the methods to acquire the idea generation and innovation conditions for that session. Thus, it is very difficult for conducting periodic assessment during collaboration process, which is essential for incremental and adaptive adjustment for better collective configuration and performance. One worthy future work is to develop advanced Idea Cloud, Idea Geography, and Idea Network which can provide intermediate visualizations in the middle of a session, with which one can periodically check the collective collaboration progress and make appropriate adjustments as needed.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant #1734147.

References

[1] S. Garold and A. Susanne, "Collective Choice, Collaboration, and Communication," *Annual Review of Psychology*, vol.71, p 589-612, Sep. 2019

[2] R.A. Weber, “Managing growth to achieve efficient coordination in large groups,” *American Economic Review*, vol. 96, no. 1, pp. 114-126, Mar. 2006.
[3] S. Faraj, and L. Sproull, "Coordinating Expertise in Software Development Teams," *Management Science*, vol. 46, no. 12, pp 1554-1568, Dec. 2000.
[4] M. Horton, P. Rogers, L. Austin, and M. McCormick, “Exploring the impact of face-to-face collaborative technology on group writing,” *J. Manage. Inf. Syst.*, vol. 8, no. 3, pp. 27–48, Dec.1991.
[5] S. C-Y. Lu, W. Elmaraghy, G. Schuh, and R. Wilhelm, “A scientific foundation of collaborative engineering,” *Annals of CIRP*, vol 56, no 2, p 605–634, 2007.
[6] J. H. Hammond, R. J. Koubek, C. M. Harvey, “Distributed Collaboration for Engineering Design: A Review and Reappraisal,” *Human Factors and Ergonomics in Manufacturing*, vol. 11, no.1, p 35-52, Dec. 2001.
[7] R. Sriram, S. Szykman, D. Durham, “Special Issue of Collaborative Engineering,” *Journal of Computing and Information Science in Engineering*, vol 6, no. 2, p 93–95, Jun. 2006.
[8] S. C.-Y. Lu, N. Jing, “A Socio-technical Negotiation Approach for Collaborative Design in Software Engineering,” *International Journal of Collaborative Engineering*, vol. 1, no. 1, p 185-209, 2009.
[9] R. L. Moreland, J. M. Levine, and M. L. Wingert, “Creating the Ideal Group: Composition Effects at Work,” *Understanding Group Behavior*, p 11–35, Oct. 2018.
[10] D. V. Knippenberg, C. W. De Dreu, and A.C. Homan, “Work group diversity and group performance: An integrative model and research agenda,” *Journal of Applied Psychology*, vol. 89, no. 6, p 1008-1022, 2004.
[11] H. van Dijk, M. L. van Engen, “A status perspective on the consequences of work group diversity,” *Journal of Occupational and Organizational Psychology*, vol. 86, p 223–241. 2013.
[12] K. A. Jehn, G. B. Northcraft, and M. A. Neale, “Why differences make a difference: a field study of diversity, conflict, and performance in workgroups,” *Administrative Science Quarterly*, vol. 44, no. 4, pp. 741–763, Dec. 1999.
[13] N. L. Kerr, R. S. Tindale, “Group performance and decision making,” *Annual Review of Psychology*, vol. 55, no. 1, p 623-655, Feb. 2004.
[14] K. A. McHugh, F. J. Yammarino, S. D. Dionne, A. Serban, H. Sayama, and S. Chatterjee, “Collective decision making, leadership, and collective intelligence: tests with agent-based simulations and a field study,” *The Leadership Quarterly*, vol. 27, no. 2, pp 218-241, April. 2016.
[15] S. D. Dionne, J. Gooty, F. J. Yammarino, and H. Sayama, “Decision making in crisis: a multilevel model of the interplay between cognitions and emotions,” *Organizational Psychology Review*, vol. 8, no. 2-3, pp. 95–124, Oct. 2018.
[16] J. Shore, E. Bernstein, and D. Lazer, “Facts and figuring: An experimental investigation of network structure and performance in information and solution spaces,” *Organization Science*, vol. 26, no. 5, p 1432-1446, Sep-Oct. 2015.
[17] W. Mason, D. J. Watts, “Collaborative learning in networks,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 3, p 764-769, Jan. 2012.
[18] F. Dansereau, F.J. Yammarino, and J. Kohles, “Multiple levels of analysis from a longitudinal perspective: Some implications for theory building,” *Academy of Management Review*, vol. 24, no. 24, p 346-357, Apr. 1999.
[19] K.L. Cullen-Lester, F.J. Yammarino, “Collective and network approaches to leadership,” *The Leadership Quarterly*, vol. 27, no. 2, pp 173-180, April. 2016.
[20] R. Ely, “A field study of group diversity, participation in diversity education programs, and performance,” *Journal of Organizational Behavior*, vol. 25, p 755-780, 2004.
[21] Y. Cao, Y. Dong, M. Kim, N. G. MacLaren, A. Kulkarni, S. Dionne, F. Yammarino, H. Sayama, “Capturing the production of innovative ideas: An online social network experiment and “Idea Geography” visualization,” Proceedings of the 2019 International Conference of The *Computational Social Science Society of the Americas*, p 341-354, Nov. 2019. arXiv:1911.06353.

[22] H. Sayama and S. D. Dionne, “Studying collective human decision making and creativity with evolutionary computation,” *Artificial Life*, vol. 21, no. 3, p 379–393, Aug. 2015.

[23] J. A. Grand, M. T. Braun, G. Kuljanin, S. W. Kozlowski, and G.T. Chao, “The dynamics of team cognition: a process-oriented theory of knowledge emergence in teams,” *Journal of Applied Psychology*, vol. 101, no. 10, p 1353–1385, Oct. 2016.

[24] F. J. Yammarino, M. D. Mumford, M.S. Connelly, and S.D. Dionne, “Leadership and team dynamics for dangerous military contexts,” *Military Psychology*, vol. 22, Suppl. 1, S15-S41, Mar. 2010.

[25] K. Girotra, C. Terwiesch, and K. T. Ulrich, “Idea generation and the quality of the best idea,” *Management Science*, vol. 56, no. 4, p 591–605, Feb. 2010.

[26] N.G.Maclaren et al., “Testing the babble hypothesis: Speaking time predicts leader emergence in small groups,” *The Leadership Quarterly*, vol. 31, no. 5, p 101409, Feb. 2020.

[27] P. Kanawattanachai, Y. Yoo. “The impact of coordination on virtual team performance over time,” *MIS Quart*, *vol.* 31, no. 4, p 783–808, Dec. 2007.

[28] A. Sapienza, Y. Zeng, A. Bessi, K. Lerman, and E. Ferrara, “Individual performance in team-based online games,” *Royal Society Open Science*, vol. 5, no. 6, p 1-14, May. 2018.

[29] S.D. Dionne, P.J. Dionne, “Levels-based leadership and hierarchical group decision optimization: A Monte Carlo Simulation,” *The Leadership Quarterly*, vol. 19, no. 2, p 212-234, April. 2018.

[30] J. R. Hollenbeck, D. R. Ilgen, D. J. Sego, J. Hedlund, D. A. Major, and J. Phillips, “Multi-level theory of team decision making: Decision performance in teams incorporating distributed expertise,” *Journal of Applied Psychology*, vol. 80, no. 2, p 292-316, 1995.

[31] A. Somech, “Relationships of participative leadership with relational demography variables: A multi-level perspective,” *Journal of Organizational Behavior*, vol. 24, no. 8, p 1003−1018, Dec. 2003.

[32] J.H. Hayes, T.C. Lethbridge, and D. Port, “Evaluating individual contribution toward group software engineering projects”, Presented at ICSE’03. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1201246

[33] S. M. Stein, T. L. Harper, “Creativity and innovation: Divergence and convergence in pragmatic dialogical planning,” *Journal of Planning Education and Research*, vol. 32, no. 1, p 5–17, Sep. 2011.

[34] I. Seeber, “How do facilitation interventions foster learning? The role of evaluation and coordination as causal mediators in idea convergence,” *Computers in Human Behavior*, vol. 94, p 176–189, May. 2019

[35] X.S. Cheng, S.X. Fu, T. de Vreede, G.J. De Vreede, I. Seeber, R. Maier, and B. Weber, “Idea convergence quality in open innovation crowdsourcing: a cognitive load perspective,” *Journal of Management Information Systems*, vol. 37, no. 2, pp. 349-376, Jun. 2020.

[36] I. Seeber, G. de Vreede, R. Maier, and B. Weber, “Beyond brainstorming: exploring convergence in teams,” *Journal of Management Information Systems*, vol. 34, no. 4, p 939-969, Jan. 2018.

[37] E.F. Rietzschel, B.A. Nijstad, and W. Stroebe, “The selection of creative ideas after individual idea generation: Choosing between creativity and impact,” *British Journal of Psychology*, vol. 101, no. 1, p 47–68, Feb. 2010.

[38] S.H. Cady, J. Valentine, “Team innovation and perceptions of consideration: What difference does diversity make?” *Small Group Research*, vol. 30, no. 6, p 730–750, Dec. 1999.

[39] J. Wang, GH-L. Cheng, T. Chen, K. Leung, “Team creativity/innovation in culturally diverse teams: A meta-analysis,” *Journal of Organizational Behavior*, vol. 40, no. 6, p 693-708, Feb. 2019.

[40] H. Chen, et al., “Automatic concept classification of text from electronic meetings*,” Communications of the ACM*, vol. 37, no. 10, p 56–73, Oct. 1994.

[41] B. Ramesh, A. Tiwana, “Supporting collaborative process knowledge management in new product development teams,” *Decision Support Systems*, vol. 27, no. 1-2, p 213–235, Nov. 1999.

[42] Z. Harris, “Distributional Structure,” *WORD*, vol. 10, no.2-3, p 146-162, 1954.

[43] D. Blei, A. Ng, and M. Jordan, “Latent Dirichlet allocation,” *Journal of Machine Learning Research*, vol. 3, p 993-1022, 2003

[44] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of phrases and their compositionality," Presented at (NIPS 2013). [Online]. Available: https://arxiv.org/abs/1310.4546
[45] Q. Le, T. Mikolov, "Distributed representations of sentences and documents," Presented at (ICML 2014). [Online]. Available:
https://arxiv.org/abs/1405.4053
[46] J. H. Lau, T. Baldwin, "An empirical evaluation of doc2vec with practical insights into document embedding generation", Presented at 1st Workshop on Representation Learning for NLP. [Online]. Available:
https://arxiv.org/abs/1607.05368
[47] IT. Jolliffe, J. Cadima, "Principal component analysis: a review and recent developments," *Phil. Trans. R. Soc*, vol. 374, no. 2065, p 20150202, April. 2016.
[48] M.A. Syakur1, B.K. Khotimah1, E.M.S. Rochman1 and B.D. Satoto, "Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster," *IOP Conf. Series: Materials Science and Engineering*, vol 336, p 012017, Nov. 2017