IIAI Open Conference Publication Series
IIAI Letters on Informatics and Interdisciplinary Research
Vol.004, LIIR170
DOI: https://doi.org/10.52731/liir.v004.170

BaleeGraph: Visualizing Co-Creation for Social Good

Kaira Sekiguchi *, Yukio Ohsawa *
Makoto Iida *, Hisashi Nakamura *

Abstract

Addressing social challenges such as the Sustainable Development Goals (SDGs) necessitates the realization of co-creation, where diverse stakeholders have shared interests and complement each other. However, understanding the state of co-creation is challenging, which can render it less effective and manageable. To address this issue, we introduce a novel visualization method called Graph of BALloon tree with Embedding-based Edges (BaleeGraph) for recognizing the co-creation state. BaleeGraph operates on the assumption that each team possesses keywords and employs embedding vectors to calculate keyword similarity among different teams, thereby illustrating their relationships. We validate BaleeGraph's effectiveness through its application to a case study of Climate change actions with Co-creation powered by Regional weather information and E-technology (ClimCORE) project. BaleeGraph enables participants to identify their strengths and shared interests with other teams. Furthermore, it uncovers paths from each team's keywords to those of the SDGs and analyzes time-series changes in the network structure. We conclude by discussing how BaleeGraph facilitates understanding co-creation for social good.

Keywords: Co-Creation, Visualization, Scenario Recommendation, BaleeGraph.

1 Introduction

Despite ongoing efforts, society continues grappling with challenges such as poverty and inequality; environmental issues pose a threat to future generations. The Sustainable Development Goals (SDGs) [1] represent one systematic approach towards addressing these issues. Solving these problems requires implementing projects that engage a wide array of stakeholders in an action known as co-creation. In this context, "stakeholder" refers to any individual or organization with an interest in or concern about a project–researchers, companies, governments or citizens–and they are referred to as "teams" when they participate in a project. A team may consist of one or more people. Co-creation [2], defined here as an action where multiple teams collaborate to produce outputs unattainable through individual efforts alone, is crucial for fostering interactive complementary relationships among teams by aligning perspectives and leveraging each other's strengths. The concept of co-creation

* The University of Tokyo, Tokyo, Japan

awareness is instrumental in executing practical actions and managing projects, acting as a form of metacognition [3]. This metacognitive approach allows teams to identify areas that require focused attention for efficient resource allocation. Such areas could include recognizing the roles of existing teams, hiring new members, or acquiring new assets.

We propose a network-style visualization tool called Graph of BALloon tree with Embedding-based Edges (BaleeGraph) to facilitate understanding how co-creation occurs in collaborative projects. Compared to existing networks, BaleeGraph offers greater flexibility and readability. BaleeGraph excels in various applications, one of which is the demonstration of the 'connecting' among teams. This connecting is dictated by the possession of similar keywords, where the subtle connections are based on keyword similarity. Another noteworthy application is the visualization of macro-level changes by delineating the distinctions within the network structure. Here, "keywords" are defined as words, phrases, clauses or short sentences but not long sentences or documents. The fundamental principle behind BaleeGraph is setting the granularity of nodes as keywords and computationally connecting these nodes based on similarity derived from embedding vectors, thereby enabling a data-driven approach. A "data-driven approach" implies that we can essentially construct a graph by automatically processing data–particularly big data–according to a specific algo-rithm.

To validate the utility of BaleeGraph, we conducted a case study involving the Climate change actions with CO-creation powered by Regional weather information and E-technology (ClimCORE) project [4][5]. The ClimCORE project is a collaboration of universities, companies, and local governments aimed at weather-related innovation. The core technology of ClimCORE is weather reanalysis data, i.e. weather data obtained by reanalyzing (simulating) using the latest weather models. The important mission of ClimCORE is to return value to society based on co-creation inside and outside the project, such as achieving SDGs. For instance, the term "weather" may appear in various keywords such as "weather reanalysis data" and "extreme weather understandings." Then, issue teams working on weather reanalysis data focus on data science from an engineering perspective while teams contemplating applications organize their ideas and visions for said applications like interpreting extreme weather events.

Applying BaleeGraph to the ClimCORE case clearly demonstrates that while each issue team shares a shared interest, e.g., weather, they each operate from their unique strengths, e.g., engineering and requirements organization. Moreover, we amalgamate BaleeGraphs from various topics to connect them. We identified scenario paths linking the key concept of ClimCORE with SDGs targets. We then visualized how each team will continue addressing their shared interest in "weather," along with how new shared interests not previously considered as keywords emerge.

The structure of this paper is as follows. Section 2 discusses related work. Details of BaleeGraph are introduced in Section 3. The settings of the ClimCORE case are described in Section 4. Representative results of the visualizations using BaleeGraph are presented in Section 5. Achievements and limitations of the current logic and implementation of BaleeGraph are discussed in Section 6. Finally, we conclude that BaleeGraph is useful for co-creation metacognition in Section 7.

2 Related Work

To our knowledge, no existing method possesses the three key characteristics that define BaleeGraph: setting nodes at a granularity larger than words, such as keywords; drawing edges between nodes based on the similarity of embedding vectors; and fitting for a data-driven approach which enables automatic visualization of large data sets. This study shows how these characteristics are effective in presenting the co-creation state in a collaborative project.

Typical visualization methods used for similar applications include KeyGraph. KeyGraph has significantly contributed to constructing the visualization field and remains effective today across various data analyses, including practicing Innovators Marketplace on Data Jackets (IMDJ) by connecting diverse metadata sets [6][7]. KeyGraph shows connections between entities by calculating infrequent, i.e., rare event nodes that bridge them. KeyGraph is also used to show network dynamics to explain changes by merging two graphs at different times as Kamishibai KeyGraph [8]. However, KeyGraph uses a word as a node and calculates co-occurrence between them to draw edges.

Then, Sebastián Lozano et al. reviewed various methods developed for edge drawing, including co-occurrence, multidimensional scaling (MDS), and clustering [9], but none utilize embedding vectors and their similarities to draw edges. Recently, Kaira Sekiguchi and Koichi Hori proposed a website for Design From the Ethics level (Dfrome), where node granularity corresponds to sentences and embedding vectors calculate similarities for scenario path recommendation [10]. Despite this innovation, Dfrome has not been successful in automatically creating graphs with large amounts of data.

We employ BaleeGraph to bridge nodes by setting a start node and an end node. Makoto Sato et al. proposed topic bridging for generating stories based on co-occurrence calculation [11]. BaleeGraph is also useful for finding such topic bridging based on similarities of embedding vectors. Various visualization studies have addressed time-series dynamics. Florian Windhager et al. explored four types of dynamic representation: animation, layer comparison, layer merging, and 2.5D view [12]. Particularly regarding graph merging–which this study adopted–some studies have used it to visualize dynamics [13][14]. However, these studies do not rely on embedding vectors to verify whether certain nodes or edges have appeared.

We demonstrate the effectiveness of BaleeGraph in visualizing the co-creation state by organizing three points and applying it to two advanced usages.

3 Details of BaleeGraph

BaleeGraph is a tool that visualizes the relationships among project teams as a network. It accomplishes this by using embedding vectors to calculate similarities between keywords, such as technologies and ideas being developed.

3.1 Visualization of Issue Teams and Keywords

In BaleeGraph, nodes are described with keyword granularity, examples include "weather reanalysis data" and "extreme weather understandings." An ownership relationship exists between each issue team and its respective keywords. Initially, we depicted this relationship as a network but decided to omit the relationship between keywords owned by the same team for simplicity's sake. Consequently, the visualization results in a graph with a balloon tree structure, as shown on the left side of Figure 1.

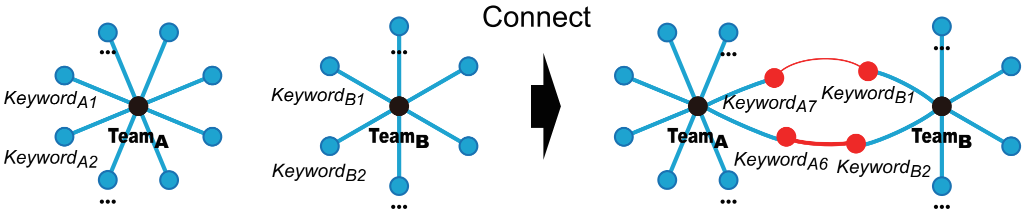

Figure 1: The concept of calculating BaleeGraph.

BaleeGraph employs nodes as keywords, thereby streamlining the interpretation process of visualization. This signifies a departure from the strategy of assigning single or compound words, each consisting of roughly three constituent words, as nodes and forming edges through the computation of their co-occurrences, a methodology that continues to prove useful in analyzing the era of COVID-19 [15][16][17]. In such scenarios, an analyst is required to aggregate multiple nodes to construct a meaningful group. For instance, if team A possesses nodes such as "organize," "research," "issues," "relate," "to," "climate," and "change," it can be inferred that the probable keyword is "organize research issues related to climate change." Given that node granularity aligns with keywords which can be a word, a phrase, a clause, or a short sentence, among others, BaleeGraph enables users to directly comprehend the relationship without necessitating the step of word combination.

3.2 Visualization of Relationships Between Keywords

To illustrate the links between issue teams, we have drawn lines (edges) representing connections based on node similarities, as depicted on the right side of Figure 1. However, a challenge arises when participants in co-creation use identical words in varying contexts. For example, "weather" could be part of the keyword "weather reanalysis data" or "extreme weather understandings." It becomes crucial to delineate the relationship between these similar yet distinct keywords. Unfortunately, this cannot be accomplished through a simplistic approach of applying co-occurrence calculations based solely on keyword matches because keywords with similar meanings but containing different words are not judged as matches. Instead, we begin with calculating the similarities between keywords and affixed edges, utilizing these similarities as our basis. To do this, we employ an embeddings technique that can transform any sentence length into a specific vector length. If a pair of nodes exhibits an index (= constant minus cosine similarity) that exceeds a certain threshold, the edge is drawn with a stroke proportional to the index. This is illustrated by the red lines in Figure 1. The thresholds were empirically determined by the analyst after testing several patterns.

Next, we propose to discover scenario paths within BaleeGraph. First, we specify a start node and an end node. Here, the start node corresponds to the means and the end node corresponds to the purpose. Then, the intermediate nodes play the role of advancing the scenario, but the specific role they play depends on the context. The crucial aspect here is ensuring that each scenario connects such that a team appears between two keywords. This approach allows us to define each team's role concerning their own keywords and those of adjacent teams. For instance, while research and development teams are tasked with realizing keywords, those responsible for concept organization (e.g., backcasting workshop, SDGs) are understood as being responsible for linking these keywords as well as connecting to the next teams.

For instance, KeyGraph demonstrated that infrequently appearing, rare nodes exhibited bridging, which in turn promoted innovation. Moreover, analyzing changes in these nodes was found to be effective in elucidating shifts in the overall structure. In BaleeGraph, the red strings and the intermediate nodes in the scenario path depicted in Figure 1 can perform an analogous function to these red nodes. Here, the variation in thickness, which corresponds to similarity, can indicate the frequent or infrequent connection.

3.3 Visualization of Time-series Changes in Network Structure

We visualized newly emerged network elements (nodes and edges) by comparing two BaleeGraph at different times. This comparison enabled us to observe whether new concepts and connections were addressed over time. Specifically, two BaleeGraphs are merged; network elements exclusive to new meetings are colored while existing elements are grayed out as depicted in Figure 2. For ease of understanding, elements already addressed are grayed out prior to merging (see upper side of Figure 2). Furthermore, embedding vector similarities were used to ascertain whether an element had previously appeared; this is represented by "$Keyword_{a2'} \sim Keyword_{a2}$" in Figure 2. Although it is probable for multiple new keywords and multiple old keywords to form a group within a certain degree of similarity, the current BaleeGraph simplifies the calculation so that there can be multiple old keywords for each new keyword within the condition. Note that edge coloring conditions can be found on the lower left part of Figure 2: "New" indicates newly appeared nodes, "Old" indicates already treated node, and "Common" indicates frequently treated node, but they are not explained here due to space limitations.

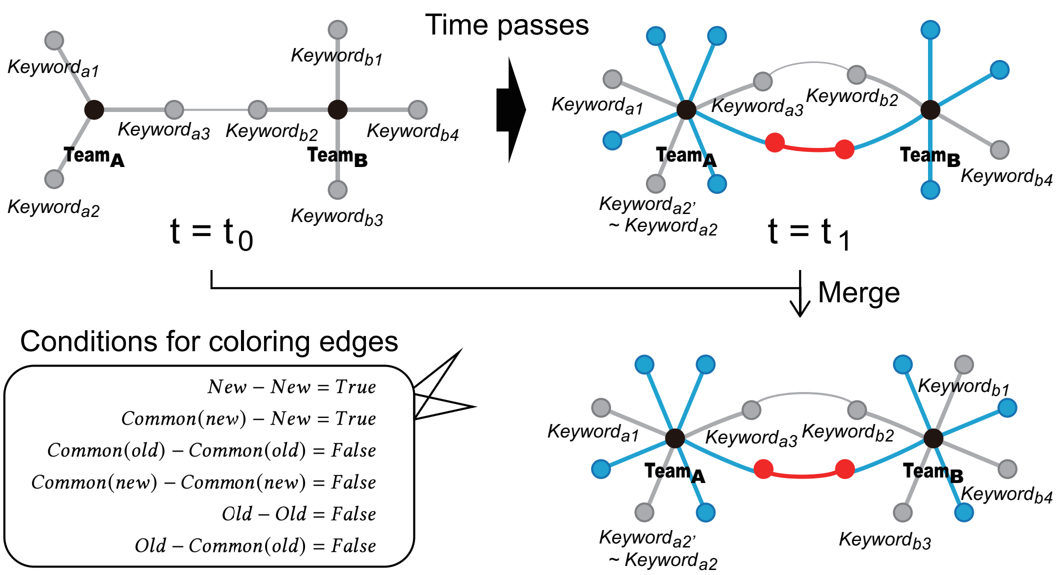

Figure 2: The concept of merging two BaleeGraphs to show time-series changes.

As an example: if keywords like "utilization of local weather data in the era of climate change" have been discussed at the former meeting, then it's likely that "climate change" or "utilization of weather data" would be judged similarly because these keywords share some degree of similarity. Thresholds for such judgments were established heuristically.

3.4 Equations

For k_a and k_b, that are the sentence vectors of keyword_a and keyword_b, the similarity is calculated as follows. Here, the $MAX\_SIMILARITY$ is a constant to ensure that the similarity value is always positive[1] . The T_u and T_l signify the upper and the lower thresholds respectively.

$$
\begin{aligned}
S(k_a, k_b) &= MAX\_SIMILARITY - \left(1 - \frac{k_a \cdot k_b}{\|k_a\|\|k_b\|}\right) \\
similarity(k_a, k_b) &= \begin{cases} S(k_a, k_b) & \text{if } T_l \leq S(k_a, k_b) \leq T_u \\ None & \text{else} \end{cases} \\
\text{where:} & \\
MAX\_SIMILARITY &= 3.00 \\
T_u &= MAX\_SIMILARITY = 3.00 \\
T_l &= 2.75
\end{aligned}
$$

The same concept and constant values are used to account for changes in the time series. Then, we rescale the similarity value for edge weighting in case the number is not None.

$$
\begin{aligned}
rescaled\_similarity(similarity) &= \frac{similarity - MIN_{old}}{MAX_{old} - MIN_{old}} \times (MAX_{new} - MIN_{new}) + MIN_{new} \\
\text{where:} & \\
MAX_{old} &= T_u = 3.00 \\
MIN_{old} &= T_l = 2.75 \\
MAX_{new} &= 3.00 \\
MIN_{new} &= 0.10
\end{aligned}
$$

This *rescaled¯similarity* is used for the edge weight between keywords. Then, the ownership of keywords by a team is supposed to have maximum similarity (=MAX_SIMILARITY); therefore, its weight was the same as that of MAX_{old} and MAX_{new} at present settings.

3.5 Adopted Technologies for Automation

In the implementation of BaleeGraph, we adopted a data-driven approach to minimize the potential for arbitrary conclusions. We aimed to eliminate as much as possible any arbi-trariness, such as manual manipulation. The current BaleeGraph implementation serves as a proof of concept, and the technologies discussed herein were selected heuristically.

Python was the programming language used in developing BaleeGraph. We utilized existing technology to extract keywords from each issue team document. Specifically, we applied the OpenAI API and selected GPT-3.5 (text-davinci-003) [18] as its model. In addition to defining keywords as described above, we set prompts in the instructions to extract keywords from input text that are valuable for societal issues. We specified that

[1]For consistency, significant figures have been standardized throughout this paper tough the actual settings in the program were simplified.

between three and ten keywords should be extracted relative to the length of the input text. If an input text's length exceeded the number of tokens that could be sent at once by the API, we divided it into smaller parts and retrieved keywords in multiple stages.

To calculate distances between keywords, we used pairs of embedding vectors. For this study, sentence-transformers [19] were employed with a multilingual model named "stsb-xlm-r-multilingual" [20]. This model can vectorize sentences of any length. It enables us to compute semantic similarity between nodes by calculating cosine similarity. Our similarity value subtracts the cosine similarity value from some constant, which is greater the closer the distance between vectors.

For visualization purposes, we used libraries called NetworkX [21][22] and Pyvis [23]. We respectively calculated "spring_layout" of NetworkX and "physics" Pyvis (which uses vis.js) simulation for positioning nodes, and the visualized results in this paper are based on Pyvis's one: nodes connected by strong edges are brought closer together while unrelated nodes are kept apart. Color settings and edge weighting were designed heuristically in this study.

4 Case Settings

This paper aims to visualize both current state and growth within co-creation efforts in the ClimCORE project which involves universities, national research institutes, companies, local governments and administrative agencies among others. Within the ClimCORE project eight teams participate in various activities ranging from development of weather-related data to application of such data towards solving social issues. Project teams are considered at granular level based on teams established for each issue. The participating teams in ClimCORE include issue teams 1A through 1D along with teams 2 through 4 plus a team conducting backcasting workshops.

ClimCORE's focus on weather clearly aligns it with SDGs target 13 "climate change" and target 7 "energy." However, it also has potential connections with other targets that are not directly related to weather, such as SDGs target 14 "sustainable management of marine resources." The Japanese version distributed by Japan's Ministry of Internal Affairs and Communications was used for SDGs reference [1]. The data for this study were sourced from materials for "general meetings" (pdf files converted from slides for the meeting). These documents were chosen due to their comprehensive coverage of all issue teams' progress, allowing us to track temporal changes. The ClimCORE project leader granted permission to use these texts, which were downloaded directly from the project's internal administrative cloud folder.

The first general meeting was held on October 1 and 7, 2021, and served as the starting point of our analysis, where most of the issues began to take shape. Then, the general meetings were held in May and November 2022 and January 2023 as the first to the third in Fiscal Year 2022 (FY2022). For this analysis, in order to realize the temporal comparison as much as possible, we decided to compare the keywords in the "third" meeting during FY2022 with the keywords that were present when each issue team first appeared as a document. In other words, we supposed a virtual BaleeGraph as the older version of the comparison, which is a mix of the general meetings before the third of FY 2022. For example, the issue team conducting backcasting workshops first appeared in the May 2022 meeting, so the older balloon of this team was about that meeting, although the Issue Team 1A balloon is from the first meeting in October 2021. Then we compared it to the target

BaleeGraph for the third meeting in FY2022. The visualization calculations were done in February 2023.

5 Results of ClimCORE Case

This section presents the results obtained using BaleeGraph. The project leader and one of the vice leaders of ClimCORE have verified these results.

5.1 Paths to the Targets of SDGs

Figure 3 illustrates how we can trace a path from a concept within ClimCORE to a particular SDG target. This graph was generated running the BaleeGraph program, and the path was deliberately chosen by the first author of this article. It exemplifies how viewing scenarios as paths provide insights into co-creation considerations. Firstly, Figure 3 shows that each issue team possesses unique strengths and shared interests, visually demonstrating a well-balanced structure among the teams in the ClimCORE project. Primarily, the lower left group of balloons corresponds with those of ClimCORE project's issue teams. Meanwhile, the upper right group represents SDG targets. Here, the unique strengths are visualized by the blue nodes; the shared interests are visualized by the red nodes. It is also important to note that SDG targets have intricate relationships with one another; they may possess hierarchical structures or share related themes. For example, it is said that the SDGs have three levels of hierarchy like a wedding cake: biosphere, society, and economy [24]; one goal of the SDGs can be a means to realize another goal, i.e., a hierarchical relationship based on causality. The network visualized in the upper right part highlights what is unique and common among the individual SDG targets.

Then, Figure 3 shows a bridging path: an example is between ClimCORE's key technology–weather reanalysis–as a starting node and sustainability related to coastal and marine aspects of SDG Target 14 as an end node. In this context, we identified unexpected scenarios and opted for an indirect path involving collaboration by four issue teams and mediation by two SDG targets (8 and 12). As detailed in Figure 3's upper left corner, this path begins with using weather reanalysis for addressing climate change; it then leads towards co-creation society, systems data use, technology development and innovation; progresses towards improving resource efficiency; advances towards sustainability implementation programs; culminating in conservation of marine resources.

In short, from the start node of Figure 3 representing the original key concepts in the project, we find a path to reach the end node representing the solutions to problems addressed in SDGs. This path goes via (a) climate change and adaptation to it which had been shared from the beginning of the project, (b) intelligence society including data plat-forms and human resources, and (c) innovations for sustainable management of resources and human activities such as sight seeing. The topics (b) and (c) are indirectly related to the ClimCORE project, which specifically focused on the development of advanced weather data with spatiotemporally high resolution, but show unexpected direction to social contribution.

5.2 Generation and Growth of Co-creation

Figure 4 illustrates the comparison between the initial presentation by each issue team and their presentation at the 3rd general meeting in the Fiscal Year 2022. This graph was also automatically made, and the focused points were manually selected by the first author of this paper. These points were selected to illustrate that by examining alterations in network structure changes, we can derive insights into the considerations relevant to co-creation. Elements introduced in these presentations are color-coded; new elements are highlighted in blue and red, while old elements are grayed out. This visualization allows us to confirm that issue teams have intensified their efforts related to weather and climate. It suggests that each interdisciplinary team has successfully integrated weather and climate-related topics into their work. Thus, we can infer that perspectives related to weather and climate data have become a central part of the ClimCORE project, as depicted on the left side of Figure 4.

In Figure 4, we find the following novelties in the new structure in comparison to the old one. In frame (a): "high-resolution weather data," "using weather data," and "scenarios of regional climate change" are unified to "weather data," whereas "effective use of weather data" is replaced by "localized weather data" and "reanalysis (a synonym of data assimilation) of weather data." In addition, "SDGs" and words relevant to reactive strategies to "climate change" are highlighted.

In frame (b): the significance of "disaster prevention" and "disaster mitigation" for "local regions" become highlighted in progress of the project. "Flood" disaster is a typical disaster due to torrential rainfall events that tend to become more likely under the global warming even in summertime Japan [25]. For example, a torrential rainfall event that occurred in early July 2020 caused disastrous flooding in the Kumamoto prefecture [26], which is collaborating with the ClimCORE project.

Disaster prevention has been one of the core research themes within ClimCORE, and the above results show the progress of this theme. Issue teams 1A, 1C, and 2 have demonstrated a diversified approach by addressing this theme, as shown on the right side of Figure 4.

6 Discussion

At first, it was observed that each group was able to work with their own unique strengths (expertise). The fact that blue clusters (keywords) were released around the black node (team) and that each issue was connected to another group partially and in a different direction (not solidly in the same direction), means that the project is progressing with the uniqueness of each individual issue ensured. And the part where the red nodes are newly connected to multiple issues (keywords such as Kumamoto and disaster) suggests the possibility of forming a unique group, and (although not the result of this analysis) shows the legitimacy of the fact that the issue setting for Kumamoto and disaster, issue team H, was drafted and the project is being managed. This is a good indication of the validity of the project's management.

It was also clarified how they are related to the common field of weather. It is inevitable that each issue is connected to weather because the ClimCORE project is for the field. However, it appears that there are still issues to be addressed regarding the analysis of the connections. This is because they are connected to other issues in terms of how weather relates to the world, and the new red nodes are also connected by the keyword about social implementation. We consider this is because weather is not just one physical phenomenon that has a strong social impact (in wind power, it is easy to understand because wind is the main factor), but it is a macro and complex phenomenon (disasters such as typhoons are multiple weather events as weather phenomena alone like rain and wind) that affects us in various ways. What we can see from this is that weather reanalysis will continue to be verified as physical phenomena (pressure, wind speed, temperature, etc.), which is most important for understanding the phenomena, but when considering social implementation and impact, it is necessary to organize and evaluate and impact in macro keywords such as the mass of weather phenomena (guerrilla downpour, tornado, frontal movement). We can consider that it is effective to consider both scales of phenomena bidirectionally. In that sense, we could give a suggestion that when there is a plan to conduct physical verification in terms of science, it is also important to consider teaming up the part that scrutinizes weather phenomena in terms of social application, such as the global warming scenarios.

By analyzing connections based on keyword similarity using embedding vectors, we could identify semantic links within BaleeGraph leading towards SDGs' targets. To actualize this path, we can further discuss that each issue team must fulfill two roles. Firstly, they need to convert input keywords into output keywords along the direction of the path. This process aligns with task completion such as technological development realization. Secondly, they must interpret keywords in a manner that connects similar terms used by other teams with their own keywords. When these roles are interconnected, viewing collaboration as a pathway reveals potential for achieving less obvious goals.

In projects like ClimCORE with substantial research budgets, it is crucial to ensure societal implementation of results. BaleeGraph facilitates redefining higher-order objectives by linking them with values such as SDGs. Although there is no direct link between the research and development discussion (weather and climate change) and SDGs target 14 (marine resources), they are connected by the keywords "sustainability" and "ecosystem conservation." This shows that they are indirectly connected to the discussion of social issues, which is the Society layer, rather than to each other in the Biosphere layer of the Wedding Cake Model of SDGs [24]. The reason why this connection is represented despite the fact that meteorological phenomena are closely intertwined with oceanographic phenomena is that the project members are relying on meteorological and social implementation issues. We can suggest two features of this visualization technique: 1) the impact on the ocean is implied through social implementation issues, even though members are unaware of it, and 2) there is a latent need for discussion and application of the ocean as a project. It was reaffirmed that the SDGs, which present representative goals in composing a sustainable world, still require mutual collaboration. Beyond SDGs alone, BaleeGraph can also aid discussions about connecting project outcomes with business interests. It does so by providing text describing business ideas from companies as input which helps visualize how these outcomes can be realized practically.

In the process of extracting keywords, it may be beneficial to incorporate manual operations as future options. One of the reasons is that the selection of keywords can vary depending on the reader's expertise and context, even when reading the same text. This variability can be accounted for in the visualization process. For instance, while the center of each balloon represents the same text, the keywords may differ from one team to another (i.e., from one balloon to another). This could aid in visualizing individual teams' interests and concerns.

The time-series visualization by BaleeGraph enabled us to observe how the ClimCORE project has evolved over time in Figure 4. We were able to visualize not only the growth of initiatives dealing with core concepts such as co-creation but also emerging seeds of co-creation related to disaster connections. This allowed us to verify whether the core concept was being integrated into the ClimCORE project while identifying future areas requiring nurturing.

Specifically, the restructuring of the ClimCORE is detailed in Figure 4, representing a major shift in the project's context. This shift emphasizes the integration and establishment of novel data to be developed within this project. The aim is to create data segments for each local region, which could be as diverse as cities, wards, villages, towns, and others. Furthermore, segments that are meaningful for regional collaborations are also considered. This approach is designed to uncover solutions for the contemporary problems that local governments are seeking to resolve.

An advanced use of BaleeGraph is to measure the quality of co-creations. BaleeGraph can provide three perspectives for this purpose. We can check whether the project realizes the good balance between uniqueness and common interests in a project by confirming the blue and red nodes; the ability to generate different scenarios through project activities to achieve external goals, especially goals such as SDGs and business ideas; the sustainable growth of the project through the highlighted nodes when comparing BaleeGraphs over different time periods.

7 Conclusion and Future Work

Our study has demonstrated that it's possible to visualize "co-creation" generation and growth with BaleeGraph in its application to the ClimCORE project. We have identified its unique strengths associated with each issue along with shared interests across issues. Our findings obtained through BaleeGraph suggest that visualizing co-creation could enhance understanding and contribute indirectly towards achieving social goods–specifically less directly related targets of SDGs. Moreover, by observing the temporal changes in the BaleeGraph, we were able to visualize the state of key concepts at different stages of generation and growth in the co-creation process. By providing this metacognition, BaleeGraph was confirmed to be able to support the achievement of social goods.

Currently, path discovery in BaleeGraph is performed manually by an analyst. However, automation is feasible as we can calculate distances between keywords or groups of keywords using embedding vectors. Given a starting point and an endpoint, we can compute all connecting paths. Subsequently, for a vector of paths with two defined endpoints, we can calculate distances for other paths. This allows us to recommend a path that is closest to a desired distance through computation.

We are contemplating offering BaleeGraph as either a web service or open source software. Interactive editing should be facilitated by enabling users to freely add and delete nodes. By showcasing various real-world applications through this platform, we aim to demonstrate its effectiveness and identify areas requiring improvement–thereby fostering development within a social cycle.

Acknowledgments

The University of Tokyo's DFIL social collaboration course, JST Grant Number JPMJPF2013 and Project Number 236100000155, and the MEXT Initiative for Life Design Innovation provided support for this study.

References

[1] Ministry of Internal Affairs and Communications, "Sustainable development goals (SDGs)," https://www.soumu.go.jp/toukei_toukatsu/index/kokusai/

02toukatsu01_04000212.html, 2009, Accessed: 2023/2, (in Japanese).

[2] K. Ueda, T. Takenaka, J. Váncza, and L. Monostori, "Value creation and decision-making in sustainable society," *CIRP Annals*, vol. 58, no. 2, pp. 681–700, 2009.

[3] M. V. J. Veenman, B. H. A. M. Van Hout-Wolters, and P. Afflerbach, "Metacognition and learning: Conceptual and methodological considerations," *Metacognition and Learning*, no. 1, pp. 3–14, 2006.

[4] ClimCORE Project Office, Research Center for Advanced Science and Technology, The University of Tokyo, "Website of Climate change actions with CO-creation powered by Regional weather information and E-technology (ClimCORE), COI-NEXT open innovation," https://climcore.rcast.u-tokyo.ac.jp/en/, 2023, Accessed: 2023/3.

[5] H. Nakamura, K. Kuma, K. Onogi, T. Miyasaka, Y. Makihara, J. Ishida, and M. Iida, "Toward high-resolution regional atmospheric reanalysis for Japan: An overview of the ClimCORE project," in *2022 IEEE International Conference on Big Data (Big Data)*, 2022, pp. 6153–6158.

[6] Y. Ohsawa, N. Benson, and M. Yachida, "KeyGraph: automatic indexing by co-occurrence graph based on building construction metaphor," in *Proceedings IEEE International Forum on Research and Technology Advances in Digital Libraries - ADL'98-*, 1998, pp. 12–18.

[7] D. Iwasa, T. Hayashi, and Y. Ohsawa, "Development and evaluation of a new platform for accelerating cross-domain data exchange and cooperation," *New Generation Computing*, vol. 38, no. 1, pp. 65–96, 2020. [Online]. Available: https://doi.org/10.1007/s00354-019-00080-0

[8] O. Yukio, I. Takaichi, and M. I. Kamata, "Kamishibai KeyGraph: Tool for visualizing structural transitions for detecting transient causes," *New Mathematics and Natural Computation*, vol. 06, no. 02, pp. 177–191, 07 2010.

[9] S. Lozano, L. Calzada-Infante, A.-D. Belarmino, and S. García, "Complex network analysis of keywords co-occurrence in the recent efficiency analysis literature," *Scientometrics*, vol. 120, no. 2, pp. 609–629, 2019.

[10] K. Sekiguchi and K. Hori, "Designing ethical artifacts has resulted in creative design: Empirical studies on the effect of an ethical design support tool," *AI & Society*, vol. 36, pp. 101–148, 3 2021,

[11] M. Sato, M. Akaishi, and K. Hori, "Topic bridging by identifying the dynamics of the spreading topic model," in *Intelligent Interactive Multimedia: Systems and Services*, T. Watanabe, J. Watada, N. Takahashi, R. J. Howlett, and L. C. Jain, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 619–627.

[12] F. Windhager, L. Zenk, and P. Federico, "Visual enterprise network analytics - visualizing organizational change," *Procedia - Social and Behavioral Sciences*, vol. 22, pp. 59–68, 12 2011.

[13] M. Ishihara, "A visualization method for the growth process and difference of dynamic networks," Master's Thesis, Graduate School of Systems and Information Engineering, University of Tsukuba, March 2007, (in Japanese; Title was translated by authors).

[14] K. Tanaka and K. Hori, "Finding diachronic objects of drifting descriptions by similar mentions," in *Knowledge Management and Acquisition for Intelligent Systems*, K. Ohara and Q. Bai, Eds. Cham: Springer International Publishing, 2019, pp. 32–43.

[15] R. Abumalloh, M. Nilashi, M. Yousoof, A. Alhargan, A. Alghamdi, A. Alzahrani, L. Saraireh, R. Osman, and S. Asadi, "Medical image processing and covid-19: A literature review and bibliometric analysis," *Journal of Infection and Public Health*, vol. 15, 11 2021.

[16] G. Cernile, T. Heritage, N. J. Sebire, B. Gordon, T. Schwering, S. Kazemlou, and Y. Borecki, "Network graph representation of covid-19 scientific publications to aid knowledge discovery," *BMJ Health & Care Informatics*, vol. 28, no. 1, 2021.

[17] S.-H. Lim, J. Chae, G. Cong, D. Herrmannova, R. M. Patton, R. Kannan, and T. E. Potok, "Visual understanding of covid-19 knowledge graph for predictive analysis," in *2021 IEEE International Conference on Big Data (Big Data)*, 2021, pp. 4381–4386.

[18] OpenAI, "OpenAI API," https://platform.openai.com/overview, 2020.

[19] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 11 2019.

[20] ——, "Making monolingual sentence embeddings multilingual using knowledge distillation," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 11 2020.

[21] NetworkX Developers, "Software for complex networks," https://networkx.org/, 2014.

[22] A. Hagberg, P. Swart, and D. S Chult, "Exploring network structure, dynamics, and function using NetworkX," in *Proceedings of the 7th Python in Science Conference (SciPy2008)*, G. Varoquaux, T. Vaught, and J. Millman, Eds., 2008, pp. 11–15.

[23] Pyvis Developers, "Pyvis: A python library for visualizing networks," https://pyvis.readthedocs.io/en/latest/, 2016.

[24] Stockholm Resilience Center, "The sdgs wedding cake," https://www.stockholmresilience.org/research/research-news/2016-06-14-the-sdgs-wedding-cake.html, Accessed: July 30, 2023.

[25] A. Shimpo, K. Takemura, S. Wakamatsu, H. Togawa, Y. Mochizuki, M. Takekawa, S. Tanaka, K. Yamashita, S. Maeda, R. Kurora, H. Murai, N. Kitabatake, H. Tsuguti, H. Mukougawa, T. Iwasaki, R. Kawamura, M. Kimoto, I. Takayabu, Y. N. Takayabu, Y. Tanimoto, T. Hirooka, Y. Masumoto, M. Watanabe, K. Tsuboki, and H. Nakamura, "Primary factors behind the heavy rain event of July 2018 and the subsequent heat wave in Japan," *SOLA*, vol. 15A, pp. 13–18, 2019.

[26] T. Horinouchi, Y. Kosaka, H. Nakamigawa, H. Nakamura, N. Fujikawa, and Y. N. Takayabu, "Moisture supply, jet, and silk-road wave train associated with the prolonged heavy rainfall in Kyushu, Japan in early July 2020," *SOLA*, vol. 17B, no. Special_Edition, pp. 1–8, 2021.

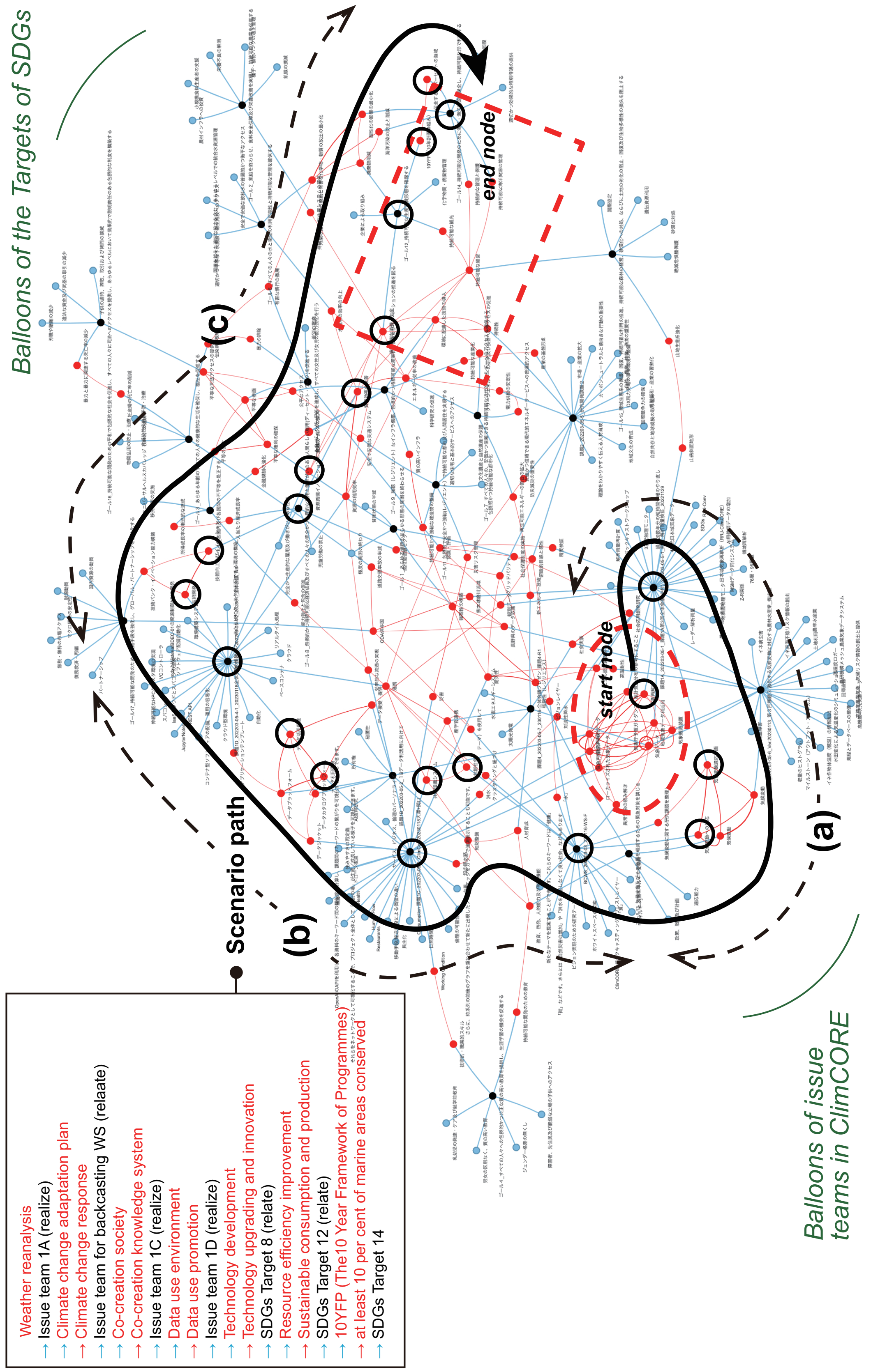

Figure 3: A Visualization Result of From the ClimCORE project keywords to the SDGs target keywords.

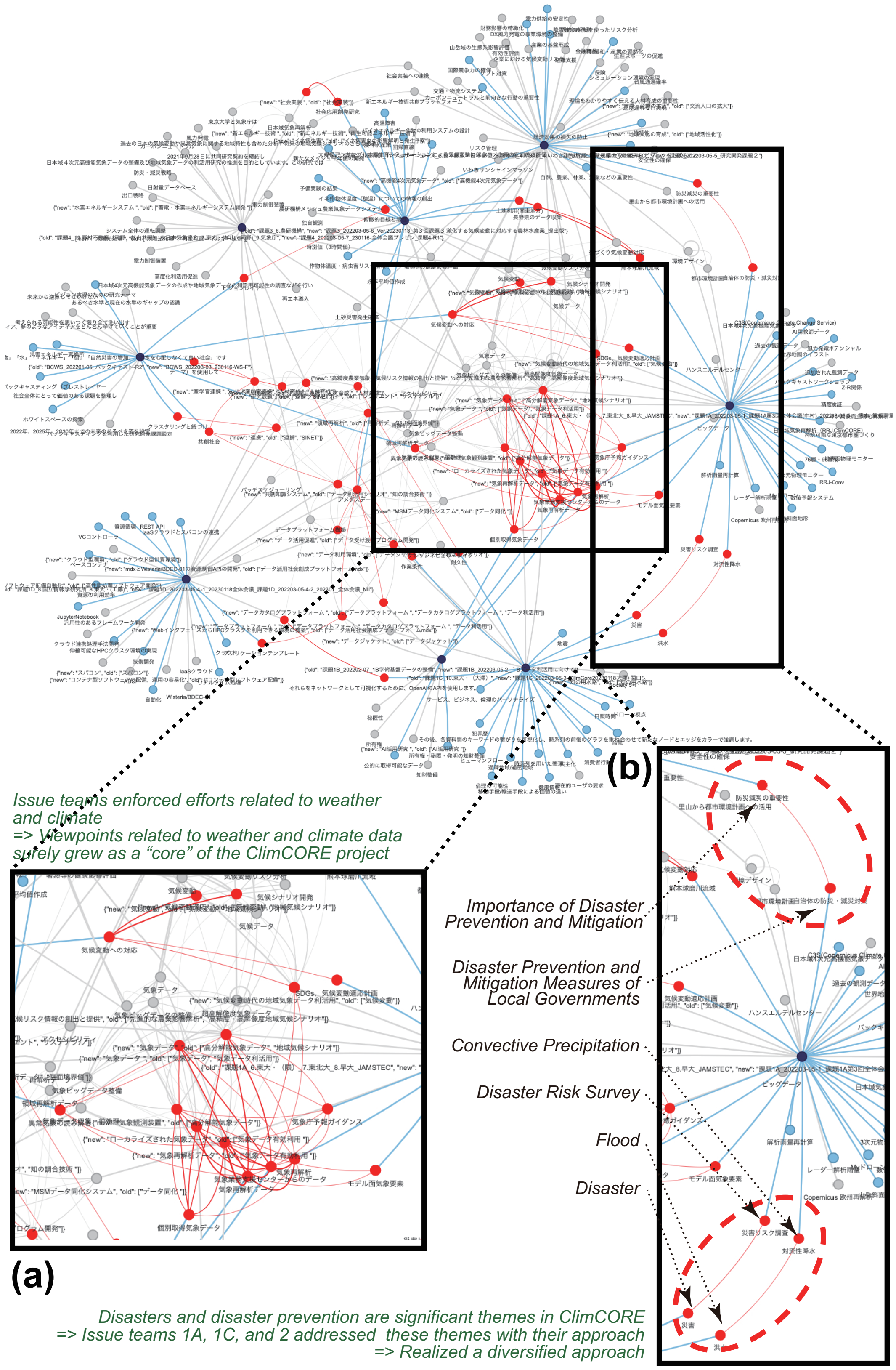

Figure 4: A Visualization Result of Time-series Changes Between the 3rd General Meetings in Fiscal Year 2022 and Balloons in Older General Meetings When Each Issue Team First Appeared.