Network Simplification and Visualization through System Dynamics-based Network Centrality

Lukas Frannek, Haruko Nagaoka, Tadasuke Nakagawa
Hitachi Ltd., Global Center for Customer Co-Creation, Tokyo

Abstract

When developing innovative products in today's business world, companies try to optimize the process of indentifying and discussing customer problems, value and opportunities as much as possible to advance faster to new business creation, proof of concept and finally contract agreement. In this paper we use system dynamics as a methodology that facilitates the modeling and visualization of causal business indicator relationships in a compact and intuitive way that clearly points to business problems and therefore leads to value and opportunity discovery. Specifically, we propose a technique that automatically simplifies a given network of nodes based on node importance, such that a fast understanding of interdependencies and problems in time-limited workshops and other occasions can be assured.

We first state the principle algorithm used to determine node importance, and then explain how the network is displayed based on node importance and user input. We illustrate the usage of this simplification and visualization method with an example in the field of visitor prediction.

1. Introduction

In the last decade the concept of customer co-creation became a widely used technique to engage clients as a group of companies, where the companies work together to create value that is specific to each client [1]. Companies want to find that specific value faster in order to be able to faster advance to later stages in the business discussions that propose a company's products and services to the client. This requires easy-to-understand ways to illustrate the client's problems and the suitable measures that fix those problems. It is vital to come to an agreement with all involved parties. System dynamics proved to be able to provide valuable tools regarding group model building in workshop-style meetings and the visualization of these models [2]. Specifically, the causal loop diagram (CLD) serves as a tool that can help to understand complex relations between a multitude of business indicators all the way up to management key performance indices (KPIs) such as profit, return on investment or sales. In a workshop and in the following meetings when explaining the outcomes of the workshop, time may be limited and understanding a large CLD is not feasible in that case. However, we may want to gather opinions and comments from persons that did not participate in the workshop, where we also may not be clear about what additions to a CLD a person can offer. On the other hand we also may need to present the workshop outcomes to various persons, where we may not know what kind of information a person is looking for.

As a means to condense a given network into a form that makes a quick understanding and exploration of that network possible, we propose a method that consists of two parts. The first part of the method determines the importance of nodes in the network with a novel calculation for the node importance. The second part of the method simplifies the network by consolidating nodes that are less likely to be relevant to the specific customer, i.e., nodes that likely have low importance to the customer.

A. Related Works

To find the importance of nodes in a network various techniques exist such as the degree centrality [3], that is determined by incoming and outgoing edges, the betweenness centrality [3], that is determined by the number of times a node acts as a bridge between other nodes, or the page rank [4], that assesses how many nodes can be connected via paths, where paths to high importance nodes add more importance. While the degree and betweenness centralities are less useful to determine the importance of nodes from a system dynamics viewpoint, the page rank takes also distant nodes and their importance into account, and thus not only assesses the topological structure. However, it still lacks the ability to grasp the main determining factors, such as feedback, in networks such as the CLD.

In the proposed method, the importance for each node in a network is found by utilizing two types of input. The first type is information based on system dynamics parameters such as loop and pattern structures. We explain what a pattern constitutes in Chapter 2. The second type of information is user input that is captured during workshops with clients or otherwise. We explain what this user input specifically constitutes in Chapter 2 as well.

The display or consolidation of nodes is based on the importance. Consolidation means that nodes that are less likely to be relevant disappear from the view of the person who views the network. Similar to the network pruning explained in references [5,6], nodes with an importance that is below some threshold are removed, but without taking any meaning away from the main structure of the network and while preserving connectivity. However, the removed nodes are not viewable after the pruning anymore, and the importance of nodes may change in subsequent meetings, which requires the nodes to be still accessible. Another technique for network simplification is given in reference [7], where nodes are rearranged in layers to represent their importance. Even though the visualization becomes cleaner, the difficulty of reconstructing the mental models of the participants of the CLD creation workshop during the workshop makes this approach unfeasible. In reference [8] the geometrical structure of a network is not recreated from scratch, but adjusted in order to preserve the mental map of a network. This approach is more suitable, but for the previously stated purpose of preserving the mental models of CLD workshop participants we refrain from any adjustments to node positions.

The display method in this paper consolidates, or hides, nodes based on importance, such that they can still be found by expanding nodes in the network to allow for an exploration of the network. In addition, the display method only adds or removes nodes and edges at the exact same positions the initial nodes and edges were located in the network. The way the consolidation works and the benefits of this approach compared to other technique for rearranging networks are explained in Chapter 3.

B. Problem Definition

As mentioned above, one part of the problem that we tackle in this paper is the time required to read and understand a CLD in time-limited meetings with unspecified providers and recipients of information. The increased time requirement when presenting the output of CLD workshops arises from the multitude of nodes and overlapping edges which impede the understanding of a large CLD and in addition also decrease aesthetics of the network as well as the patience and interest from participants in time-limited meetings. This hinders the effective usage of CLDs as decision making support and the adoption of system dynamics as a valid tool in business. Usually post-processing is necessary, and even then the result leaves a lot to be desired, practically making the extraction of the main information and creation of separate content necessary. However, at an early stage in a project the creation of separate content might not be feasible since there were not enough discussions or available key members. In addition, we want to utilize the CLD directly since the way of thought of the CLD creators can be often retraced from the CLD.

The second part of the problem we tackle in this paper is related to the possibility to use the CLD as a story building board. To this end, the possibility to expand consolidated nodes is very important since it allows for an exploration of the network and retracing of the mental models of the CLD creators.

Note that the creation of the overall method for simplification and visualization was developed as a direct response to business needs that arose during the usage of the CLD in business situations.

2. Node Importance

A. Preamble

As a preamble to the introduction of the calculation method for node importance in this chapter, we describe the general setting where this tool is intended to be used. The creation of a CLD in a typical workshop uses paper as the main medium. Workshop participants use post-its to write down their ideas and arrange them on a large piece of paper or a white board. There are various drawbacks to this way of conducting workshops, such as interruptions of the train of thought, unwieldy workshop utensils or illegible writing. Hitachi Ltd. developed a framework that supports this process through computerized tools, the NEXPERIENCE framework [9]. Within this framework, CLDs can be created digitally, alleviating many drawbacks of creating CLDs on paper. Furthermore, sharing and presenting digital CLDs becomes easier, and an automatic analysis becomes possible. In this paper we deal with CLDs that have been created digitally.

The points that can be leveraged to improve an insufficient visualization are three-fold. First, many nodes that are not relevant are displayed too, and while they are necessary to draw the CLD they are not necessary to quickly explain the gist of the CLD. Secondly, the loops might not be sufficiently clear, and especially in larger CLDs this is a challenge. Thirdly, we use certain common patterns in CLDs that consist of particular arrangements of nodes and edges. These patterns are industry-independent and indicate certain problems or situations in a business structure [10,11]. There exist various patterns which were defined beforehand and stored in some sort of database, such that they can be accessed and used as a reference to identify instances of these patterns in a given CLD. An example of a pattern is the “Limits to Success” pattern in Fig. 1.

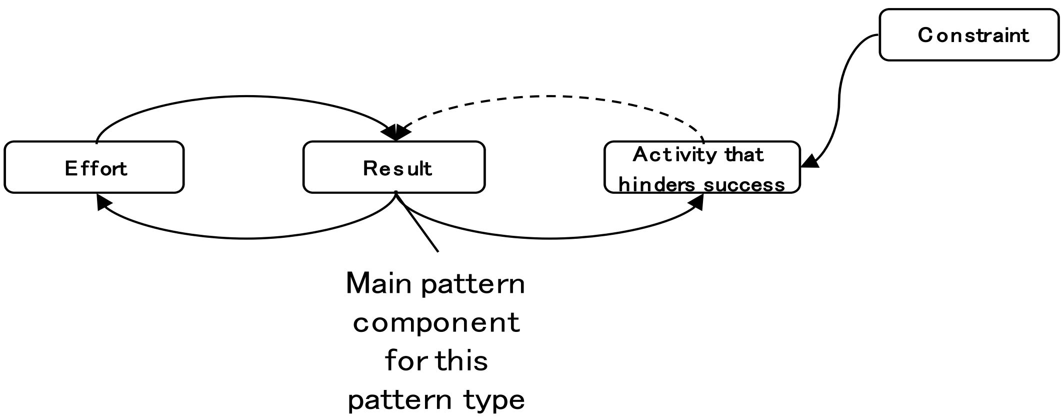


Figure 1: ‘Limits to Success’ Pattern Structure

In this pattern a ‘Result’ node is increased by a reinforced loop through an ‘Effort’ that can be contributed to directly, whereas the ‘Result’ can usually not be directly influenced from a business point of view. Opposed to the ‘Effort’ is an ‘Activity that hinders success’ that decreases the ‘Result’ through a negative feedback loop. The ‘Activity that hinders success’ can usually be directly influenced or indirectly through the ‘Constraint’. Note that each node, except the ‘Result’ that is the main node and defined as tha main pattern component, can consist of multiple nodes that were summarized. There is always just one ‘Result’ node.

A more complex pattern is the ‘Success to the Successful’ pattern in Fig. 2.

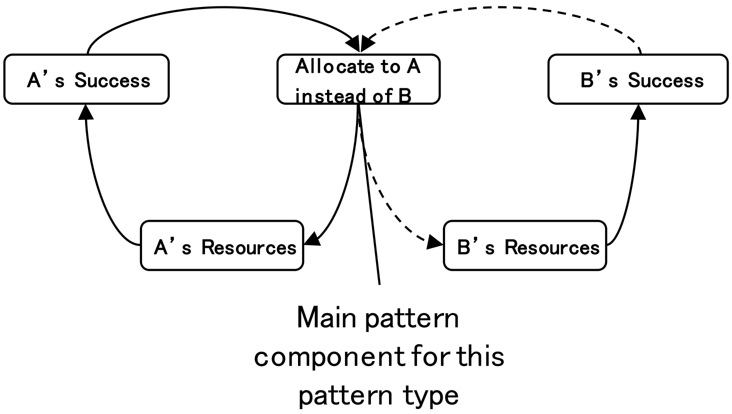


Figure 2: ‘Success to the Successful’ Pattern Structure

In this pattern two entities, A and B, use the same resources which means that the success of the first entity leads to a higher share of the resource for the first and a lower share of the resources for the second entity. The challenge when using patterns is that they can also be considerably complex, which makes them even more difficult to find and understand.

Note that, just as loops, patterns are created, or not created, naturally in the CLD creation process, without any particular attention from, for instance, the CLD workshop facilitator.

B. System Dynamics Centrality

In the calculation for the node importance we take the information regarding loop membership, pattern membership and pattern main component membership of a node and the average importance of nodes in a loop into account.

We define a loop as a set of edges and nodes that connects an arbitrary node with itself with at least one other node in between. Patterns are defined as particular arrangements of edges and nodes, such as in Fig. 1 and Fig. 2, and are stored in a data file to be compared with a given CLD. The algorithm for the identification of loops and patterns reads in a new CLD and assigns IDs to each node in the CLD. With a depth-first search the algorithm finds node-edge-node combinations, such as Effort-edge-Result in Fig. 1 when searching for 'Limits to Success' patterns, and sets this not yet completely identified pattern before continuing to the next node-edge-node combination, in this case Result-edge-Activity that hinders success, until the pattern has been completely identified or until the search ends due to preset conditions. Note that the Effort-edge-Result combination in Fig. 1, for instance, can consist of more than 2 nodes due the summary of multiple nodes. A more thorough explanation of the algorithm is given in reference [12], while references [9,13] contain information on how the algorithm is intended to be used.

The outputs of this algorithm are arrays for each loop and each pattern that contain the IDs of the included nodes, where each ID in a pattern array is given the suitable label, such as 'Result' in Fig. 1, thus identifying the main pattern components. We convert this information into a form where each loop, pattern and main pattern component is given a vector that has a length equal to the overall number of nodes in a given CLD. We denote these vectors

$$l_i \in \mathbb{L},$$
$$p_i \in \mathbb{P},$$
$$m_i \in \mathbb{M},$$

for loops, patterns and main pattern components respectively, where i is an index, $\mathbb{L}$ is the set of loops, $\mathbb{P}$ is the set of patterns and $\mathbb{M}$ is the set of main pattern components in the network. If a node is a member of a loop or pattern, the corresponding entry in the vector for that loop or pattern is set to 1.

Note that patterns, such as in Fig.1 and Fig. 2, include loops as well. In this paper we make the assumption that loops that are used in a pattern, are not again included in the set of loops $\mathbb{L}$. This is assured when converting the output of the algorithm to the vectors above.

Based on the information from the three matrices, we can define the intermediate importance IM as

$$\mathrm{IM} = \sum_{\mathrm{i} \in \mathbb{L}} \mathrm{l_i}\, a + \sum_{\mathrm{j} \in \mathbb{P}} \mathrm{p_j}\, b + \sum_{\mathrm{k} \in \mathbb{M}} \mathrm{m_k}\, c,$$

where the parameters a, b and c are weights that indicate the importance we assign to loops, patterns and main pattern components. Generally main pattern components are more important in the overall network structure since they are by experience centrally located and connected to various loops. Nodes that are not main pattern components but still members in patterns have a relatively high importance since they represent important feedback structures, whereas loops are slightly less important. If a node in the network is neither a member of a loop nor a member of a pattern it can be ignored unless it was selected as a KPI node or a controllable node during the CLD creation. We define a KPI node as a node that was identified to have significant impact on the customer's business and a controllable node as a node that was identified to have significant potential to change the current business structure due to the customer's ability to change that node. The non-membership of nodes does not occur very often since CLDs are commonly created around KPI nodes and controllable nodes in order to understand interdependencies and problems better.

After the calculation of the intermediate importance we calculate the system dynamics centrality (SDC) with

$$\mathrm{SDC} = \mathrm{IM} + \sum_{\mathrm{i} \in \mathbb{L}} \mathrm{l_i^T} \frac{\mathrm{IM}}{\mathrm{sum}\,(\mathrm{l_i})} \mathrm{l_i}.$$

In this equation we calculate the average loop importance for each loop by summing up the importance of each node in the loop and then dividing that sum by the number of nodes in that specific loop. We add the resulting number to each node in that loop. After conducting this operation for each loop in the CLD we receive the SDC.

The reason for this calculation is that we want to emphasize the few loops that have a

very big impact on the business. Nodes may not receive a high importance only based on the IM, since they are not part of patterns. However, if other very high importance nodes are contained in the same loop as the low importance nodes, the low importance nodes have an impact on the high importance nodes, thus becoming more important themselves. This relationship is reflected in the SDC and is related to the page rank algorithm. The elements of the SDC vector are the values for each node's importance.
The second type of input used to determine importance is user input. This user input consists of KPI node designations and controllable node designations. KPI nodes and controllable nodes are decided before or in a workshop for CLD creation. These designations do not change the SDC but are considered during the drawing of the consolidated CLD.

3. Node Consolidation

Taking the SDC vector as the input, the decision of whether a node should be visible or not is made based on the average node importance, given by

$$\text{Average Node Importance} = \sum_{i=1}^{N} \frac{SDC_i}{N},$$

where the total number of nodes in the CLD is denoted as N. Depending on the use case, the average node importance is multiplied by a factor that increases or decreases the threshold that determines if a node is displayed or not. If a node's importance is below that threshold it is not displayed. Instead, it is replaced by a symbol such as a dot at the exact position in the diagram where the node was before, Fig. 3.

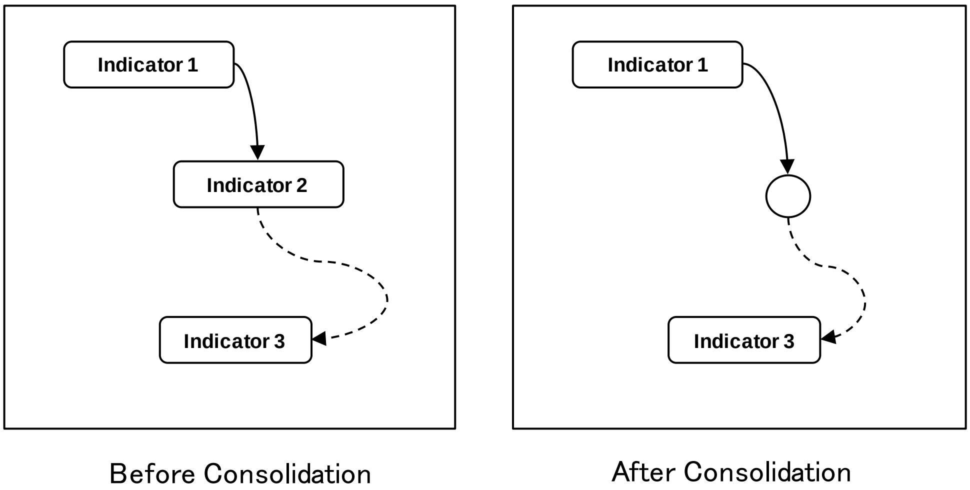


Figure 3: Consolidation by Representing Nodes with Low Importance as a Dot

After conducting this check for all nodes, the actual consolidation of dots starts. Starting at the dot with the lowest importance, all directly connected dots are consolidated into one dot, pair by pair, Fig. 4.

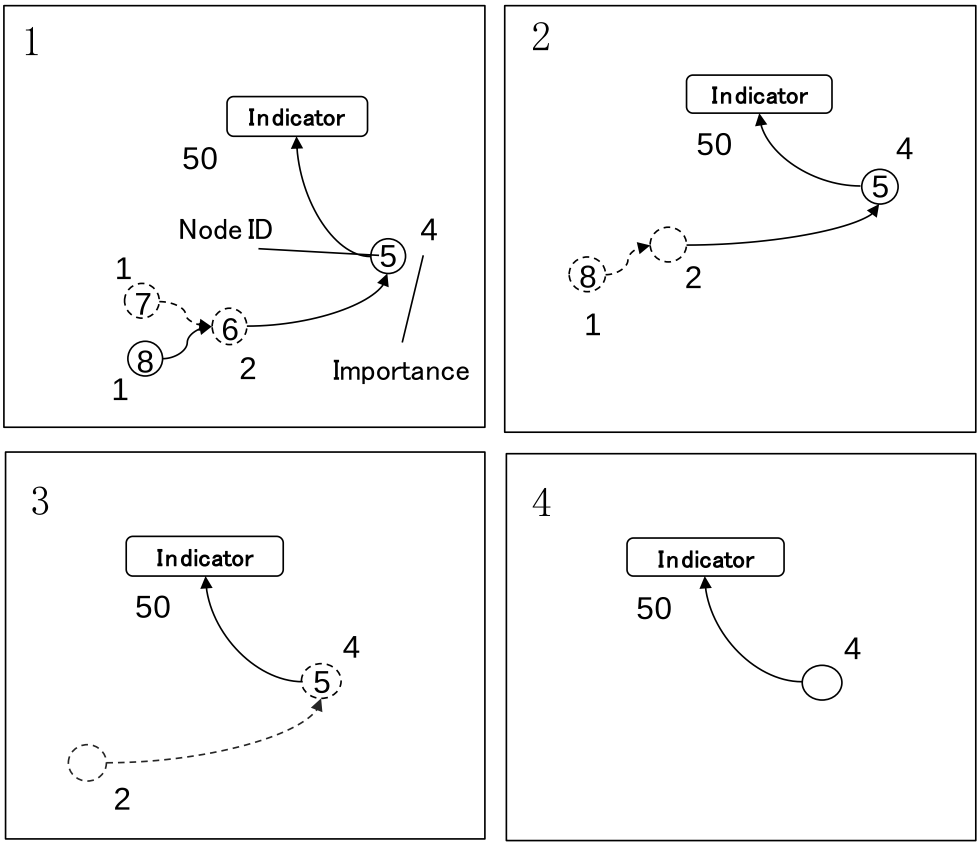


Figure 4: Step-by-Step Consolidation of Nodes

The edges between dots are hidden as well, while the edges between dots and nodes

remain. In the end this creates several clusters that are each represented by one dot and separated by nodes with above-average importance. The idea is then to expand dots that are connected to nodes of interest pair by pair, where the expansion starts with the node that has the highest importance, such that a meaningful exploration of clusters of interest can take place. An illustration of this exploration is given in Appendix 1-3. The important part is that nodes and dots are not moved from their position, only hidden, such that the CLD creators' train of thought can be better understood.
The user of this visualization simplification can, after the initial consolidation, expand and contract dots at will, by using the appropriate symbols in the user interface.

4. Example Case

A. Network Simplification Process

As an illustrating example for the simplification technique proposed in this paper we chose the example of visitor prediction for a department store. Assume the output of an initial workshop is given in Fig. 5.

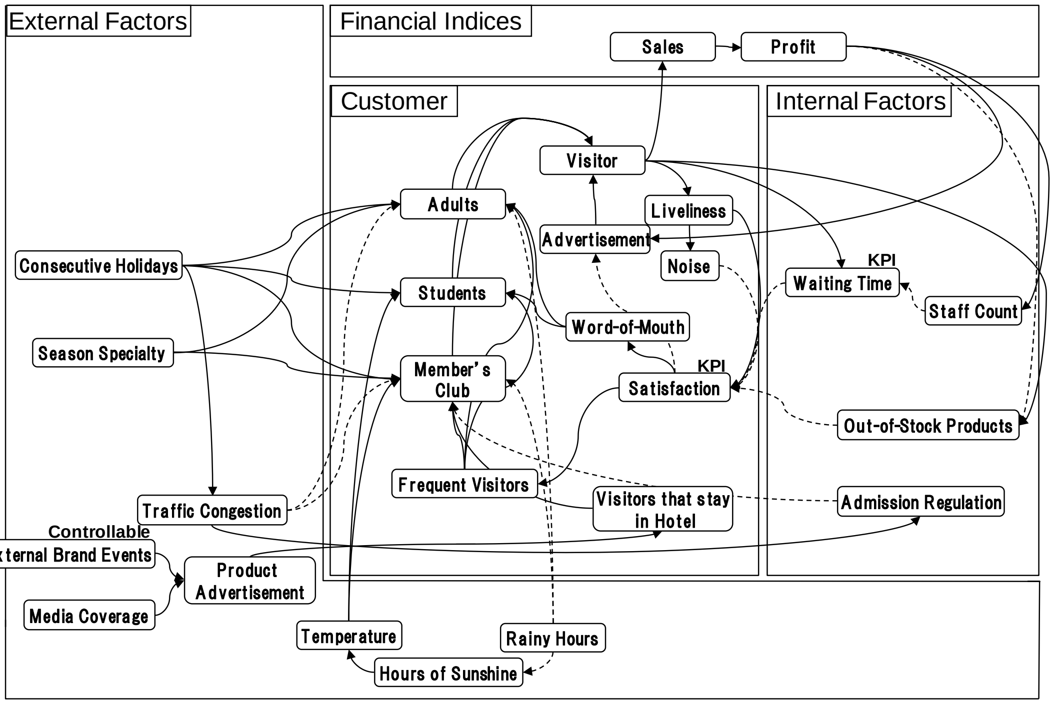


Figure 5: CLD for Understanding Department Store Visitor Prediction

The CLD in Fig. 5 is not very complex but sufficient to illustrate how the proposed simplification method works. During the CLD workshop it was found that the client in this specific case is interested in implementing measures that reduce the waiting time for various shop related activities, such that the overall satisfaction of customers who come to the department store is increased. Hence, the waiting time and the satisfaction are set as KPIs. As a controllable node external brand events were identified, thus marking the corresponding node as controllable. In this example we can discover two system patterns. One of them is 'Limit to Success' as given in Fig. 6.

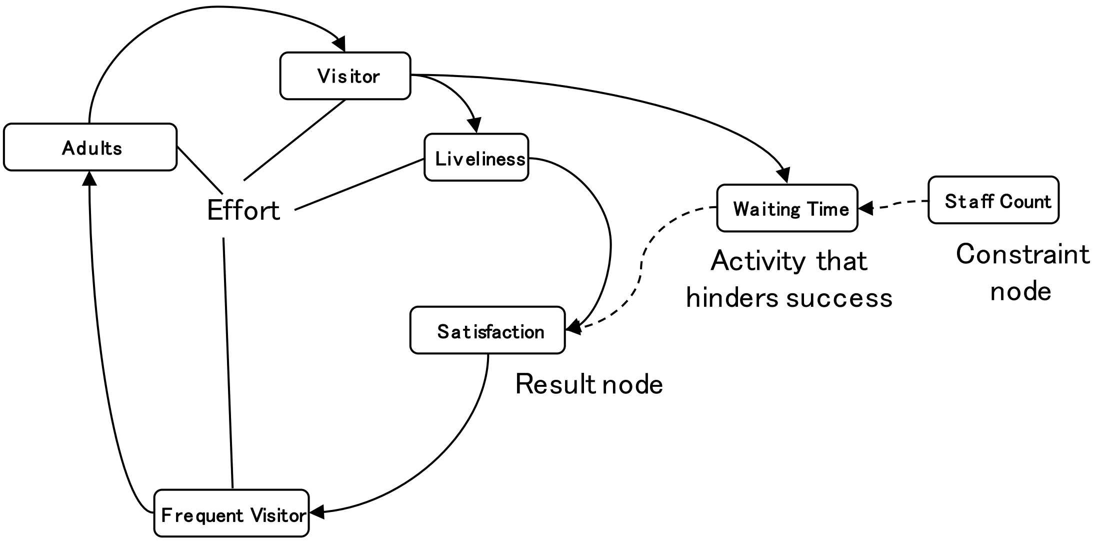


Figure 6: 'Limit to Success' Pattern in the Example CLD

Note that the 'Effort' can take other forms as well in this CLD, thus forming multiple occurrences of the 'Limit to Success' pattern. Also note that the satisfaction node is the main pattern component in each of these occurrences.

The second pattern we can find is the 'Success to the Successful' pattern, given in Fig. 7.

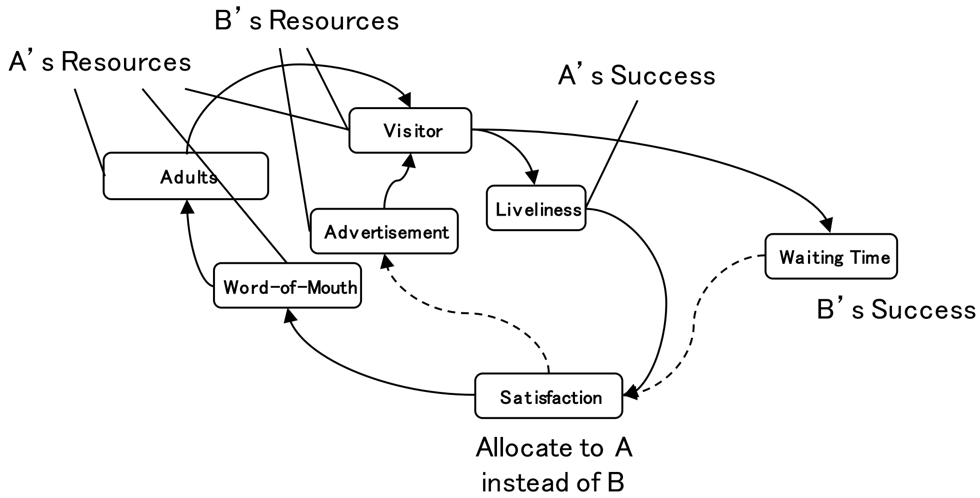


Figure 7: 'Success to the Successful' Pattern in the Example CLD

This pattern also occurs multiple times, increasing the importance of the satisfaction node even more. An example of a loop that is not used in a pattern, but includes very important nodes, is given in Fig. 8.

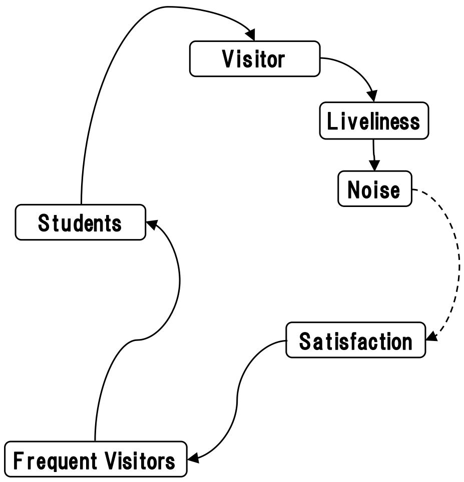


Figure 8: Increased Node Importance of Nodes outside Patterns through the Assessment of the Average Loop Importance

Since satisfaction and visitors are both very important nodes, which can be seen in the previously displayed patterns, noise receives a higher importance, increasing the possibility of being displayed and not hidden.

To calculate the importance for each node we set the parameters a, b and c to 1,2 and 3 respectively. The importance numbers are given in Fig. 9.

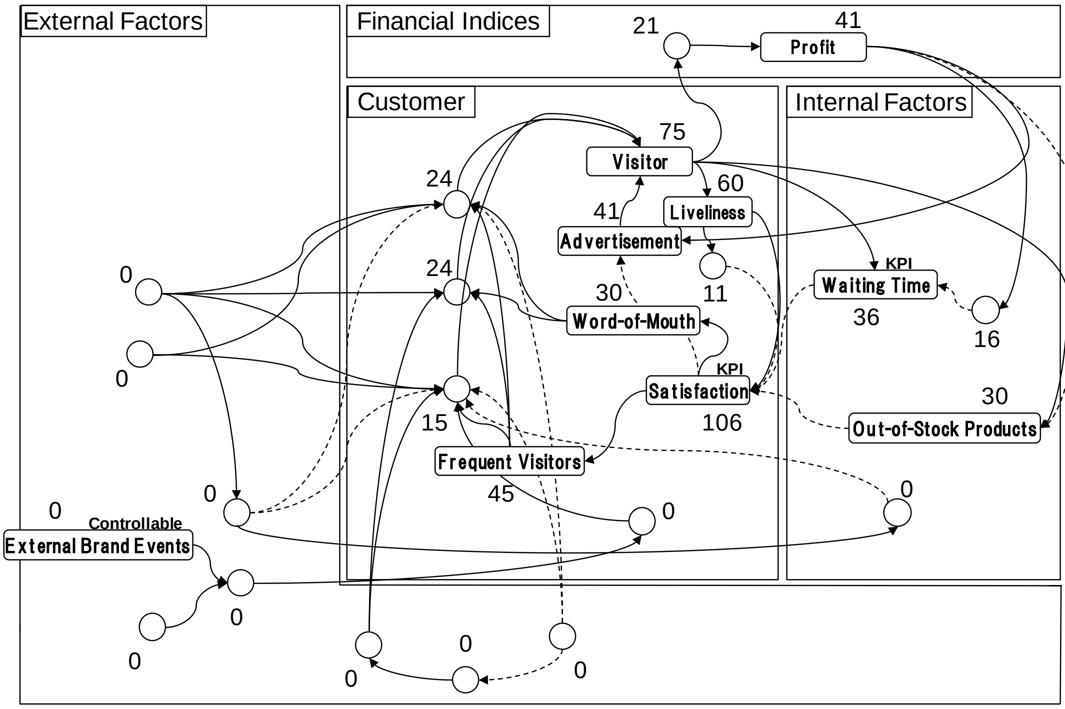


Figure 9: SDC Display and Representation of Unimportant Nodes as Dots

After consolidating all nodes the final CLD is given as displayed in Fig. 10.

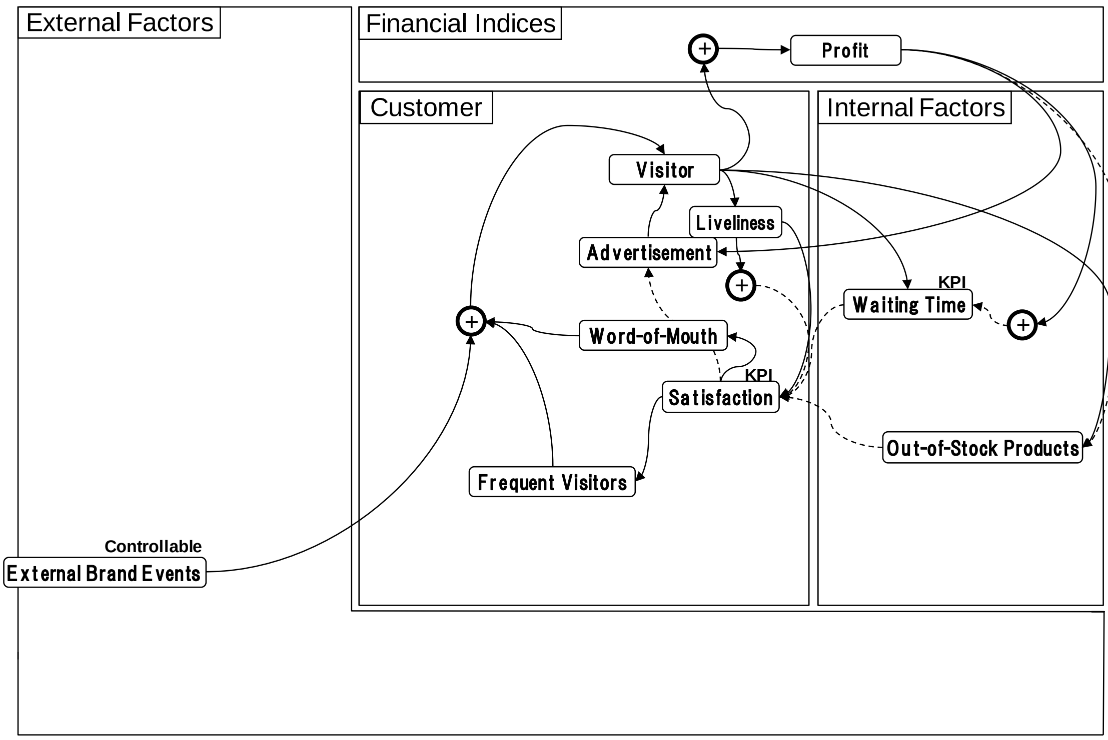


Figure 10: Final Output of the Simplification Method

In this specific case the average node importance is multiplied by 1.2, setting the threshold that determines if nodes are displayed or not to 27.1.

B. Evaluation

With regards to the reduction of time that is necessary to understand the CLD, persons who view this CLD for the first time can quickly indentify KPI and controllable nodes, and the nodes that influence them. From this basic understanding that acts as a platform, other pathways can be easily explored, for instance, by management level staff who want to get a more detailed picture of financial indices or department store layout planners who are able to add important factors that could reduce waiting time. Since the viewer of the network is confronted with less nodes and edges, compared to the original view of the CLD, the positions in the network where the financial indices and waiting time-related indices can be found, are easy to determine as well.

The majority of nodes that were consolidated are in the 'External Factors' category, which mainly contains nodes that cannot be controlled or influenced such that the satisfaction or waiting time would improve. This circumstance is represented by the fact that these nodes are not part of any pattern or loop, which therefore results in a low importance and the associated consolidation of these nodes, thus reducing view-cluttering nodes and edges. 'External brand events' are an exception since they were specifically identified as a controllable node because, even though external brand events have little influence in the dynamics of the CLD, they are important to the customer. It is easily imaginable that a subsequent meeting could update the CLD by connecting the 'Profit' to 'External brand events', which would result in a different view.

With regards to the usage of the simplified CLD as a story building board, related persons can be easily led along story paths in the CLD after those persons gained a basic understanding of KPI, controllable nodes and their interdependencies. As an example, the path from 'External brand events' to 'Visitors' and 'Satisfaction' could be explored by branching out the dot connected to 'External brand events'. Expanding this path would make it easier to understand the mental model of the CLD creators, and if the mental model was understood, meaningful additions and changes could be made

such as the connection between ‘External brand events’ and ‘Profit’ mentioned above.
In the end, the time needed to understand the CLD and find nodes of interest, consolidated or not, was reduced in the case of this explanatory CLD. CLDs that are usually created in business situations contain significantly more nodes, increasing the number of expandable clusters and their content to a great extent, thus making the proposed method even more effective.

5. Conclusion

In this paper we proposed a method for simplifying the visualization of networks from a system dynamics viewpoint. In Chapter 1 we gave the motivation for this research, surveyed related techniques, and explained our contribution. We continued to give details regarding the importance calculation in Chapter 2 and the network display in Chapter 3. Chapter 4 contained an example case that was used to illustrate the usage of the proposed method.
The main purpose of the proposed method is to simplify cluttered networks such that less time needs to be spent to understand the network in a system dynamics sense, for operators that may contribute to the network through their knowledge of low level processes as well as for management staff that wants to receive a quick assessment of the business situation. As the example case indicates, the gist of the customer’s business structure can be assessed and discussed quicker with the simplified CLD, compared to the CLD before the simplification. This can potentially lead to faster understanding of the business structure, goals and problems by various parties, and an accelerated consensus building, two important elements for achieving an effective customer co-creation.
Another advantage the proposed method has is the ability to explore a network, an ability that supports the story telling in a proposal process. This is best combined with an appealing user interface that might add additional information to the view of the network. From a business standpoint, further developments involve a more appealing presentation with more information, and dynamic real time network changes that allow for a better interaction with the customer and related persons.

6. References

[1] Prahalad, C.K., Ramaswamy, V., “Co-Opting Customer Competence”, Harvard Business Review, 2000

[2] Sterman, J.D., *Business Dynamics: Systems Thinking and Modeling for a Complex World*, McGraw Hill/Irwin, 2000

[3] Freeman, L. C., “Centrality in social networks: Conceptual clarification”, *Social Networks,* Vol. 1, pp. 215-239, 1978

[4] Page, L., Brin, S., Motwani, R. and Winograd, T., “The PageRank citation ranking: Bringing order to the Web”, Technical Report, 1998

[5] Zhou, F., Mahler, S. and Toivonen, H., “Simplification of networks by edge pruning”, *Bisociative Knowledge Discovery*, vol. 7250, No. 13, pp. 179–198, 2012

[6] Hennessey, D., Brooks, D., Fridman, A. and Breen,D., “A simplification algorithm for visualizing the structure of complex graphs”, Proceedings of the 2008 12th International Conference Information Visualization, pp. 616-625, 2008

[7] Sugiyama, K., Tagawa, S., Toda, M., “Methods for visual understanding of hierarchical system structures”, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 11, No. 2, pp. 109-125, 1981

[8] Misue, K. *et al.*, “Layout Adjustment and the Mental Map”, *J. Visual Languages and Computing,* Vol. 6, No. 2, pp. 183-210, 1995

[9] Nagaoka, H., Nakamura, T., Nakagawa, T., Kaneda, M., “Development of methods for visualizing customer value in terms of people and management”, *Hitachi Review*, Vol. 65, No. 2, 2016, to be published

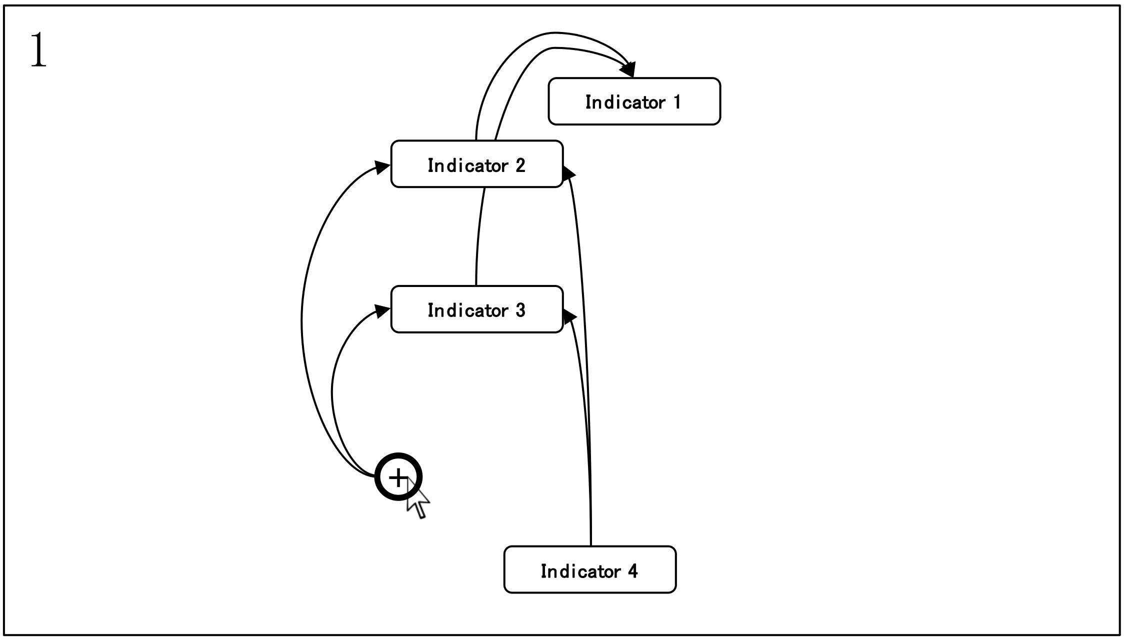
[10] Wolstenholme, E., “Using generic system archetypes to support thinking and modeling”, *Syst. Dyn. Rev.*, Vol. 20, pp. 341–356, 2004

[11] Braun, W., *The System Archetypes,* The Systems Modeling Workbook, 2002

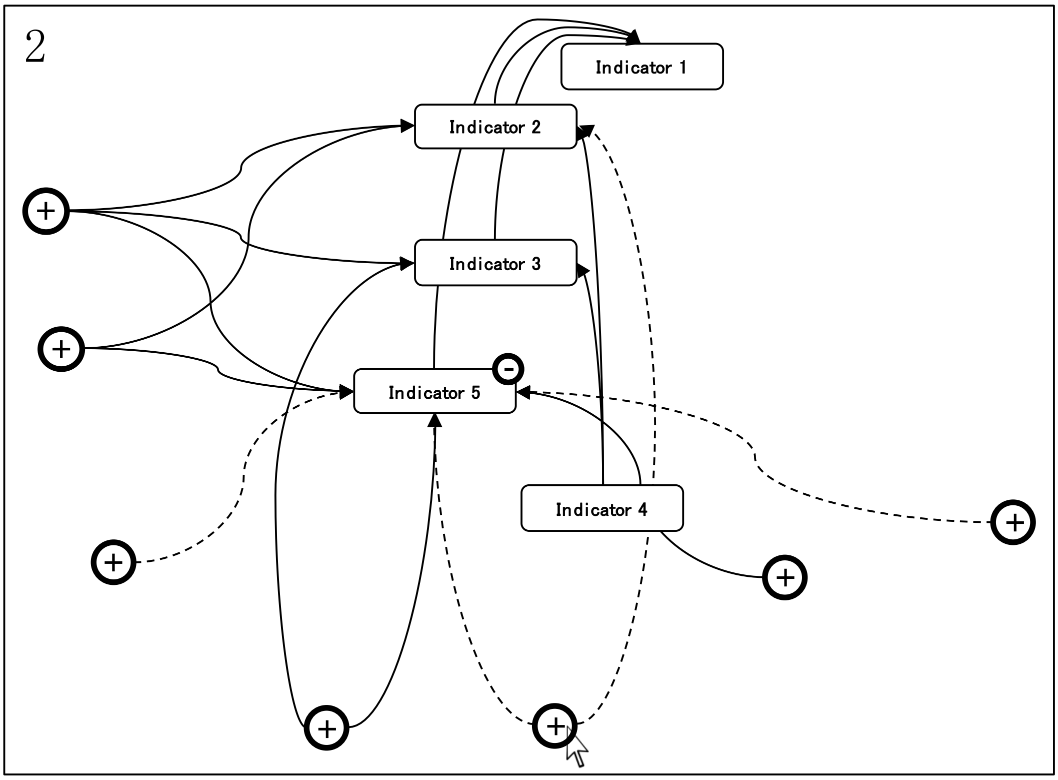
[12] Aoshima. Y., Nagaoka, H., Nakagawa, T., Tanaka, E., 問題構造抽出装置および問題構造抽出方法 [Problem structure extraction apparatus and problem structure extraction method], Japan patent application number 2014-204903, October 2014

[13] Nagaoka, H., Tanaka, E., Miyamoto, K., “Performance evaluation model for service business involving multi-stakeholders”, Asian Conference on Information Systems, 2014

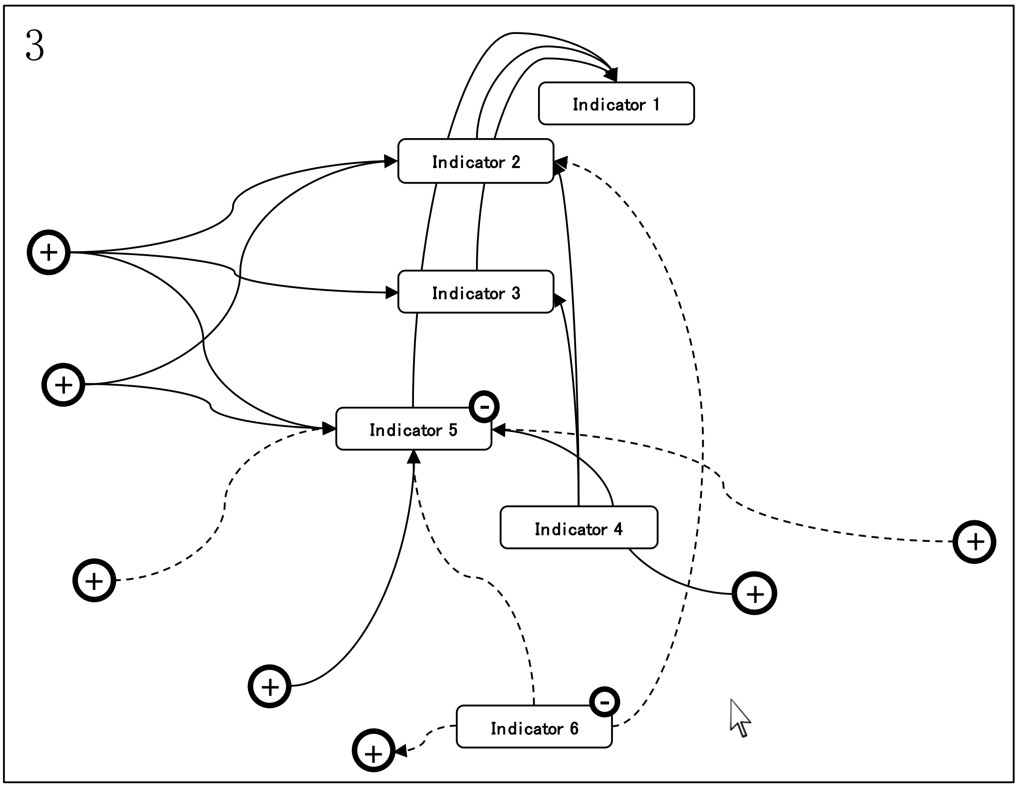
Appendix



Appendix 1: Illustration of Node Expansion 1



Appendix 2: Illustration of Node Expansion 2



Appendix 3: Illustration of Node Expansion 3