

Graph-Based Open-Ended Survey on Concerns Related to COVID-19

Tatsuro Kawamoto (✉ kawamoto.tatsuro@aist.go.jp)

National Institute of Advanced Industrial Science and Technology

Takaaki Aoki

Kagawa University

Michiko Ueda

Waseda University

Research Article

Keywords: COVID-19, multiple-choice survey, interconnectedness

Posted Date: December 14th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-124312/v1>

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

[Read Full License](#)

Graph-based open-ended survey on concerns related to COVID-19

Tatsuro Kawamoto^{1,*}, Takaaki Aoki², and Michiko Ueda³

¹Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan

²Faculty of Education, Kagawa University, Takamatsu, Japan

³Faculty of Political Science and Economics, Waseda University, Tokyo, Japan

*kawamoto.tatsuro@aist.go.jp

ABSTRACT

The COVID-19 pandemic is an unprecedented public health crisis with broad social and economic consequences. We conducted four surveys between April and August 2020 using the graph-based open-ended survey (GOS) framework, and investigated the most pressing concerns and issues for the general public in Japan. The GOS framework is a hybrid of the two traditional survey frameworks that allows respondents to post their opinions in a free-format style, which can subsequently serve as one of the choice items for other respondents, just as in a multiple-choice survey. As a result, this framework generates an opinion graph that relates opinions and respondents. We can also construct annotated opinion graphs to achieve a higher resolution. By clustering the annotated opinion graphs, we revealed the characteristic evolution of the response patterns as well as the interconnectedness and multi-faceted nature of opinions.

Introduction

The COVID-19 pandemic is an unprecedented event with a myriad of consequences. Without a doubt, it is one of the most serious public health crises in recent history. However, it is more than a public health challenge. Various measures that were introduced to prevent the spread of the virus have caused major disruptions in economic activities and social life. Reduced business activities and associated job losses, school closures, movement restrictions, and social distancing have affected multiple aspects of our lives, presenting new challenges that every single member of society has to face in addition to the disease itself. Thus, the COVID-19 pandemic has given rise to multi-faceted issues and concerns for many individuals, ranging from the consequences of the infection itself to the economic and social ramifications of the pandemic. Throughout this paper, we refer to these concerns and issues as opinions.

Using the graph-based open-ended survey (GOS) framework¹, this paper seeks to understand such multi-faceted opinions expressed by the general population in Japan between April and August 2020. The GOS framework is a hybrid of the two traditional survey frameworks: multiple choice and free format. In a multiple-choice survey, respondents are presented with a list of items from which they could choose one or more that were most applicable to them. While this is a widely used framework owing to its ease of implementation, the potential answers must be anticipated and compiled as a list by those conducting the survey. Thus, the multiple-choice survey may not be the most appropriate framework to use if the purpose is to gauge complex and multi-faceted opinions, such as people's concerns during an unprecedented event like the COVID-19 pandemic. Their concerns and opinions may not be clear a priori because of the complex and rapidly developing nature of the event. The second framework is the free-format survey in which the respondents post their opinions as texts. However, analysing such texts can be arbitrary to a certain extent and challenging in a large-scale survey, including this study.

In the GOS framework, respondents can post their opinions in a free-format style, which can subsequently serve as one of the choice items for other respondents, just as in the multiple-choice survey. Because survey respondents can select others' opinions that they agree with, the GOS framework constitutes a graph (or network) of responses at the end. This graph makes statistical evaluation of the free-format responses feasible, while preserving the diversity of respondents' opinions that increases spontaneously as respondents express their own opinions.

A major contribution of this paper is to reveal the most pressing issues and concerns of the general public through the opinions expressed by the respondents. In addition, by conducting the same survey using the GOS framework several times, we were able to capture the evolution of those opinions.

Our analysis shows that the variety of major opinions rapidly changed during the study period, and that several opinions were strongly related. Traditional survey methods are less suitable to capture such characteristic behaviours. These results indicate that studying opinions under a rapidly changing situation, such as the COVID-19 pandemic, is indeed an ideal

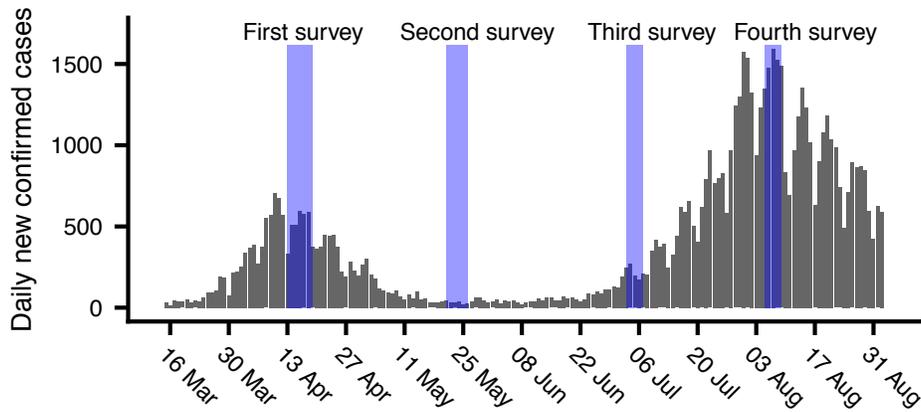


Figure 1. Survey periods and the daily new confirmed cases of COVID-19 in Japan.

application of the GOS framework.

Our study period corresponded to the middle of the first wave of cases in Japan (April 2020), the period in which the number of cases temporarily diminished (May 2020), and the beginning and the middle of the second wave (July and August 2020, respectively), which was larger than the first one in terms of the number of cases, but resulted in fewer deaths (see Fig. 1 for the evolution of the confirmed cases). The first case of COVID-19 in Japan was reported on January 16. The Japanese government declared a state of emergency in major metropolitan areas on April 7, and expanded the coverage to the entire country on April 16. The Japanese government implemented neither a lockdown nor strict movement restrictions. Non-essential businesses were urged to either temporarily close or operate for reduced hours, but non-compliance was not penalised. The state of emergency was lifted on May 25. As of August 31, the total number of positive cases was 67,077, and the number of COVID-19-related deaths was 1,278, which amounted to 10.15 deaths per one million people².

Methods

We conducted four rounds of an online survey of the Japanese adult population aged 20–59 between April and August 2020. The respondents were asked about their most pressing concerns. Specifically, the question asked (in Japanese) was, “What is the concern and the difficulty that you face in your daily life and economic activities? Please provide concrete and succinct answers.” This study was approved by the Ethics Review Committee on Human Research of Waseda University (approval number: 2020-050). Further details on data collection can be found in Supplementary Information.

Graph-based open-ended survey

The results of the GOS were used to generate a bipartite graph consisting of two sets of vertices representing the opinions and respondents, respectively. Each edge connects a pair of opinions and respondent vertices, and indicates that a respondent supports the opinion. We refer to this graph as an *opinion graph* (see Supplementary Information for a comment on the definition of the opinion graph).

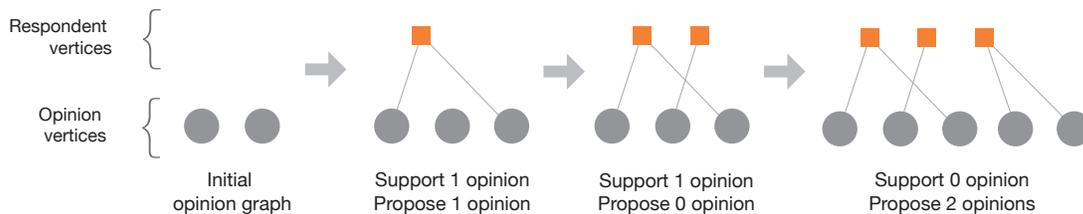


Figure 2. Example of the generation process of an opinion graph.

The generation process of the opinion graph is illustrated in Fig. 2. At the beginning of the survey, we prepared a set of opinion vertices (which can be an empty set) as the initial opinions (choices). The description of the initial opinions is

provided in Supplementary Information. When a respondent responds to a question, several opinions are sampled uniformly and randomly, and presented to the respondent; the minimum number of samples is eight, while respondents can opt to refer to up to 24 opinions. Whenever a respondent supports a presented opinion(s), the GOS generates a respondent vertex connected to the selected opinion vertices. When a respondent expresses new opinions, the GOS generates corresponding opinion vertices connected to the newly added respondent vertex. The numbers of opinion vertices and respondent vertices increase simultaneously as the GOS is carried out. The above procedure constitutes an opinion graph.

We expect a set of similar opinion vertices to be connected via respondent vertices. Similarly, we expect the respondent vertices to be classified into groups, in which the vertices have a similar response pattern. The classification of opinions and respondents is realised through clustering of an opinion graph, which we refer to as the opinion groups and respondent groups, respectively.

Clustering of the opinion graphs

A number of clustering algorithms have been proposed for graphs and bipartite graphs in the literature³⁻⁷. Here, we used the Markov chain Monte Carlo method, which was implemented in a software called *graph-tool*⁸ that performs a Bayesian inference of groups under the assumption of a stochastic block model⁹. The algorithm identifies a statistically significant group structure and estimates the number of groups in a non-parametric manner. All the clustering results in this study are inferences of unnested (i.e. non-hierarchical) stochastic block models. We confirmed that, while the nested variant does not raise the resolution considerably, it often subdivides a large opinion group into smaller sets, which is apparently an overfit.

Annotated opinion graphs

In addition to the clustering of raw opinion graphs, we also considered annotated opinion graphs to achieve a higher resolution by adding more information externally through annotation.

Based on the collected opinions, we asked three annotators to classify the opinions into 10 groups: (1) *infection risk*, (2) *social pressure & future prospect*, (3) *financial issues*, (4) *(restriction of) travel*, (5) *government policies*, (6) *mask (shortage)*, (7) *mask (discomfort)*, (8) *other issues*, (9) *no concerns*, and (10) *invalid responses*. The opinions that may not exclusively belong to one of these groups are left unannotated. The last item refers to opinions that are not directly related to the question that was asked. More detailed descriptions of these groups are provided in Supplementary Information.

The annotators classified opinions independently, and we distinguished annotations according to different annotators. The maximum number of annotation labels was 30 for each survey. We denoted the total number of annotation labels as K , which can be less than 30 if some labels are not used.

As shown in Fig. 3, based on annotators' decisions, we constructed a K -dimensional vector that represents a prior probability distribution of the group assignment for each opinion vertex. The k th element is η ($\gg 0$) if the vertex has the k th annotation label; otherwise, ε ($0 < \varepsilon \ll \eta$), where ε and η are determined to normalise the vector to unity. Therefore, the annotated vertices have biased prior distributions, whereas all unannotated vertices have a uniform prior distribution.

These prior distributions were incorporated into the Bayesian inference using the *bfield* parameter in *graph-tool*. The set of opinion vertices with similar prior distributions tends to be classified into the same group. Because we used a non-parametric inference method, the algorithm eventually identifies the parsimonious number of opinion groups. Alternatively, one could use an approach discussed in Ref.¹⁰ to incorporate annotations in the inference.

Results

Response patterns

The number of respondents who participated in the four rounds of the survey were 2103, 1516, 1729, and 1659, respectively. In each round, 191 (9.1%), 127 (8.4%), 119 (6.9%), and 117 (7.1%) respondents expressed their own opinions, while the remaining respondents selected opinion(s) that were applicable to them from a list presented to them.

Figure 4 shows the clustering results of the raw opinion graphs. For the first and fourth surveys, no group structure was identified. In contrast, for the second and third surveys, opinion graphs were partitioned into two groups. In both cases, the vertices in the smaller respondent group are highly connected to the opinion vertices that can be coded as *no concerns*, while the respondents in the other group supported the opinions on various issues, such as infection risk and financial concerns.

The clustering in Fig. 4 is the most agnostic result based purely on the collected responses. However, their resolution is insufficient, as the raw opinion graph is only classified into a few groups or not classified at all. Figure 5 shows the response patterns of respondents that were obtained through the clustering of annotated opinion graphs. The (i, j) element is the group-wise propensity (normalised in each column) towards opinion group i for respondent group j . More precisely, this is the fraction of the number of edges between groups i and j . Although the number of identified opinion groups in each survey was nine (first survey), six (second survey), and seven (third and fourth surveys), we added empty opinion groups in Fig. 5 to obtain the same number of rows in all surveys. In contrast to the results reported in Fig. 4, where we observed only one group

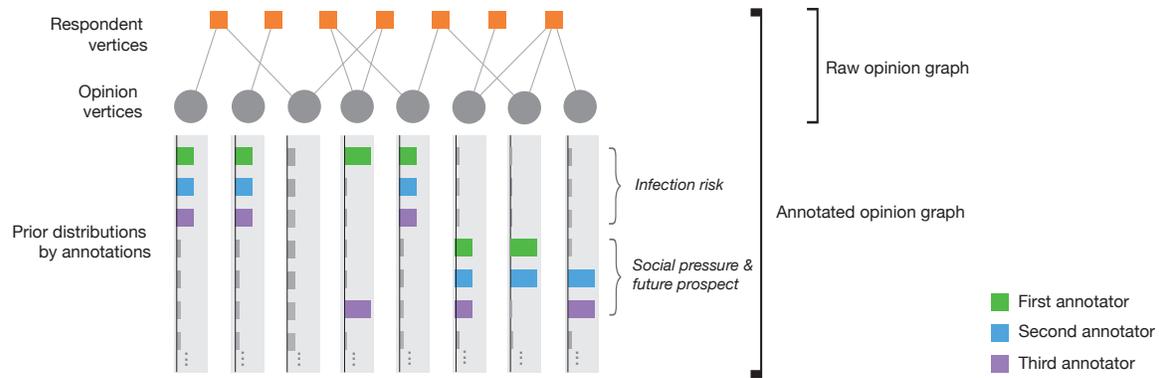


Figure 3. Schematic of an annotated opinion graph. A raw opinion graph is shown at the top, and the prior distribution for each vertex is shown as a bar plot at the bottom.

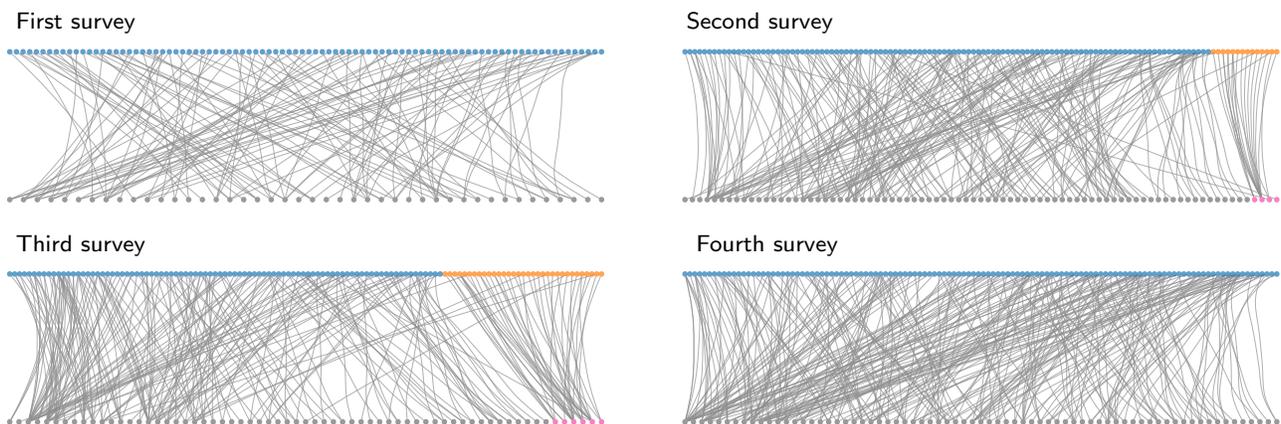


Figure 4. Clustering results of the opinion graphs. These graphs only show randomly sampled connected subgraphs instead of the entire dataset, for illustration purposes. The opinion vertices are aligned at the bottom, while the respondent vertices are aligned at the top. Both types of vertices are sorted based on the inferred group assignments. The colour of each vertex represents the group assignment. The pink vertices are the set of opinions that can be coded as *no concerns*.

of respondents for the first and fourth surveys, we can now observe two groups, analogous to the results of the second and third surveys.

To further reveal the details about response patterns, we show the palette diagrams^{11,12} in Fig. 6. The palette diagram is essentially a stream plot, which is a stack plot with varying origin axes. The vertical axis represents the normalised propensity for each opinion group that a respondent supports. The set of normalised propensities (i.e. vertical thickness is unity) for all respondents are optimally aligned horizontally so that the global distribution of the response pattern can be better understood. The palette diagrams in Fig. 6 show the microscopic response patterns of respondents; the tables in Fig. 5 show how these patterns can be summarised at a macroscopic level.

The results presented in these two figures suggest that there was a sizable block of respondents who did not express any specific concerns throughout the study period. These respondents were identified as a distinct group, relatively isolated from others, through graph clustering. The evolution of the number of respondents in this group is negatively correlated with the number of daily cases, which peaked during our fourth survey, as expected. As indicated in Fig. 6, these respondents were less likely to support other opinion groups.

For the other opinion groups, it should be noted that many of these groups, including *government policies* and *travel*, emerged spontaneously from the opinions expressed by respondents themselves, as they were not included in the initial

	First survey		Second survey		Third survey		Fourth survey	
	A	B	A	B	A	B	A	B
Infection risk	17.3	4.4	15.8	1.2	7.9	0.8	44.8	4.3
Social pressure & future prospect	22.1	9.3	26.9	6.1	29.3	5.5	22.3	2.7
Financial issues	21.7	6.6	31.1	4.3	27.4	3.6	20.1	8.2
Travel	6.2	0.0	18.7	4.0	20.9	3.0	9.4	4.3
Government policies	15.2	0.5	0.0	0.0	0.0	0.0	0.0	0.0
Mask (shortage)	4.5	1.6	0.0	0.0	0.0	0.0	0.0	0.0
Mask (discomfort)	0.0	0.0	0.0	0.0	7.2	1.7	1.8	7.1
Other issues	10.2	2.7	6.5	0.9	6.5	0.0	1.1	0.0
No concerns	0.4	62.1	0.9	83.5	0.8	85.4	0.6	73.4
Invalid responses	2.3	12.6	0.0	0.0	0.0	0.0	0.0	0.0

Figure 5. Response patterns of the four surveys. The rows indicate opinion groups and the columns indicate respondent groups. In each table, the two respondent groups are indicated as groups A and B.

opinions displayed at the beginning of the survey. On the other hand, the opinions belonging to *infection risk*, *social pressure & future prospect*, and *financial issues* were included in the initial opinions. It should also be noted, however, that the initial opinions become less likely to be displayed to the respondents as more respondents post their own opinions. Therefore, the initial opinions are unlikely to have a significant influence on the result.

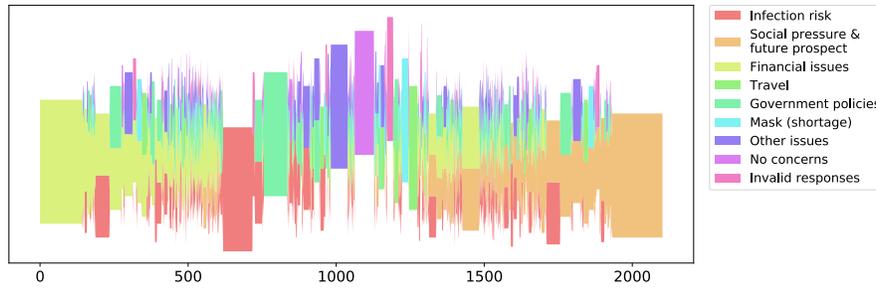
In all four surveys, the opinions of respondent group A (shown in Fig. 5) tended to focus on the following opinion groups: *infection risk*, *social pressure & future prospect*, *financial issues*, and *travel*. The size of these opinion groups varied considerably across surveys. For example, *travel*, which covers issues associated with difficulties with traveling and stay-at-home “requests,” was not a major issue in April, as there were no formal movement restrictions in Japan. It became more prevalent in the second and third surveys (May and July) when the number of cases diminished temporarily. By August, however, it no longer constituted a major opinion as the infection risk again became the dominant concern with the rising number of cases.

In the first survey (April 2020), our respondents also tended to complain about the government’s handling of the pandemic (*government policies*) as well as express concerns regarding the status of hospitals that were overwhelmed by the rapidly rising number of patients; this was captured in the *other issues* group. However, as indicated in Fig. 5, the fraction of such concerns quickly decreased in subsequent surveys as the focus of respondents shifted to other issues, such as financial concerns.

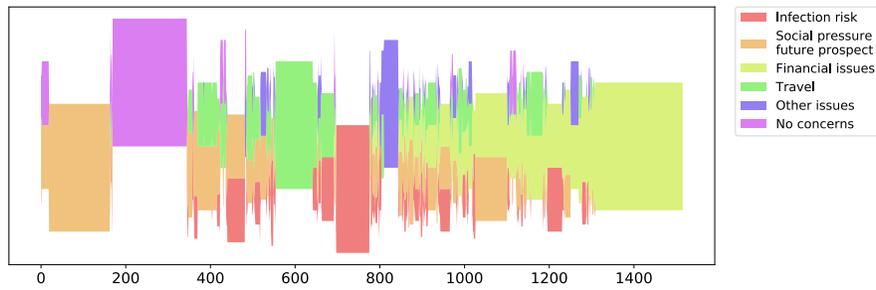
It is noteworthy that *social pressure & future prospect* always constituted one of the major opinion groups throughout our study period. This group mainly refers to the negative *social* consequences of contracting the disease, rather than the physical consequences. Many respondents expressed fears that they might be ostracised by others if they contract the disease; thus, this opinion group largely represents the social pressure that respondents felt even before testing positive for COVID-19. Similarly, *financial issues* was another group that appeared as a sizable block in all of the surveys, indicating that the economic consequences of the pandemic were among the major concerns of respondents.

Fig. 6 reveals that many respondents expressed multiple concerns and opinions, instead of selecting a single issue as their most pressing concern. While *financial issues*, *social pressure & future prospect*, and *no concerns* tended to appear on their own, other types of issues were more likely to appear in combination with other opinion groups, suggesting that many of our

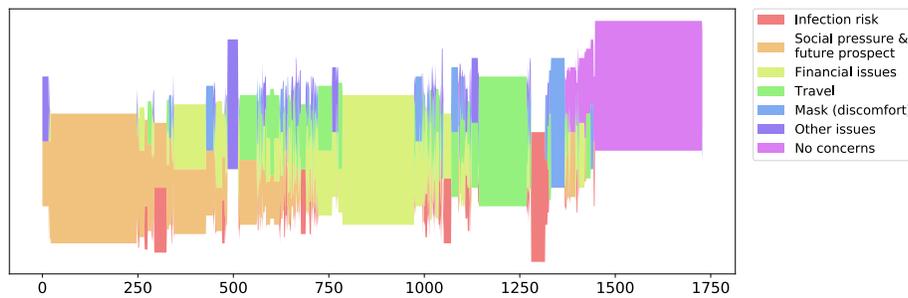
First survey



Second survey



Third survey



Fourth survey

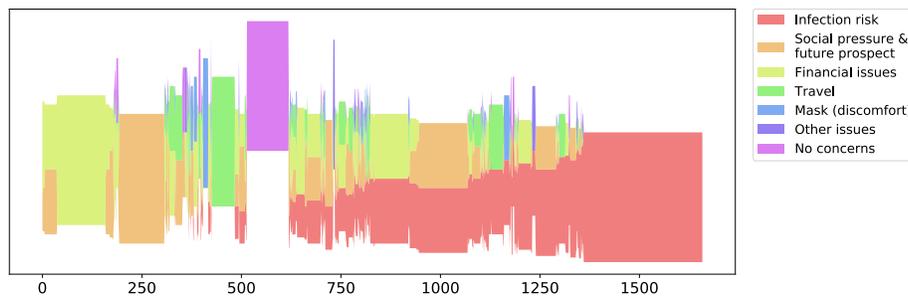


Figure 6. Palette diagrams of the four surveys.

respondents faced multiple issues and concerns during the pandemic.

Finally, our results show that infection risk became the dominant concern by the time we conducted the fourth survey, which coincided with the peak period of the second wave of cases (August). As shown in Fig. 6, although the *infection risk* group was one of many other opinions for most respondents until the third survey, it became a prominent group in the fourth survey. The number of respondents who specified infection risks as their only concern was highest in the fourth survey, which

suggests that infection risk became the dominant issue for a large number of respondents.

Discussion

In this paper, we surveyed the general adult population of Japan four times using the GOS framework to investigate the issues and concerns of people in the midst of the COVID-19 pandemic. We classified the opinions of respondents by aggregating the response patterns and annotations through graph clustering. As a result, we revealed the characteristic evolution of the response patterns, particularly on infection risk.

The survey results indicated that the most pressing concerns of the general public changed through the various phases of the pandemic. Many of our respondents also had multiple issues and concerns. Such a fine evolution of people's opinions as well as the multi-faceted nature of these opinions may not be adequately captured by the traditional multiple-choice survey method, which required us to predict the opinions to be included in a list of potential choices. Presenting survey respondents with a pre-determined set of responses can be particularly challenging when the situation changes on a daily basis, as in the case of the current pandemic. Therefore, this is an ideal situation for an open-ended survey, in which the group labels are determined a posteriori. As a hybrid of the two traditional survey methods, the GOS framework takes advantage of both methods and successfully reveals the multi-faceted and rapidly changing features of people's opinions.

Finally, let us discuss the role of annotation in the clustering of opinion graphs. If every respondent supports only one issue or concern, then the resulting opinion graph would have a relatively simple group structure, in which each respondent group would be densely connected to the corresponding opinion group. However, as observed in Figs. 5 and 6, the opinion graphs in the present survey have more complex group structures because a significant number of respondents support multiple opinions with different meanings. The annotations were needed to achieve a higher resolution because a graph with a complex group structure is difficult to distinguish from a uniformly random graph, that is, pure noise.

Although the annotations made the group structures clear, it should also be noted that clustering annotated opinion graphs is far from trivial. First, because the annotators' classifications were performed independently, just as the responses of the surveys, the decisions of the annotators are subjective and do not coincide precisely (see Supplementary Figure S1 for the consistency among the annotators). In addition, not all opinion vertices were annotated, and there were no annotations for respondent vertices. Considering the fact that the responses to open-ended questions are often classified (i.e. annotated) in a completely manual manner, our statistical inference approach that brings together all the information from respondents and annotators is a considerable improvement.

Acknowledgements

This study was financially supported by JSPS Grants-in-Aid for Scientific Research Grant Nos. 18K18604 (T.A. and T.K.) and 20H01584 (M.U.), and Research and Regional Cooperation for Crisis Management Shikoku (T.A.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Kawamoto, T. & Aoki, T. Democratic classification of free-format survey responses with a network-based framework. *Nature Machine Intelligence* **1**, 322–327 (2019).
2. Ministry of Health Labour and Welfare. On the number of COVID-19 cases. (2020). URL https://www.mhlw.go.jp/stf/covid-19/kokunainohasseijoukyou.html#h2_1.
3. Dhillon, I. S. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '01*, 269–274 (Association for Computing Machinery, New York, NY, USA, 2001). URL <https://doi.org/10.1145/502512.502550>.
4. Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788–791 (1999).
5. Kluger, Y., Basri, R., Chang, J. T. & Gerstein, M. Spectral biclustering of microarray data: coclustering genes and conditions. *Genome research* **13**, 703–716 (2003).
6. Larremore, D. B., Clauset, A. & Jacobs, A. Z. Efficiently inferring community structure in bipartite networks. *Phys. Rev. E* **90**, 012805 (2014).
7. Gerlach, M., Peixoto, T. P. & Altmann, E. G. A network approach to topic models. *Science Advances* **4** (2018).
8. Peixoto, T. P. The graph-tool python library. *figshare* (2014). URL http://figshare.com/articles/graph_tool/1164194.

9. Peixoto, T. P. Bayesian stochastic blockmodeling. *"Advances in Network Clustering and Blockmodeling"*, edited by P. Doreian, V. Batagelj, A. Ferligoj, (Wiley, New York, 2019) (2017).
10. Hric, D., Peixoto, T. P. & Fortunato, S. Network structure, metadata, and the prediction of missing nodes and annotations. *Phys. Rev. X* **6**, 031038 (2016).
11. Noguchi, C. & Kawamoto, T. Evaluating network partitions through visualization (2019). [1906.00699](#).
12. Noguchi, C. & Kawamoto, T. Palette diagram: A python package for visualization of collective categorical data (2020). [2011.01934](#).

Figure legends

1. Survey periods and the daily new confirmed cases of COVID-19 in Japan.
2. Example of the generation process of an opinion graph.
3. Schematic of an annotated opinion graph. A raw opinion graph is shown at the top, and the prior distribution for each vertex is shown as a bar plot at the bottom.
4. Clustering results of the opinion graphs. These graphs only show randomly sampled connected subgraphs instead of the entire dataset, for illustration purposes. The opinion vertices are aligned at the bottom, while the respondent vertices are aligned at the top. Both types of vertices are sorted based on the inferred group assignments. The colour of each vertex represents the group assignment. The pink vertices are the set of opinions that can be coded as *no concerns*.
5. Response patterns of the four surveys. The rows indicate opinion groups and the columns indicate respondent groups. In each table, the two respondent groups are indicated as groups A and B.
6. Palette diagrams of the four surveys.

Author contributions statement

All authors contributed to the design of the work and interpretation of the results. T.K. created the software used in the work and analysed the data. M.U. conducted the data collection. T.K. and M.U. wrote the manuscript. All authors reviewed the manuscript.

Additional information

The authors declare no competing interests.

Figures

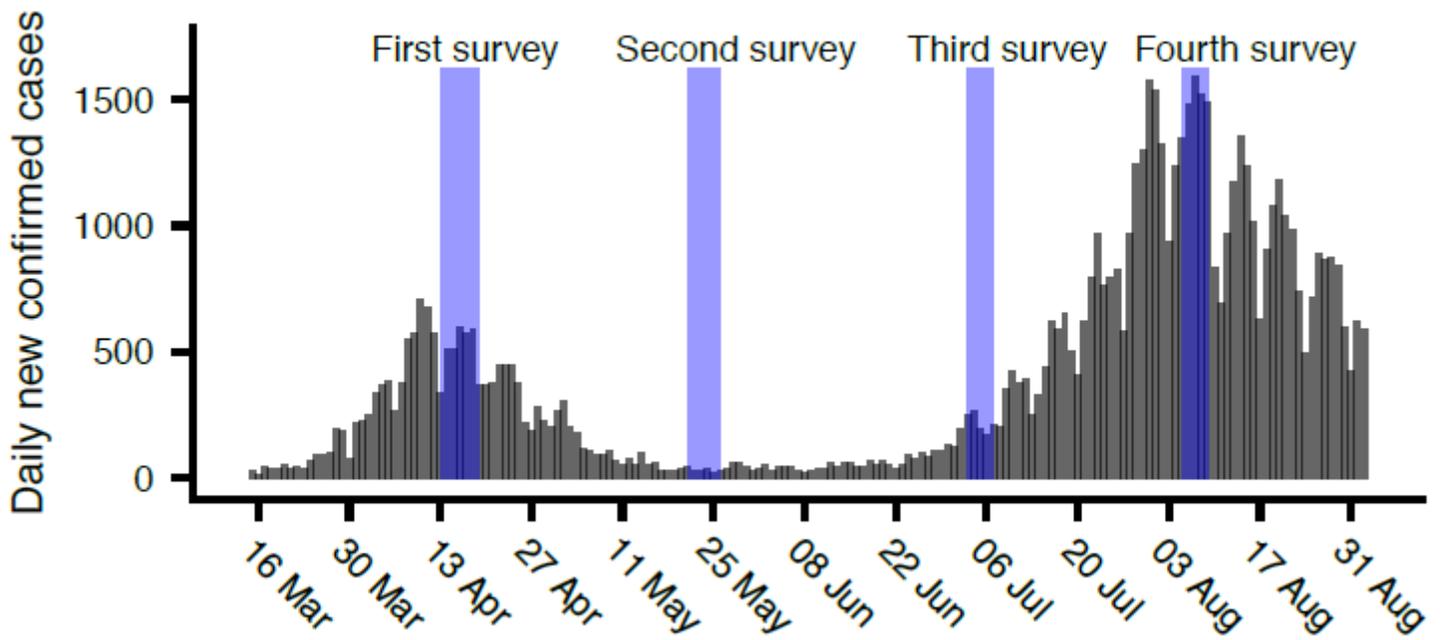


Figure 1

Survey periods and the daily new confirmed cases of COVID-19 in Japan.

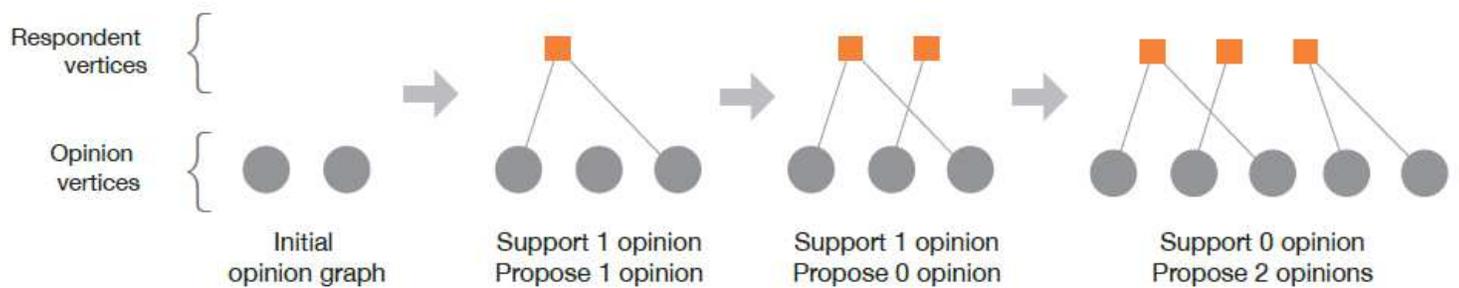


Figure 2

Example of the generation process of an opinion graph.

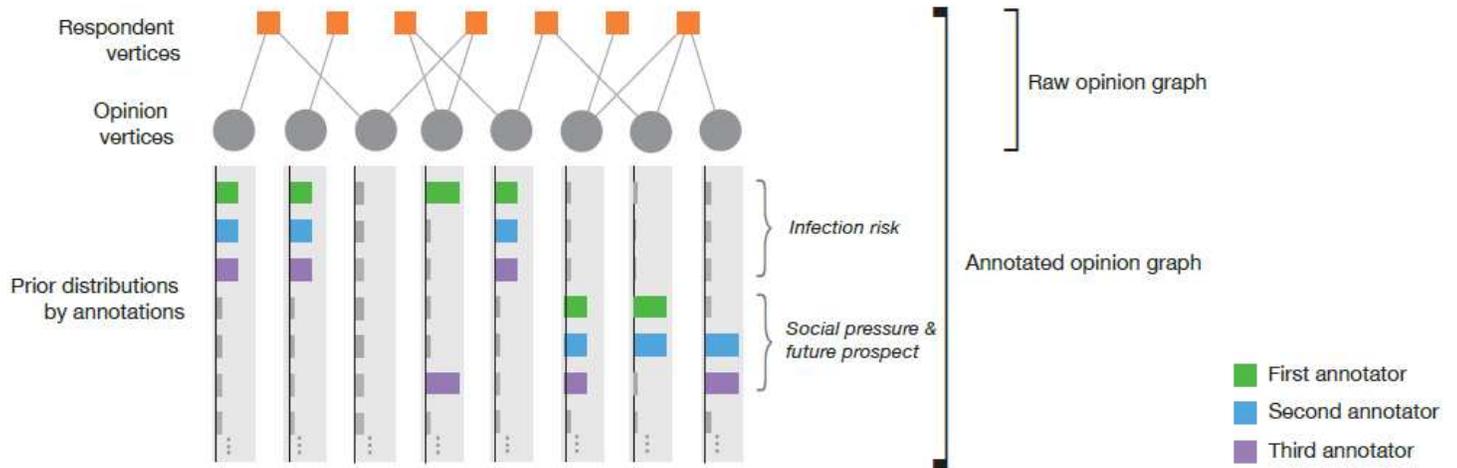


Figure 3

Schematic of an annotated opinion graph. A raw opinion graph is shown at the top, and the prior distribution for each vertex is shown as a bar plot at the bottom.

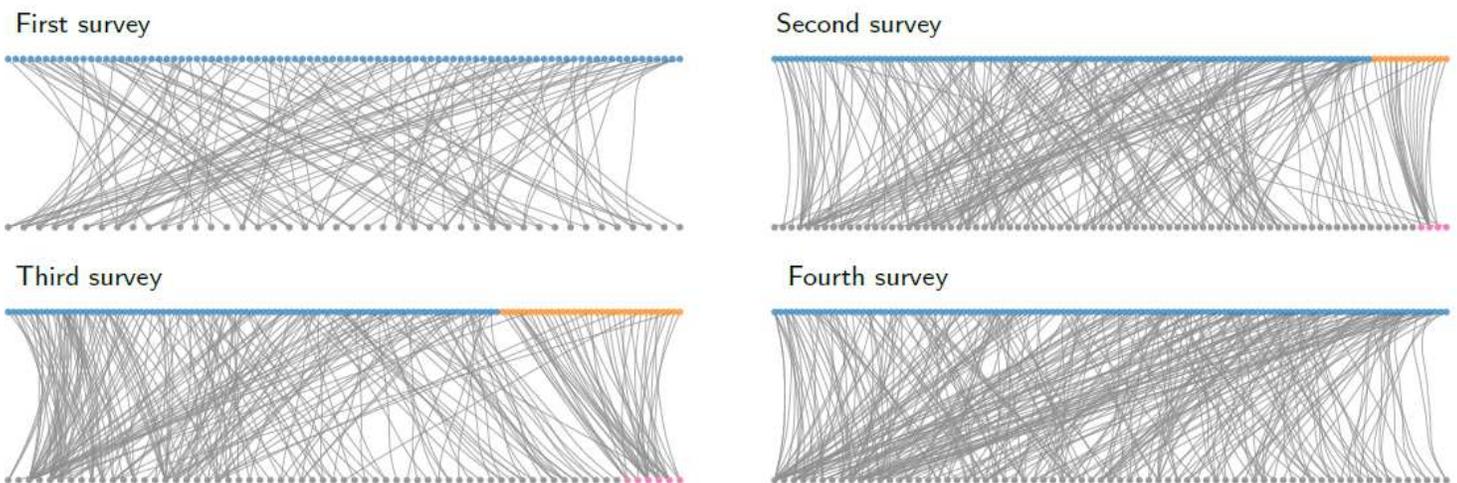


Figure 4

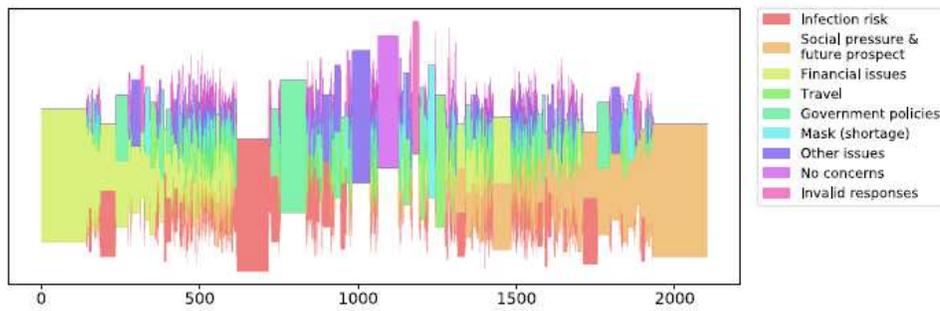
Clustering results of the opinion graphs. These graphs only show randomly sampled connected subgraphs instead of the entire dataset, for illustration purposes. The opinion vertices are aligned at the bottom, while the respondent vertices are aligned at the top. Both types of vertices are sorted based on the inferred group assignments. The colour of each vertex represents the group assignment. The pink vertices are the set of opinions that can be coded as no concerns.

	First survey		Second survey		Third survey		Fourth survey	
Infection risk	17.3	4.4	15.8	1.2	7.9	0.8	44.8	4.3
Social pressure & future prospect	22.1	9.3	26.9	6.1	29.3	5.5	22.3	2.7
Financial issues	21.7	6.6	31.1	4.3	27.4	3.6	20.1	8.2
Travel	6.2	0.0	18.7	4.0	20.9	3.0	9.4	4.3
Government policies	15.2	0.5	0.0	0.0	0.0	0.0	0.0	0.0
Mask (shortage)	4.5	1.6	0.0	0.0	0.0	0.0	0.0	0.0
Mask (discomfort)	0.0	0.0	0.0	0.0	7.2	1.7	1.8	7.1
Other issues	10.2	2.7	6.5	0.9	6.5	0.0	1.1	0.0
No concerns	0.4	62.1	0.9	83.5	0.8	85.4	0.6	73.4
Invalid responses	2.3	12.6	0.0	0.0	0.0	0.0	0.0	0.0
	A	B	A	B	A	B	A	B

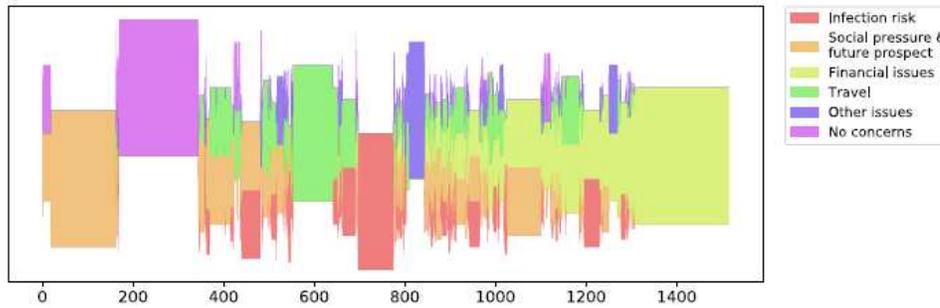
Figure 5

Response patterns of the four surveys. The rows indicate opinion groups and the columns indicate respondent groups. In each table, the two respondent groups are indicated as groups A and B.

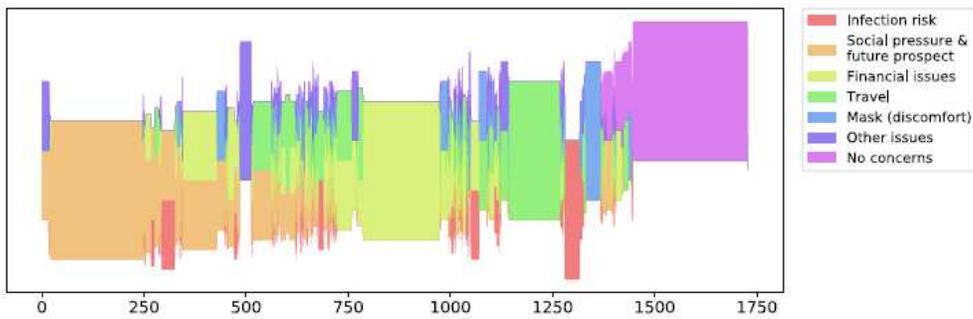
First survey



Second survey



Third survey



Fourth survey

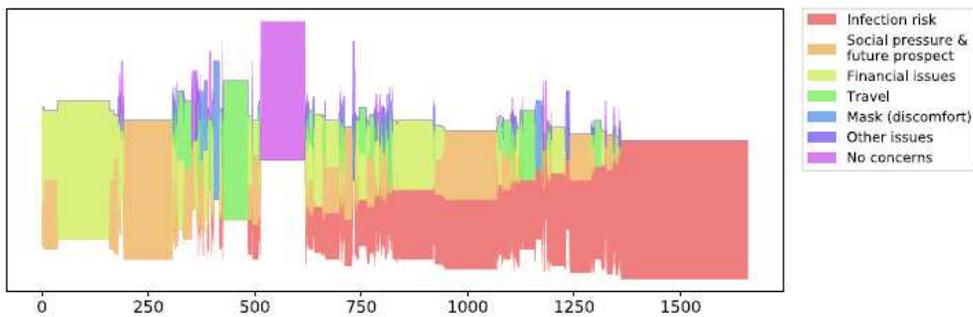


Figure 6

Palette diagrams of the four surveys.

Supplementary Files

This is a list of supplementary files associated with this preprint. [Click to download.](#)

- [covid19suppl.pdf](#)