Democratic classification of free-format survey responses with a network-based framework

Tatsuro Kawamoto1* and Takaaki Aoki2

Social surveys have been widely used as a method of obtaining public opinion. Sometimes, it is more ideal to collect opinions by presenting questions in free-response formats than in multiple-choice formats. Despite their advantages, free-response questions are rarely used in practice because they usually require manual analysis. Therefore, classification of free-format texts can present a formidable task in large-scale surveys and can be influenced by the interpretation of analysts. In this study, we propose a network-based survey framework in which responses are automatically classified in a statistically principled manner. This can be achieved because, in addition to the text, similarities among responses are also assessed by each respondent. We demonstrate our approach using a poll on the 2016 US presidential election and a survey taken by graduates of a particular university. The proposed approach helps analysts interpret the underlying semantics of responses in large-scale surveys.

A survey is a method of collecting ideas, opinions and other information from people. They are widely used in many contexts, including social research, patient feedback for medical care, customer surveys of products and services, and opinion polls for elections and government policies. Survey questions are typically presented as multiple-choice questions (that is, closed-ended questions) or free-response questions (that is, open-ended questions).

A question like ‘Are you a scientist?’ can be answered with ‘yes’ or ‘no’, and therefore can be easily considered in a multiple-choice format. However, some questions must be asked in a free-response format1,2. For example, a question such as ‘Why did you become a scientist?’ cannot be easily considered in a multiple-choice format. To create a list of multiple choices, an analyst must predict several representative responses. This task can be quite challenging because possible major responses may be missing from the list3–5. In such a situation, the outcome of the survey can be significantly influenced by the decision of an analyst. Surveys with free-response questions are free from such difficulties and limitations, directly revealing the logic and thought processes of respondents3–5.

Despite their advantages, surveys rarely include free-response questions because they require manual analysis. To perform statistical analyses, the various opinions collected in the form of text data must be classified into a manageable number of categories without introducing any loss in their intrinsic variety and reflecting underlying semantic relations among responses. This is termed coding7 and is considered a significant scientific issue. This task is subjective and time-consuming even for a trained analyst3,8,9 and is particularly complicated in the case of large-scale surveys. In fact, in the literature10, the limitations of free-response surveys are viewed as follows: ‘... the lack of standardization means that responses are likely to be more difficult to compare statistically, and this limits attempts to draw general conclusions from the study.’

Here we propose a network-based (or graph-based) approach that helps analysts perform coding by automatically classifying responses in a statistically principled manner. This is illustrated in Fig. 1a. In Step 1, for a given free-response question, the respondent inputs his/her own response. In Step 2, a small subset of the responses input by other respondents is randomly drawn and presented to the respondent. The respondent then selects those responses that he/she feels are similar to his/her own response, if any. This is similar to the manner in which ‘likes’ are input in social network services.

After the above procedure has been executed by many respondents, we obtain a dataset that we call the opinion graph (schematically depicted in the far left-hand side of Fig. 1b). The vertices denote the responses and the edges encode the ‘positive’ (chosen as similar by one of the respondents) or ‘negative’ (presented to a respondent, but not chosen) relations among them. In other words, this is a network in which the measurement process is taken into account11–13. When there is a consensus among the respondents, the resulting network exhibits a modular structure in which vertices within a group are likely to be connected via positive edges (an assortative structure), and vertices in different groups are likely to be connected via negative edges (a disassortative structure). We statistically infer such a modular structure using a graph clustering algorithm, as described in the Materials and Methods. Therefore, in the method that we propose, classification is performed by an algorithm that inputs respondent choices rather than by human analysts. To assign a label to a group, we extract keywords from several typical opinions from that group (middle, Fig. 1b). We can then perform categorical data analysis for those opinion groups analogously to the case of a survey with multiple-choice questions (far right-hand side, Fig. 1b).

Estimating the number of groups is also an important problem. For example, the network may not exhibit a clear modular structure. This occurs when the opinions do not exhibit any obvious separation of ideas. An advantage of our approach is that we can assess the statistical significance among opinion groups.

To demonstrate our network-based framework, we implemented an online-survey system. We discuss the results we obtained from two surveys in the following.

Results

Poll on the 2016 US presidential election. The first application of our framework is a poll on the presidential election of the United States in 2016. A total of 117 responses were collected.

1Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Koto-ku, Tokyo, Japan. 2Faculty of Education, Kagawa University, Takamatsu, Japan. *e-mail: kawamoto.tatsuro@gmail.com
from 1 October to 8 November (election day) 2016. Although the respondents were not restricted to any particular group, a majority of opinions were collected from people who visited the University of Nevada, Las Vegas on 18 and 19 October—the day before and the day of the final presidential debate. (Note that the University of Nevada, Las Vegas was the venue of this debate.)

The question we asked was '#NeverHillary or #NeverTrump?' As shown in Fig. 2a, we could extract not only the groups of opinions expressed by the supporters of Donald Trump (group 1) and Hillary Clinton (group 2) but also a group of opinions expressed by those who supported neither (group 3). In Fig. 2a, red and blue edges denote positive and negative edges, respectively. Here, we use a colour gradient ranging between red (group 1), blue (group 2) and white (group 3) to express estimated assignment probabilities ($p_1$, $p_2$, $p_3$) with respect to those three groups. The colour indicates the proxim- ity to the three groups; that is, more red means closer to the red group, more blue means closer to the blue group and more white (a paler shade) means closer to the white group. In addition, vertices that are selectively localized within a single group (max $p_i \geq 0.9$) are represented by deep red, deep blue and pure white, and have black borders; these vertices can be regarded as the typical opinions of a group.

Note that whether a respondent supports either of the candidates may not be expressed explicitly in a response. For example, the dataset contains responses such as ‘We need insurance’ and ‘Gary Johnson’. Several such responses can be appropriately classified based on prior knowledge of the candidates. However, the distinction between Clinton supporters and people who support neither are often unclear, whereas Trump supporters can usually be easily distinguished. In such a situation, classification would be extremely difficult if we were to classify opinions using only natural language processing, unless sufficient domain knowledge of the subjects of the survey were incorporated. Instead, what we should really believe in is the decisions made by each individual, that is, the opinion graph.

The alluvial diagram in Fig. 2b depicts flows of group assignments as we increase the number of groups. A bundle represents a set of responses that are classified as the same group, and the diagram depicts the flow of group assignments as the number of groups changes. An alluvial diagram is an extension of a Sankey diagram. The diagram expresses the significance of group assignments with colour shading: a bundle of vertices that are selectively localized in a single group with max $p_i \geq 0.9$ is represented by a deep colour,
and others are represented by paler colours. The groups that can be interpreted as Trump supporters, Clinton supporters and neither correspond to red, blue and grey, respectively.

Alluvial diagrams provide a useful tool to visually determine the appropriate number of groups (that is, the model selection)\(^{15}\) to be used for a given survey. In the present case, it can be observed that groups 2 and 3 form a single group in the case of bipartition, and they split in the case of tripartition. When the network is partitioned into four groups, the responses in deep colours no longer split. Moreover, a significant fraction of vertices in different groups merge and, as a result, the hierarchical structure no longer exists. For these reasons, we selected the partition with three groups as the final result. Further details of the assessment of the proper number of groups are provided in the Supplementary Information.

**Social survey of the Faculty of Education in a particular university.** As the second application of our framework, we conducted a survey focusing on graduates of the Faculty of Education of a particular university. The questions we asked are the following: 'What is your career?' (Q1); 'What reason made you choose your current job?' (Q2); 'What is the most valuable experience that you will take away with you from your time at the university?' (Q3).

We announced the survey via postal mail to people who graduated from our university. The questions we asked are the following: 'What is your career?' (Q1); 'What reason made you choose your current job?' (Q2); 'What is the most valuable experience that you will take away with you from your time at the university?' (Q3).

We conducted a survey focusing on graduates of the Faculty of Education of a particular university. The questions we asked are the following: 'What is your career?' (Q1); 'What reason made you choose your current job?' (Q2); 'What is the most valuable experience that you will take away with you from your time at the university?' (Q3).

We announced the survey via postal mail to people who graduated from our university. The questions we asked are the following: 'What is your career?' (Q1); 'What reason made you choose your current job?' (Q2); 'What is the most valuable experience that you will take away with you from your time at the university?' (Q3). We received 258 responses. (The response rate was \(\sim 10\%\).)

Figure 3a presents the opinion graph for Q1, 'What is your career?' The network is plotted in the same manner as that in Fig. 2a. To determine the appropriate number of groups, we again used the alluvial diagram, as in the case of Fig. 2b, including the criteria based on cross-validation estimates of edge-prediction errors (see Supplementary Information for details).

Figure 3b depicts the structures that the algorithm learned with respect to the positive and negative edges. It is observed that except for group 4, the positive edges exhibit an assortative structure, as intended. However, it must be noted that negative edges exhibit a structure that is more complicated than a disassortative structure.

Therefore, it is evident from our result that the negative edges must not be strictly interpreted as dissimilar relationships.

Once the responses have been classified, we can analyse the categorical data using standard tools. The bar chart for each question is presented in Fig. 3c. Among the four groups that emerged from the responses to Q1, 'teachers' (group 1) has a clear majority; over half of the respondents are assigned to this group. This is reasonable, because all respondents are graduates of the Faculty of Education. The other three groups include 'civil servants' (group 2), 'office workers' in private firms (group 3) and 'change of career' (group 4). (It must be noted here that only group 4 is a carrier path, instead of a name of the job.) These groups reflect a similarity that the graduates recognize, which is distinct from simple similarities between text or the standard classification. The respondents would have responded based on their own experiences, the education they have received, the career paths of their friends, parents, neighbours and so on.

Two groups emerged from the responses to Q2 (career motivation): 'dream, rewarding, application of learned skills' and 'working conditions'. The responses to Q3 (valuable experience) were classified into three groups: 'academic knowledge and friendships established', 'certification and expertise' and 'self-discovery and social knowledge'. Although some of the typical responses within a given group may appear quite distinct at first glance, these responses form a peer group with the same statistical tendencies of positive and negative edges.

We can again confirm that the opinion graphs reflect a similarity based on the underlying semantics, instead of simple text similarities. For example, in the responses of Q1 ('What is your career?'), a response 'a school janitor of a primary school' is positively connected to 'a municipal officer', while it is negatively connected to 'a primary school teacher'. As another example, in Q3 ('What is the most valuable experience that you will take away with you from your time at the university?'), it can be observed that a response 'I found what I really wish to do in my life' is positively connected to 'I met many people and I could have time to face myself'.
human analysts read all the responses and classified them into several categories using their own interpretations. We observed that the identified groups are in good agreement with classifications performed by some of the analysts, in the sense of normalized mutual information (NMI) and adjusted Rand index (ARI). The results of the comparison are presented in the Supplementary Information.

The Sankey diagram in Fig. 3d depicts the flows of respondents among the opinion groups for Q1, Q2 and Q3. It appears that the respondents in the group ‘teachers’ for Q1 mostly flow into the group ‘dream, rewarding, application of learned skills’ for Q2, and a relatively large fraction of them flow into the group ‘academic knowledge and friendships established’ for Q3. This observation suggests some correlations between the factors queried in Q1, Q2 and Q3. This can be further confirmed by performing regression analysis. A more detailed analysis of the above surveys is provided in the Supplementary Information.

Discussion

To overcome the problems that limit the usefulness of free-response questions in large-scale surveys, we proposed a network-based survey framework. The advantages of the present approach are discussed in the following.

Tractability and scalability: the conversion of a large set of raw text data into an opinion graph makes the analysis much more tractable and facilitates coding for numerous responses.

One classical prescription to facilitate coding is to conduct surveys twice, first conducting a small-scale survey using free-response questions to determine the major responses and then conducting a large-scale survey using multiple-choice questions whose possible responses are based on the major opinions obtained previously. However, not only does this two-step survey require additional effort, but there still exists a possibility of missing major responses in the large population. Another approach that can be used for coding is to perform some natural language processing (NLP). However, such an approach is effective only when one has a dataset that contains the appropriate domain knowledge of the survey. To confirm if this is the case for the datasets, we conducted the NLP-based classifications shown in the Supplementary Information. In contrast, once the opinion graph is constructed in our framework, the responses are automatically classified without additional costs or the need for any prior knowledge. This makes it possible to conduct a large-scale survey with free-response questions in an efficient manner.

Statistically principled and democratic: the classification reflects the collective evaluations made by the respondents themselves, instead of the decisions by a small number of analysts. Thus, the result is not subjective or completely objective, but is democratic.

The difficulty in drawing general conclusions is often explained as a weakness of free-response surveys. By contrast, with our framework, it is possible to draw general conclusions from free-response questions. This is possible because (1) the diversity of responses is assessed in a statistically principled manner in terms
of the network configuration and (2) the network is provided by the subjective evaluations of relationships among responses accumulated from the respondents. In contrast, as often done in standard crowd sourcing\textsuperscript{2,3}, we could let employed workers perform step 2 as hypothetical respondents. In that case, however, the workers' coding abilities would play a critical role. We would then no longer have the same objective that we consider in this Article.

Factor analysis is a method frequently used in survey analysis, particularly in the field of psychology. Using the responses to a series of multiple-choice questions, the analysts are able to extract a few factors that characterize those responses from various aspects. It should be noted, however, that factor analysis suffers from the same problems that the multiple-choice questions have in general.

While this framework paves the way for a number of investigations using free-response surveys, it also includes novel problems concerning both survey design and the statistical theory of networks. For example, there is a trade-off between the effort made by the respondents and the degree to which the groups of responses are resolved. In general, the sparser the network, the more information-theoretically difficult it will be to identify the appropriate opinion groups. If the framework requires a respondent to consider a large number of responses, the resulting opinion graph will be dense. Evidently, demanding too much effort from respondents is unrealistic. On the other hand, if each respondent only considers a small number of responses, the resulting network will be sparse. This trade-off is related to the certainty of consensus behind the responses; a set of clearly distinguishable responses can be detected even in a very sparse network, while a dense network is needed to detect subtle differences between opinion groups. The limit of distinguishability of a modular structure of sparse networks is called the detectability limit or the detectability threshold\textsuperscript{25,26} in statistical theory of networks. The difficulty associated with coding in large-scale surveys is directly related to this limit.

A possible bias in step 2 should also be mentioned. The respondents may be influenced by the responses of others. When this point is a critical issue, we can maintain independence among responses by simply hiding step 2 until step 1 is completed according to the procedure depicted in Fig. 1a. However, a more important issue in free-response surveys is the problem of ‘frame of reference’\textsuperscript{2,4}, that is, the risk of failing to collect sufficiently diverse opinions while keeping the coherence among them. Our framework is more suitable for the type of survey in which the respondents are expected to be well-informed\textsuperscript{27} and speculate deeply in reference to the opinions of others. In such a situation, displaying other responses can be considered beneficial rather than detrimental.

Although evaluating the diversity of responses is problematic in many senses, it is always needed whenever a survey seeks information from the public. We hope that our framework will play a useful role in society and help researchers acquire new knowledge in various fields of science.

Materials and methods

Graph clustering based on statistical inference. We perform the partitioning of an opinion graph using graph clustering, which is a component of probabilistic machine learning. This task is also known as community detection, in more restricted cases. Although numerous frameworks and algorithms have been proposed in the literature, in this study, we employ the statistical inference approach.

A random graph model with a modular structure called the stochastic block model\textsuperscript{28,29} is a canonical generative model for the inference of a modular structure. The group label is denoted by \( \sigma \in \{1, \ldots, q\} \), where \( q \) is the number of groups, and the set of group assignments is denoted by \( \{\sigma\} \). The model parameters, \( \gamma, \omega^+, \omega^- \) specify the macroscopic structure of the network: \( \gamma \) is a \( q \times q \) dimensional vector that represents the fractions of group sizes, and \( \omega^+ \) and \( \omega^- \) represent the \( q \times q \) density matrices (also known as the affinity matrices) that determine the connection probabilities within/between groups with respect to the positive and negative edges, respectively. An instance of the stochastic block model, that is, the adjacency matrix \( A \), is generated as follows. For each vertex, we randomly determine the group assignment such that the fraction of the group \( \sigma \) is equal to \( \gamma_{\sigma} \) on average. Then, for all pairs of vertices, we generate edges independently and randomly on the basis of group assignments. For example, when vertices \( i \) and \( j \) have the assignments \( \sigma_i = \sigma_j = \sigma \), they are connected by a positive edge with probability \( \omega^+_{\sigma_i \sigma_j} \), connected by a negative edge with probability \( \omega^-_{\sigma_i \sigma_j} \), and not connected with probability \( 1 - \omega^+_{\sigma_i \sigma_j} - \omega^-_{\sigma_i \sigma_j} \). (This type of stochastic block model is sometimes referred to as a labelled stochastic block model.) It is important to note that we can enforce the sampling of vertices referred by a respondent to be random. Therefore, the probabilistic nature of the edge generation process assumed in the stochastic block model is guaranteed. It is not necessary that the generation process of real data resembles that of the model for the stochastic block model to be applicable. However, the results will be more reliable when they are similar.

The likelihood function of the stochastic block model with \( N \) vertices is given by

\[
p(A, \{\sigma\} | \gamma, \omega^+, \omega^-) = \prod_{i=1}^{N} \prod_{j<i} |1 - (\omega^+_{\sigma_i \sigma_j} + \omega^-_{\sigma_i \sigma_j})|^{\delta_{A_{ij}}} \times |\omega^+_{\sigma_i \omega^-_{\sigma_j}}|^{\delta_{A_{ij}}} \times |\omega^-_{\sigma_i \omega^+_{\sigma_j}}|^{\delta_{A_{ij}}}
\]

This stochastic block model assumes that all of the vertices within a given group are statistically equivalent. However, in the opinion graph, even within a group of similar responses, some responses may be more popular than others. This heterogeneity is manifested as a hub structure in the opinion graph, which cannot be modelled by equation (1). A variant of the stochastic block model that properly considers a hub structure is called the degree-corrected stochastic block model\textsuperscript{30}. The corresponding likelihood function in this model is given by

\[
p(A, \{\sigma\} | \gamma, \omega^+, \omega^-) = \prod_{i=1}^{N} \prod_{j<i} |1 - (d^+_{\omega^+_{\sigma_i \sigma_j} + d^-_{\omega^-_{\sigma_i \sigma_j}}) + d^-_{\omega^-_{\sigma_i \sigma_j} d^+_{\sigma_j \sigma_i}}|^{\delta_{A_{ij}}} \times |d^+_{\omega^+_{\sigma_i \sigma_j} d^+_{\sigma_j \sigma_i}}|^{\delta_{A_{ij}}} \times |d^-_{\omega^-_{\sigma_i \sigma_j} d^-_{\sigma_j \sigma_i}}|^{\delta_{A_{ij}}}
\]

where \( d^+ \) and \( d^- \) represent the degrees, that is the number of neighbours connected via positive and negative edges from vertex \( i \), respectively. The degrees of vertices can be readily obtained from \( A \). (Note also that the density matrices are renormalized compared to those in equation (1), owing to the degree distribution.)

Based on equation (2), we infer the group assignment \( \sigma_i \) by solving for its marginal posterior distribution, \( p(\sigma_i | A, \gamma, \omega^+, \omega^-) \), and the values of the model parameters are determined so that the marginal likelihood \( \sum_{i} p(A, \{\sigma\} | \gamma, \omega^+, \omega^-) \) is maximized. Unfortunately, precise computations of the marginal posterior distributions and model parameters are computationally demanding. A commonly used approximation method is the expectation-maximization algorithm\textsuperscript{31}, and belief propagation\textsuperscript{32,33} is known to be an efficient and accurate algorithm for its E-step. We implemented this approach for the stochastic block model. We note here that other methods such as the Monte Carlo method can also be used, as long as they are computationally feasible and applicable to the opinion graphs.
There are three reasons for employing the inference algorithm described above. First, using this algorithm, instead of a single partition that optimizes a certain objective function, we can obtain a degree of certainty as the probability of group assignment for each opinion vertex. Second, the properties of the stochastic block model are theoretically well known\textsuperscript{2,8,26,34–36} and efficient algorithms exist for the sparse case. Third, the algorithm can learn an arbitrary connection pattern for each edge type; that is, we do not need to assume that negative edges strictly exhibit a disassortative structure.

Although we considered only positive and negative edges, it is possible to allow more than two labels and make finer distinctions among responses. In principle, the opinion graph would then acquire more information. However, it must be noted that this does not necessarily imply better identification of groups in practice, because it might be very difficult to devise an algorithm that can stably and accurately treat opinion graphs with a large number of edge types\textsuperscript{37}.

To determine the appropriate number of groups, we employed leave-one-out cross-validation estimates of prediction errors, which can be evaluated efficiently\textsuperscript{38} using belief propagation. Note that we should not expect such estimates to be very accurate. Their role is merely to provide a rough estimate using a mathematically principled approach. For the final determination of the number of groups, we examined the manner in which the network is actually partitioned using the alluvial diagram.

**Data availability**

The network datasets that support the findings of this study are available in a GitHub repository at https://github.com/tatsuro-kawamoto/opinion_graphs. The graph clustering code that supports the findings of this study is available in a GitHub repository at https://github.com/tatsuro-kawamoto/graphBIX.

Received: 30 November 2018; Accepted: 7 June 2019; Published online: 9 July 2019.

**References**

1. Kahn, R. L. & Cannell, C. F. *The Dynamics of Interviewing: Theory, Technique, and Cases* (Wiley, 1957).
2. Schuman, H. & Scott, J. Problems in the use of survey questions to measure public opinion. *Science* **236**, 957–959 (1987).
3. Schuman, H. & Presser, S. Questions and Answers in Attitude Surveys: Experiments on Question Form, Word, and Context (Sage, 1996).
4. RePass, D. E. Issue salience and party choice. *Am. Polit. Sci. Rev.* **65**, 389–400 (1971).
5. Kelley, S. J. *Interpreting Elections* (Princeton Univ. Press, 2014).
6. Geer, J. G. What do open-ended questions measure? *Public Opin. Q.* **52**, 365–371 (1988).
7. Singleton, R. & Straitis, B. C. *Approaches to Social Research* 6th edn (Oxford Univ. Press, 2017).
8. Schuman, H. The random probe: a technique for evaluating the validity of closed questions. *Am. Sociol. Rev.* **31**, 218–222 (1966).
9. Lombard, M., Snyder-Duch, J. & Bracken, C. C. Content analysis in mass communication: assessment and reporting of intercoder reliability. *Human Commun. Res.* **28**, 587–604 (2002).
10. Giddens, A. & Sutton, P. W. *Sociology* 7th edn (Polity Press, 2013).
11. Aicher, C., Jacobs, A. Z. & Clauset, A. Learning latent block structure in weighted networks. *J. Complex Netw.* **3**, 221–248 (2015).
12. Newman, M. E. J. Network structure from rich but noisy data. *Nat. Phys.* **14**, 542–545 (2018).
13. Peixoto, T. P. Reconstructing networks with unknown and heterogeneous errors. *Phys. Rev. X* **8**, 041011 (2018).
14. Rosvall, M. & Bergstrom, C. T. Mapping change in large networks. *PLoS ONE* **5**, 1–7 (2010).
15. Kawamoto, T. & Kabashima, Y. Comparative analysis on the selection of number of clusters in community detection. *Phys. Rev. E* **97**, 022315 (2018).
16. Danon, L., Diaz-Guilera, A., Duch, J. & Arenas, A. Comparing community structure identification. *J. Stat. Mech.* **2005**, P09008 (2005).
17. Rand, W. M. Objective criteria for the evaluation of clustering methods. *J. Am. Stat. Assoc.* **66**, 846–850 (1971).
18. Hubert, L. & Arabie, P. Comparing partitions. *J. Classif.* **2**, 193–218 (1985).
19. Simon, A. F. & Xenos, M. Dimensional reduction of word-frequency data as a substitute for intersubjective content analysis. *Polit. Anal.* **12**, 63–75 (2004).
20. Hopkins, D. J. & King, G. A method of automated nonparametric content analysis for social science. *Am. J. Polit. Sci.* **54**, 229–247 (2010).
21. Roberts, M. E. et al. Structural topic models for open-ended survey responses. *Am. J. Polit. Sci.* **58**, 1064–1082 (2014).
22. Benoist, K., Conway, D., Lauderdale, B. E., Laver, M. & Mikhailov, S. Crowd-sourced text analysis: reproducible and agile production of political data. *Am. Polit. Sci. Rev.* **100**, 278–295 (2016).
23. Lind, E., Gruber, M. & Boomgaard, H. G. Content analysis by the crowd: assessing the usability of crowdsourcing for coding latent constructs. *Commun. Methods Meas.* **11**, 191–209 (2017).
24. Jacobson, M. R., Whyte, C. E. & Azzam, T. Using crowdsourcing to code open-ended responses: a mixed methods approach. *Am. J. Eval.* **39**, 413–429 (2018).
25. Decelle, A., Krzakala, F., Moore, C. & Zdeborová, L. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84**, 066106 (2011).
26. Moore, C. The computer science and physics of community detection: landscapes, phase transitions, and hardness. Preprint at https://arxiv.org/abs/1702.00467 (2017).
27. Fishkin, J. S. *When the People Speak: Deliberative Democracy and Public Consultation* (Oxford Univ. Press, 2011).
28. Holland, P. W., Laskey, K. B. & Leinhardt, S. Stochastic blockmodels: first steps. *Soc. Netw.* **5**, 109–137 (1983).
29. Wang, Y. J. & Wong, G. Y. Stochastic blockmodels for directed graphs. *J. Am. Stat. Assoc.* **82**, 8–19 (1987).
30. Karrer, B. & Newman, M. E. J. Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83**, 016107 (2011).
31. Bishop, C. M. *Pattern Recognition and Machine Learning* (Information Science and Statistics) (Springer, 2006).
32. Mézard, M. & Montanari, A. *Information, Physics, and Computation* (Oxford Univ. Press, 2009).
33. Decelle, A., Krzakala, F., Moore, C. & Zdeborová, L. Inferring and phase transitions in the detection of modules in sparse networks. *Phys. Rev. Lett.* **107**, 065701 (2011).
34. Kawamoto, T. Algorithmic detectability threshold of the stochastic block model. *Phys. Rev. E* **97**, 032301 (2018).
35. Abbe, E. Community detection and stochastic block models: recent developments. Preprint at https://arxiv.org/abs/1703.01046 (2017).
36. Peixoto, T. P. Bayesian stochastic blockmodeling. Preprint at https://arxiv.org/abs/1703.10225 (2017).
37. Kawamoto, T. Algorithmic infeasibility of community detection in higher-order networks. Preprint at https://arxiv.org/abs/1701.08816 (2017).
38. Kawamoto, T. & Kabashima, Y. Cross-validation estimate of the number of clusters in a network. *Sci. Rep.* **7**, 5327 (2017).

**Acknowledgements**

The authors thank H. Tokioka and S. Shinomoto for discussions. The authors are also grateful to J. Park and M. Rosvall for their comments. Finally, the authors appreciate all the people who contributed to the poll on the 2016 US presidential election and acknowledge support from the Faculty of Education in Kagawa University and the reunion of the faculty. T.K. was supported by JSPS (Japan) KAKENHI grant no. 26011023. T.A. was supported by the Research Institute for Mathematical Sciences, a joint research centre at Kyoto University, and open collaborative research at the National Institute of Informatics (NII) (Japan (FY2017). T.K. and T.A. acknowledge financial support from JSPS KAKENHI grant no. 18K18604.

**Author contributions**

T.K. and T.A. designed the survey framework, analysed the data and wrote the manuscript. T.K. implemented the online survey system. T.K. conducted a survey of the poll on the 2016 US presidential election and T.A. mainly conducted a survey focusing on graduates of the Faculty of Education of a particular university.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s42256-019-0071-y.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to T.K.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019