Ito and Ohnishi *Appl Netw Sci* (2021) 6:52
https://doi.org/10.1007/s41109-021-00393-4

Applied Network Science

RESEARCH

Open Access

Co-occurrence network of TV advertisements revealing Japanese lifestyle

Mariko I. Ito* and Takaaki Ohnishi

*Correspondence: marikoito.rnu1@gmail.com
Graduate School of Artificial Intelligence and Science, Rikkyo University, Tokyo, Japan

Abstract

The relationship between culture and the appeals in TV advertisements has been extensively studied. We attempted to reveal the image structure produced by TV commercials in Japan, which may show the cultural features of the country, and to evaluate its temporal change. For this purpose, we constructed and analysed a co-occurrence network of keywords related to TV commercials by using immense data that include the records of all TV commercials aired in the Kanto area in Japan including Tokyo for a period of 15 years. We found a strong heterogeneity of the co-occurrence relationship, where a few keywords, e.g., 'woman', 'man', 'animation', and 'logo', co-occur with a huge number of other keywords every year. A community on a co-occurrence network can be regarded as a set of keywords that are mutually associated with each other through TV commercials. We examined the characteristics of the communities by associating them with categories of advertised products and found a temporal change in which the relationship between the communities possessing the image of entertainment and children and the category of PC and A/V gradually increases in strength. However, there was a consistent tendency in the examined period for the product categories related to communities that include 'man' to be less associated with those that include 'woman' and vice versa, which implicates a gender role inequality underlying the various appeals in TV commercials.

Keywords: TV commercials, Co-occurrence networks, Community structure, Image structure, Gender role inequality

Introduction

A network representation is an effective way to determine a hidden characteristic in the objects observed (Newman 2003). For example, we can find several key persons if we construct a network in which each node represents a person and each edge represents the interaction between two individuals, and consider the role of each node (an individual) within the network structure. Various meanings and quantities are assigned to the edges in the network representation, for example, friendship or influence among individuals (Watts and Dodds 2007; Castellano et al. 2009), regulatory relationships between genes or neurons (Gross and Blasius 2008; Socolar and Kauffman 2003), and the correlation between two time series of, for example, stock returns (Mantegna 1999; Mizokami and Ohnishi 2018). The co-occurrence of two items is one of the relationships that can be represented by an edge (Radhakrishnan et al. 2017; Su and Lee 2010; Zhang et al. 2012; Özgür et al. 2008).

Keyword co-occurrence networks, where nodes denote the keywords in articles, have been constructed and analysed to investigate the knowledge structure in the academic fields (Radhakrishnan et al. 2017; Su and Lee 2010). In these networks, an edge represents the co-occurrence of two keywords in the same article, and an edge occasionally has a weight that represents the frequency of their co-occurrence. A node with a high degree or strength, which is the sum of the weights of edges connected to the focal node, can be regarded as the core of the knowledge structure in the sense that it significantly appears with various topics in the field. Previous studies have investigated such cores in the knowledge structure and their neighbouring nodes to reveal the trends or important topics in the field.

Such a co-occurrence network representation can also be applied to the investigation of the image structure. Ito and Ohnishi (2020) constructed a co-occurrence network of keywords that describe the content of TV commercials and examined the image structure produced by Japanese TV commercials during the last 4 years. Each node in the network is a keyword, and the weight of an edge represents the diversity of the product categories among which two keywords appear in the same commercial. The community structure in the co-occurrence network, where each community can be regarded as a set of images mutually associated through TV commercials, can represent a characteristic of Japanese culture. For example, the keywords 'man' and 'woman', both of which were the cores in the image structure, were assigned to different communities. The community including 'man' was composed of 'laugh', 'talk', and 'office' among other terms, whereas that including 'woman' includes 'look back', 'room', and 'cafe'. Moreover, these communities are related differently to categories of products advertised in commercials.

Indeed, among the extensive studies on the effect of TV commercials (Hekkert et al. 2013; Carreón et al. 2019; Davtyan and Cunningham 2017; Boyland and Halford 2013; Pan 2011), the discussion on the cultural difference of viewer preferences, the appeal of TV commercials, and the response of viewers to commercials have a long history (Okazaki and Mueller 2007; De Mooij and Hofstede 2011; Moon and Chan 2005; Pham et al. 2013; Liu et al. 2019; Milner and Collins 2000; Bresnahan et al. 2001). Previous studies have classified cultures or countries based on the extent of individualism or collectivism, the strength of uncertainty avoidance or differentiation of male and female roles, and other factors (De Mooij and Hofstede 2011; Okazaki and Mueller 2007). It has been discussed that the feelings of viewers evoked by the items shown in commercials vary significantly according to culture. For example, people in some countries tend to regard direct advertisements as aggressive, whereas people in other countries find them informative (Liu et al. 2019; Okazaki and Mueller 2007; De Mooij and Hofstede 2011). This means that, in turn, the culture of a country can be observed at least partially through TV commercials.

In many investigations conducted on the cultural differences in TV commercials and responses of viewers, there has been an issue of selection bias caused by a limitation or shortage of the collected data (Moon and Chan 2005; Milner and Collins 2000; Bresnahan et al. 2001). The aforementioned study of the image structure of TV commercials (Ito and Ohnishi 2020) addressed this issue by analysing immense amounts of data, including all TV commercials aired in the Kanto area of Japan, including Tokyo, from 2017 to 2020, by using the knowledge of complex networks. However, this study was based only on a single network representing keywords co-occurrence in TV commercials from 2017 to 2020. Here, the question is whether the observed characteristics of the image structure used in the TV commercials in such a study are only attributed to the examined period or are robust for longer periods.

In the present study, we analysed the data of Japanese TV commercials aired over a period 15 years, from 2006 to 2020. For each year, we constructed a co-occurrence network of keywords describing TV commercials and examined the features of the image structure by analysing the co-occurrence network. We in particular investigated the temporal change of the image structure, i.e., the difference or common features in these co-occurrence networks in the examined years. We conducted a community detection for each year's network and associated communities in networks of consecutive years by evaluating the flow of nodes between them. By the analysis on the temporal change of community structure, we found that the community whose nodes are associated with the keyword 'woman' seems to have a significant relationship to that associated with 'product'. Moreover, the results imply a social issue of gender role inequality: The community whose nodes are associated with 'man' and the community whose nodes are associated with 'woman' exhibited completely different characteristics in the sense of how the nodes in each community are related to the categories of the advertised products during almost all years within the examined period.

Materials and methods

Data description

We analysed the data of TV commercials provided by M Data Co., Ltd. (https://mdata.tv/en/). Such data include information on TV commercials aired on five TV stations, Fuji TV, Nippon TV, TBS TV, TV Asahi, and TV Tokyo, in the Kanto area of Japan. The data recorded TV commercials aired from January 1, 2006 to December 31, 2019, and those aired from January 1 to June 30, 2020. The scenario of each commercial is described by keywords such as 'mother and child' and 'nursery'.

Each commercial is classified according to the type of product advertised in the commercial. The classification has three levels, large, middle, and small. We use the large and middle classifications in this study, which are labelled *category* and *subcategory*, respectively. Therefore, a product advertised in a commercial is classified into a subcategory, which further belongs to a category. For example, a certain product can be classified into the subcategory of "health drink", which belongs to the category of "drink". Other subcategories belonging to "drink" category are "tea", "fruit juice drink", and so on. Although the number of categories differs in different years according to actually aired commercials, the following categories are used in every year examined: cup noodles, pet food, food, machines, credit cards, finance and insurance, beer, liquor, logistics, communication, car, medicine, snack, oil and tire, infomercial, toy, distribution industry A, distribution industry, detergent, apparel, appliance, household goods, cosmetics, estate, roadshow, AV software, PC and A/V, canned coffee, drink, tobacco, sports, camera and watch, interior, publication, others. Note that though distribution industry and distribution industry A have similar names, they are different categories—the former one includes restaurants, retailers, and specialised stores, and the latter one includes convenience stores, department stores, and supermarkets. The number of subcategories also varies in the analysed period. Table 1 shows a summary of the analysed data: the number of commercials, the mean number of co-occurring keywords per commercial, and the numbers of categories and subcategories in each year. Regarding the number of analysed commercials, those that share the same scenario but are aired at different times are counted as different commercials.

Analysis of co-occurrence network of TV commercials and image structure

We constructed a weighted co-occurrence network G_y of the keywords in TV commercials for the year $y \in \{2006, 2007, \ldots, 2020\}$. First, for year y and subcategory κ, we defined an unweighted network $G_y^{(\kappa)}$ in which each node represents a keyword, and an edge is drawn between two nodes if they co-occur at least once in the same TV commercial of the product belonging to the subcategory in year y (Fig. 1). Subsequently, we constructed a weighted network G_y by merging these unweighted networks $G_y^{(\kappa)}$ in year y as follows. The nodes in network G_y are all keywords that appear in TV commercials in year y, and the weight of an edge between two nodes denotes the proportion of the number of networks for subcategories $G_y^{(\kappa)}$ where the edge exists between these two nodes against the total number of subcategories in year y. Thus, the weight of an edge between two nodes increases when they co-occur in the same commercials of various types of products.

The resulting network G_y indicates that, for example, the nodes of ‘mother’ and ‘child’ are connected by an edge with a weight of 0.9, which shows that these images are frequently used in the same commercial to advertise almost all types of products. Let $A_{i,j}^{y}$ be the weight matrix of G_y. The strength of node i, $s_i = \sum_j A_{i,j}^{y} (\in \mathbb{R})$, is the

Table 1 Analysed data

Year	# commercials	# co-occurrence keywords per commercial	# categories	# subcategories
2006	1,591,712	3.30	39	132
2007	1,610,762	3.51	39	129
2008	1,590,339	3.64	40	129
2009	1,592,488	3.81	40	129
2010	1,652,888	4.29	40	130
2011	1,657,522	4.42	40	130
2012	1,676,052	4.60	40	129
2013	1,676,414	4.69	38	128
2014	1,672,136	4.71	38	129
2015	1,683,645	5.42	40	129
2016	1,700,535	6.06	39	129
2017	1,682,171	6.57	39	130
2018	1,672,548	7.29	38	131
2019	1,657,577	7.58	37	128
2020	804,827	7.37	37	128

The number of commercials, the mean number of co-occurrence keywords per commercial, and the numbers of categories and subcategories are exhibited for each year

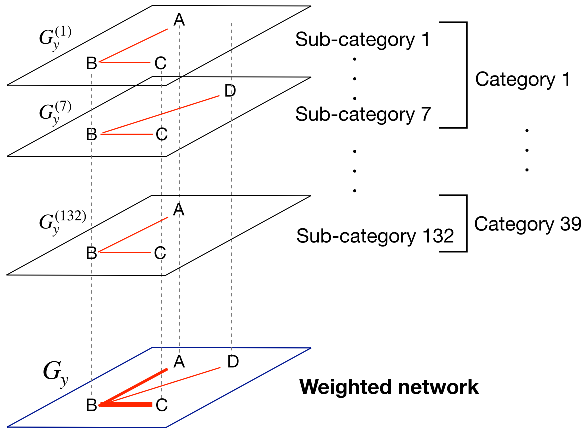

Fig. 1 Schematic image of the construction the co-occurrence network G_y. The figure was adapted from Ref. (Ito and Ohnishi 2020)

extent to which node i co-occurs with other keywords over various subcategories, whereas the degree k_i of node i represents the number of keywords that have co-occurred with node i at least once.

We applied community detection in G_y through the modularity maximisation (Fortunato 2010). Modularity Q is defined as follows:

$$Q = \frac{1}{2m}\sum_{i,j}\left(A^y_{i,j} - \frac{s_i s_j}{2m}\right)\delta(c(i), c(j)), \tag{1}$$

where $2m = \sum_{i,j} A^y_{i,j}$, δ denotes Kronecker's delta, and the community to which node i is assigned is denoted as $c(i)$. Because the term $s_i s_j/2m$ in Eq. (1) is the expected weight between nodes i and j in a random network, where the strength distribution is the same as that of G_y, the modularity Q measures the extent to which nodes within a community are connected tightly compared to the null model for a given graph partition. We obtained a graph partition by (locally) maximising the modularity Q with the Louvain heuristic (Blondel et al. 2008; Aynaud 2020). The Louvain heuristic is a fast algorithm for locally maximising the modularity and is applicable to networks with large sizes, as in our case. We applied the graph partitions 10 times using the Louvain heuristic and adopted the partition that resulted in the highest modularity. In our case, a community consists of keywords that significantly co-occur and are mutually associated over various subcategories.

We also examined the relationship between the communities in G_y and the product categories. Note that a node, that is, a keyword, can appear in commercials of multiple (sub)categories, whereas it is assigned to a single community. We first evaluated the extent to which a node is related to each category as follows: Let $N_{i,k}$ be the number of subcategories of category k in which node i appears at least once. For example, the maximum value of $N_{i,k}$ is 7 when there are seven subcategories in category k. We normalised $N_{i,k}$ for all nodes appearing in category k as $n_{i,k} := N_{i,k}/\sum_j N_{j,k}$. The value of $n_{i,k}$ represents the extent to which node i is related to category k. Subsequently, we summed $n_{i,k}$ of nodes belonging to community l as follows:

$$W_{l,k} = \sum_{\text{node } i \in \text{community } l} n_{i,k}, \tag{2}$$

and normalised it using all nodes belonging to community l and defined as $w_{l,k}$:

$$w_{l,k} = \frac{W_{l,k}}{\sum_j W_{l,j}}, \tag{3}$$

which represents the extent to which community l is composed of nodes related to category k, and thus denotes the strength of the relationship between category k and community l.

Results

Cores of image structure and temporal changes

In Table 2, we summarise the features of the co-occurrence network G_y of the keywords found in TV commercials. The mean degree and the mean strength gradually increase with the year. The reason for this increase may not be attributed to the meaningful change in the characteristics of the TV commercials or of the culture, but presumably to the editorial aspects of the dataset, because the mean number of co-occurrence keywords per commercial continuously increases with the year (Table 1). Rather, it should be noted that we can observe power-law-like distributions in degree and strength in any year, showing a strong heterogeneity (Fig. 2, Additional file 1: Section S.1). Note that the large strength of a node indicates that it has co-occurred with many nodes through TV commercials of various products. By contrast, the degree only represents the number of other nodes that have co-occurred with the node at least once in the same commercial. Therefore, there were a few nodes that co-occurred frequently with other nodes regardless of the variety of advertised products, whereas many nodes co-occurred with a few nodes. This characteristic was robust during the examined period. Figure 2c shows the relationship between the degree and the strength of each node. We can see a clear tendency in which nodes with a large degree also have a large strength. Hereafter, we regard nodes with a large strength as the cores of the image structure represented by G_y.

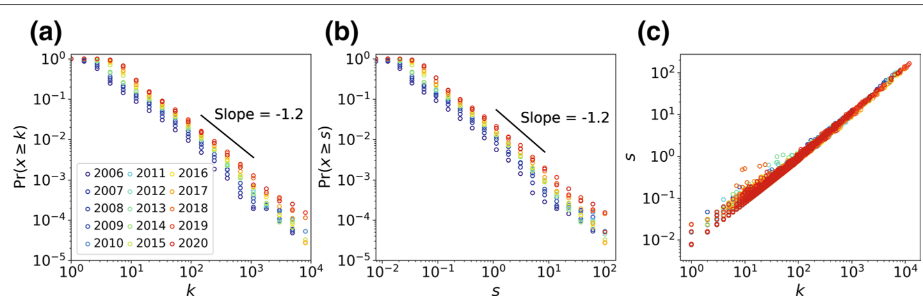

Fig. 2 Complementary cumulative distribution function (CCDF) of **a** the degree k and **b** strength s (for the fitting of power law, see Additional file 1: Section S.1). Each colour represents each year. The relationship between the degree and strength of each node is shown in **c**

Table 3 exhibits the nodes having the first to the fifth-largest degree and strength in each year. We can see 'woman' and 'man' in the top-five every year, and thus these nodes can be regarded as robust cores in the image structure of the TV commercials. The keywords 'cinema scope', 'animation', and 'man and woman' were in the higher ranks during the earlier years, but 'product', 'logo', and 'white back' had higher ranks in the later years.

Temporal change of community structure

Subsequently, we demonstrate the characteristics of the community structure in the co-occurrence network of TV commercials G_y and its temporal changes. The resulting modularity of the graph partition and the number of communities are listed in Table 2.

Table 4 exhibits the node with the largest strength and the size of each community whose size is within the seventh largest during each year. Hereafter, we refer to such a node with the largest strength in each community as a *representative node*. Core nodes with the largest strength within the whole network, as shown in Table 3, were mostly assigned to different communities, and represented the communities, each of which is a subset of the image structure. Regarding the communities marked with a star in Table 4, seven nodes with the largest strength in each community are shown in Tables 5 and 6 , and the nodes with the largest strength in communities other than those shown in Tables 5 and 6 are shown in Additional file 1: Tables S.2 to S.6.

The diagram in Fig. 3, called the Sankey diagram, visualises the flow of nodes between communities during two consecutive years. Each node represents a community, which is labelled by a representative node, and each flow exhibits the number of nodes that move from the source community to the target community. Herein, we show only communities that have from the first to the seventh-largest size in each year and show only flows between them in the Sankey diagram. In addition, to clarify the mainstream, the flows are removed if their proportion to the total outflow from the source is less than 20%.

Table 2 Feature of the co-occurrence network of each year

Year	# nodes	# edges	Density	Mean degree	Mean strength	# com	Q
2006	20,980	79,095	3.59×10^{-4}	7.54	0.06	751	0.38
2007	18,915	84,788	4.74×10^{-4}	8.97	0.08	433	0.34
2008	17,383	88,294	5.84×10^{-4}	10.16	0.09	263	0.31
2009	16,926	95,191	6.65×10^{-4}	11.25	0.11	147	0.30
2010	19,110	125,658	6.88×10^{-4}	13.15	0.12	82	0.28
2011	20,275	135,329	6.58×10^{-4}	13.35	0.12	116	0.30
2012	22,491	147,912	5.85×10^{-4}	13.15	0.12	156	0.32
2013	22,524	148,693	5.86×10^{-4}	13.20	0.12	118	0.32
2014	22,597	149,861	5.87×10^{-4}	13.26	0.12	133	0.33
2015	30,811	236,738	4.99×10^{-4}	15.37	0.14	151	0.36
2016	37,107	303,259	4.40×10^{-4}	16.35	0.14	177	0.37
2017	36,312	328,386	4.98×10^{-4}	18.09	0.16	120	0.35
2018	35,121	371,264	6.02×10^{-4}	21.14	0.19	66	0.30
2019	32,766	371,899	6.93×10^{-4}	22.70	0.21	52	0.28
2020	19,895	211,498	1.07×10^{-3}	21.26	0.20	33	0.28

The network size, number of edges, edge density, mean degree and strength, number of communities, and modularity of community detection are exhibited

Table 3 Five nodes with the largest strength during each year

2006	Cinema scope 60.97 (5212)	Woman 39.34 (3238)	Man 23.78 (2138)	Animation 21.68 (2153)	Man and woman 9.14 (930)
2007	Cinema scope 68.12 (5362)	Woman 54.18 (3920)	Man 35.74 (2813)	Animation 24.48 (2190)	Man and woman 11.73 (1086)
2008	Cinema scope 74.84 (5571)	Woman 58.29 (4026)	Man 40.69 (3005)	Animation 25.92 (2236)	Man and woman 13.42 (1193)
2009	Cinema scope 99.13 (7045)	Woman 65.67 (4385)	Man 47.45 (3425)	Animation 28.74 (2359)	Man and woman 16.19 (1,385)
2010	Cinema scope 117.15 (8205)	Woman 87.88 (5809)	Man 66.08 (4752)	Animation 29.27 (2460)	Man and woman 23.17 (1,908)
2011	Woman 98.23 (6441)	Man 70.57 (4964)	Animation 30.46 (2593)	Cinema scope 24.60 (2145)	Man and woman 21.87 (1,837)
2012	Woman 103.53 (7033)	Man 71.06 (5076)	Animation 32.52 (2761)	Man and woman 24.54 (2057)	Product 22.77 (1,938)
2013	Woman 99.20 (6772)	Man 71.94 (5265)	Animation 33.05 (2787)	Product 24.95 (2111)	Man and woman 24.60 (2034)
2014	Woman 94.76 (6520)	Man 67.91 (5024)	Animation 32.53 (271)	Product 24.77 (225)	Man and woman 21.20 (1832)
2015	Woman 116.56 (8513)	Man 85.24 (6468)	Product 52.27 (4489)	Animation 41.30 (3565)	Man and woman 30.27 (2612)
2016	Woman 123.52 (9514)	Man 86.73 (7096)	Product 65.33 (5656)	Animation 46.16 (410)	Cinema scope 43.98 (4100)
2017	Woman 128.12 (9810)	Product 97.50 (8075)	Man 92.86 (7372)	Cinema scope 57.67 (5086)	Logo 53.55 (4768)
2018	Woman 132.80 (9703)	Product 132.19 (10,539)	Logo 123.76 (9813)	Man 105.29 (8082)	White back 102.80 (8133)
2019	Logo 165.55 (11,951)	Product 153.51 (11,365)	White back 145.06 (10,609)	Woman 139.83 (9693)	Man 116.83 (8368)
2020	Logo 110.14 (8008)	Product 100.18 (7540)	White back 97.17 (7215)	Woman 89.66 (6329)	Man 74.35 (5396)

The degree of each node is also shown in parentheses

We found several major streams where many nodes moved together between communities in the examined period using the Sankey diagram. First, we can observe the stream of communities represented by 'woman' during the period from 2006 to 2016 and 'product' from 2017 to 2020, which is called Stream 1. Second, the stream of communities represented by 'animation' for every year except 2019, and by 'black back' in 2019, is also significant, and is called Stream 2. Third, we found a robust stream that consists of communities represented by 'family' and 'eat'. Here, the size of the community represented by 'eat' in 2020 was ranked lower than the seventh largest, and thus it is not shown in the Sankey diagram (the rank was the ninth largest as shown in Additional file 1: Fig. S.5). Considering, however, that the flow from the community represented by 'eat' in 2019 to that represented by 'eat' in 2020 is non-negligible, we set Stream 3 as that configured with the communities represented by 'family', 'family', 'indoor', 'family', 'eat', 'eat', 'family', 'family', 'family', 'eat', 'eat', 'eat', 'eat', 'eat', and 'eat' in the years from 2006 to 2020, respectively. As mentioned before, the nodes of 'man' and 'woman' were ranked within the top-five highest strength for G_y every year. Moreover, we found that these nodes were never assigned to the same community in the analysis period and were the

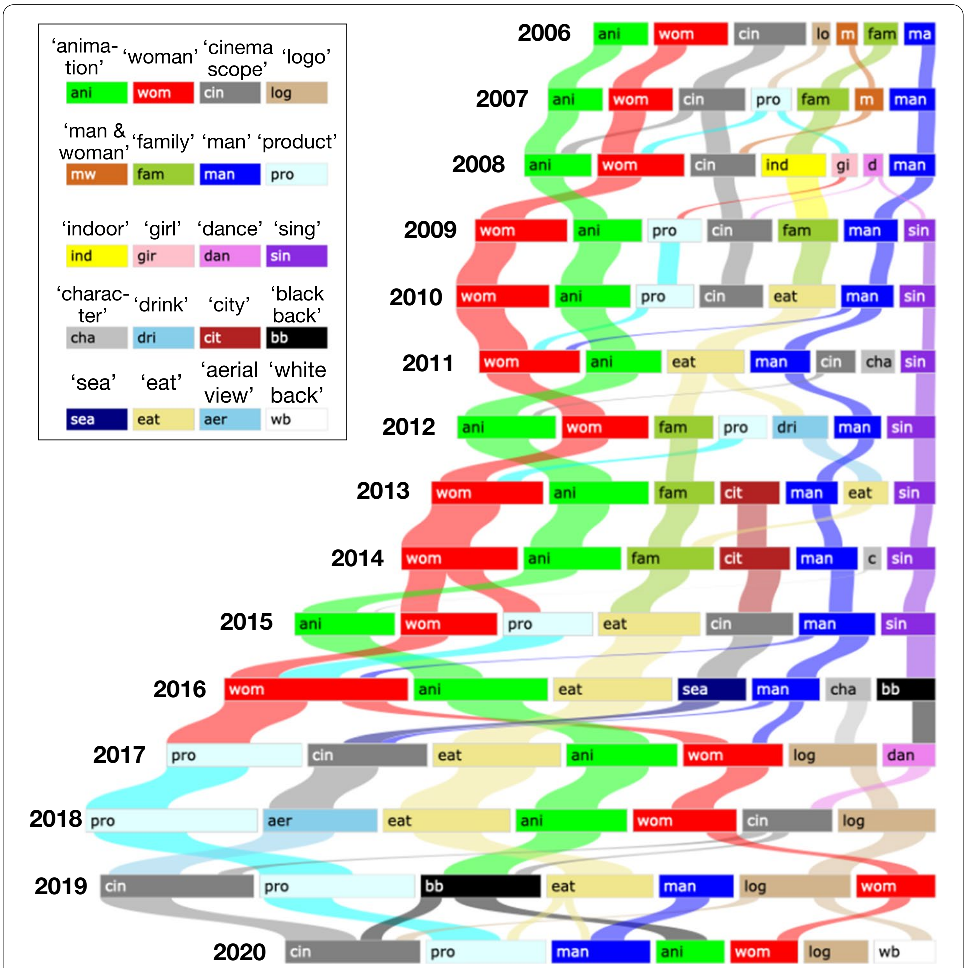

Fig. 3 Sankey diagram showing the flow of nodes between communities. Each colour denotes the community with each representative node, as shown in the legend

representative nodes in their communities. Therefore, we also observed the following two streams configured by the communities represented by 'man' every year and that represented by 'woman', which are called Streams 4 and 5, respectively. Seven nodes with the largest strength in the communities belonging to Streams 1, 2, and 3 and those belonging to Streams 4 and 5 are shown in Tables 5 and 6 , respectively.

Many nodes in the community represented by 'woman' in 2016 moved to either the community represented by 'product' or the community represented by 'woman' (Fig. 3). Such a split of nodes was presumably caused because 'product' was in the community represented by 'woman' in 2016 (Table 5). The node of 'product' was frequently in the same community as that of 'woman', implicating a significant relationship between the image shared with 'woman' and that with 'product' in TV commercials. In Stream 2, the communities, on the whole, include keywords associated with entertainment, children, and things kids like, e.g. 'animation', 'game screen', 'boy', and 'girl'. Communities in Stream 3 almost always represented by 'family' and 'eat'. They also include keywords relating to family members and 'kitchen' and 'cooking' and so on. Therefore, it is inferred that the images of family and behaviour related to eating have a strong relationship. Regarding

Stream 4, the communities represented by 'man' share keywords that evoke a positive expression of feelings or communication such as 'laugh', 'surprise', and 'conversation'. Regarding Stream 5, which is the sequence of communities represented by 'woman', the involved communities were the same as that in Stream 1 until 2016, as mentioned before. It should be noted that the communities in Stream 5 always include 'room' or 'indoor', except in the years 2008, 2019, and 2020.

We further investigated the characteristics of these streams of communities by evaluating the relationship between the communities and the categories, $w_{l,k}$, as described in the Materials and methods section. The heatmaps in Fig. 4 exhibit the value of $w_{l,k}$ in each year for Streams 1, 2, and 3. The horizontal and vertical axes show the category and year, respectively, and we can determine the extent of the relationship $w_{l,k}$ between category k and community l which configures the stream in the year. Figure 5 shows the value of $w_{l,k}$ for Streams 4 and 5 in the same manner as Fig. 4. The value of $w_{l,k}$ for the other communities with the first to the tenth-largest size in each year are summarised in Additional file 1: Figs S.1 to S.5. In Figs. 4 and 5 , the red (blue) colour indicates that the value of $w_{l,k}$ is higher (lower) than the mean value of $w_{l,k}$ in each stream.

In Streams 1, 2, and 3, we can find unique relationships between the involved communities and categories. Communities in Stream 1 have a strong relationship between categories of medicine, detergent, appliances, household goods, and cosmetics almost every year. By contrast, the values of $w_{l,k}$ in Stream 1 for the categories of infomercial, medicine, detergent, and household goods became more pronounced in the later years, particularly since 2016 or 2017. A reason for this feature in the later years can be inferred as follows. Many nodes that belonged to Stream 1, which originally contains the images of not only 'woman' but also 'product', moved to the community represented by 'product' in 2017 and configured Stream 1 during 2017 to 2020. Images shared by 'product' should be associated with categories of infomercial, medicine, detergent and household goods. The heatmap for Stream 2 exhibits a completely different nature from that of Stream 1. Categories of toy, roadshow, av software, and publication are consistently salient in Stream 2, and PC and A/V gradually increase the extent of the relationship to the stream. These categories related significantly to Stream 2 were less related to Stream 1. Regarding Stream 3, which is associated with the images of 'family' and 'eat', the categories of cup noodle, pet food, food, and appliance are strongly related to the communities in the stream during almost all periods. We can observe that at approximately 2014 and 2015, the relationship of Stream 3 to the categories of beer, liquor, snack, distribution industry A, and distribution industry became strong, whereas those of detergent, household goods, and tobacco became weak. Here, distribution industry (distribution industry A) is a category that includes restaurants, retailers, and specialised stores (convenience stores, department stores, and supermarkets). Therefore, the communities in Stream 3 shifted to a stronger image of foods and drinks in later years.

Figure 5 shows the relationship $w_{l,k}$ between categories (k) and communities (l) represented by 'man' and 'woman', i.e., Streams 4 and 5, respectively. Interestingly, the value of $w_{l,k}$ tends to be large in stream 4 whereas it is small in stream 5, and vice versa. For example, although cup noodle and pet food are both food-related, cup noodle is related only to communities represented by 'man' on the whole, and pet food is related only to communities represented by 'woman'. Machine, credit card, finance and insurance, beer,

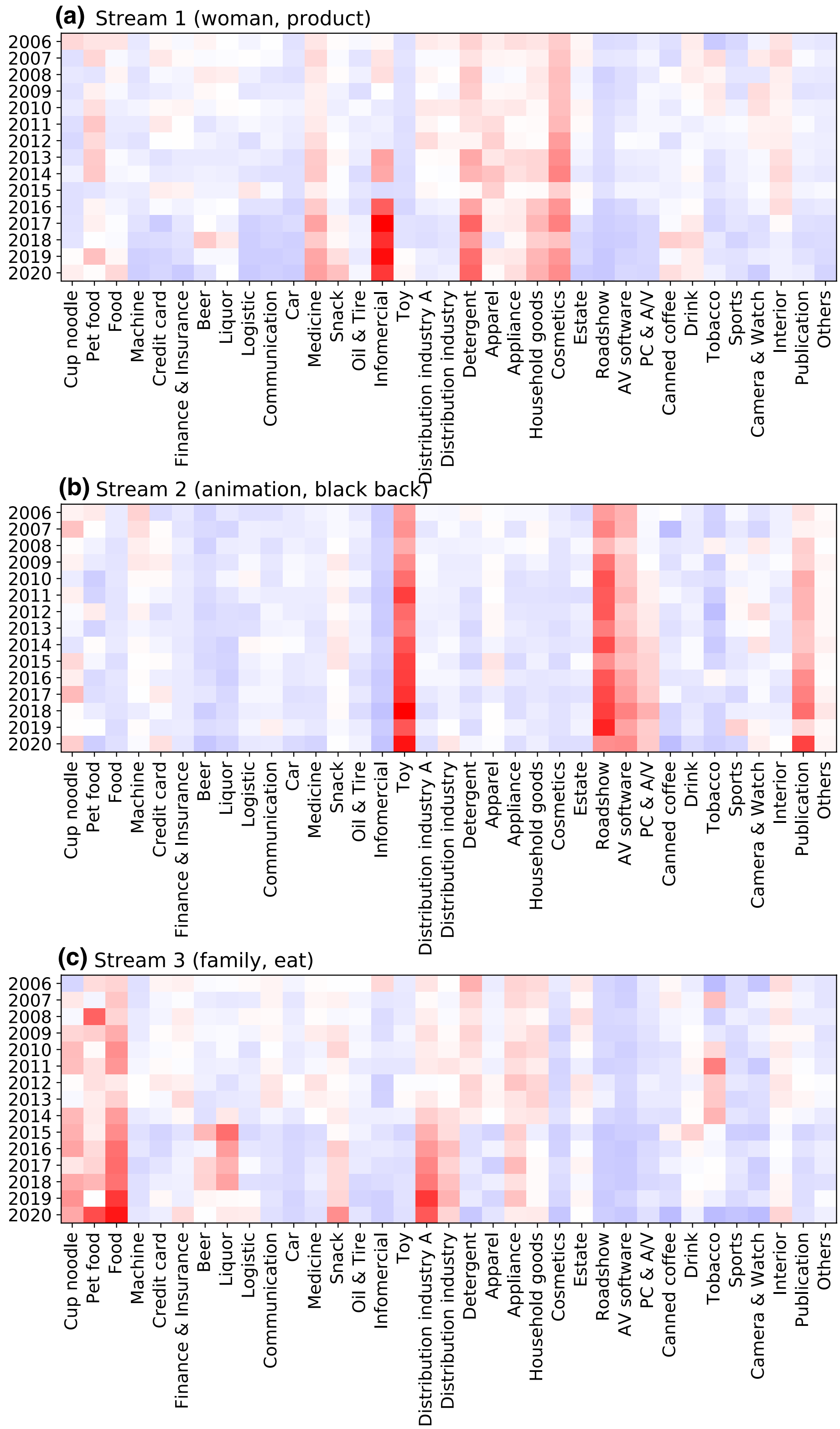

Fig. 4 Relationship $w_{l,k}$ between the categories and the communities included in **a** Stream 1, **b** Stream 2, and **c** Stream 3, for each year

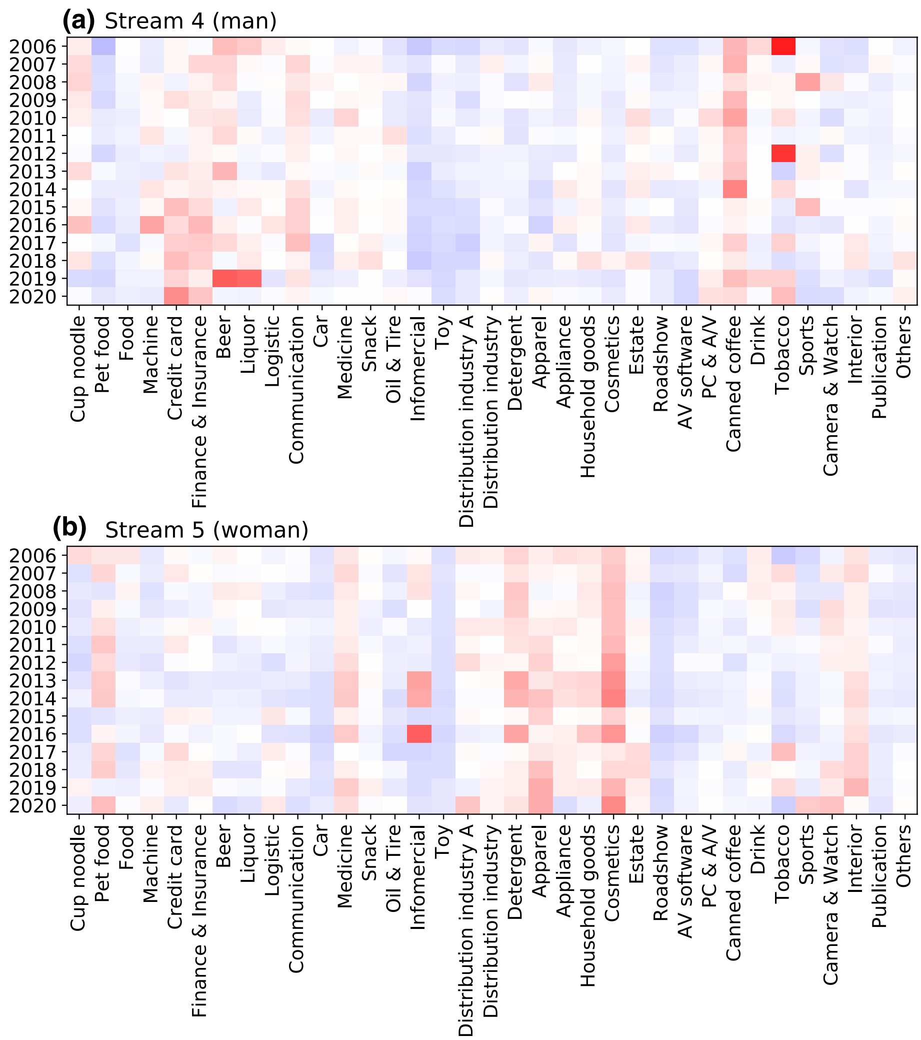

Fig. 5 Relationship $w_{l,k}$ between the categories and the communities included in **a** Stream 4 and **b** Stream 5, for each year

communication, canned coffee, and tobacco exhibit strong relationships to Stream 4 of the communities represented by 'man', whereas pet food, medicine, detergent, apparel, appliance, household goods, cosmetics, and interior are strongly related to Stream 5 of the communities represented by 'woman'.

Discussion

In the literature studying the effects of TV commercials, cultural differences in the appeal of TV commercials have been extensively investigated (Okazaki and Mueller 2007; De Mooij and Hofstede 2011). This suggests, in turn, that we can observe the culture in a country through its TV advertisements. We attempted to reveal the characteristics of the image structure produced based on the appeal in TV advertisements by representing such characteristics through a co-occurrence network of keywords. Here, each node represents a keyword, and the weight of an edge indicates the variety of products of

Table 4 Community structure

2006	Cinema scope	Woman*	Animation*	Man*	Family*	Logo	Man and woman
	3189	2760	2219	1325	1238	975	911
2007	Cinema scope	Woman*	Animation*	Man*	Product	Family*	Man and woman
	2386	2114	1891	1825	1762	1716	1175
2008	Woman*	Cinema scope	Animation*	Indoor	Man*	Girl	Dance
	2652	2566	2255	2086	1607	1044	737
2009	Woman*	Cinema scope	Animation*	Product	Man*	Family*	Sing
	2451	2382	2099	1972	1775	1702	1146
2010	Woman*	Product	Cinema scope	Animation*	Eat*	Man*	Sing
	2966	2375	2356	2335	2001	1958	1606
2011	Woman*	Eat*	Animation*	Man*	Character	Cinema scope	Sing
	3201	2409	2390	2021	1567	1528	1468
2012	Animation*	Woman*	Sing	Product	Drink	Man*	Family*
	3335	2864	2112	1962	1903	1837	1696
2013	Woman	Animation*	Man*	Family*	City	Sing	Eat
	3702	3381	2160	1984	1969	1741	1563
2014	Woman*	Animation*	Family*	Man*	City	Sing	Character
	3848	3317	2700	2196	1955	1787	1214
2015	Animation*	Woman*	Cinema scope	Eat*	Product	Man*	Sing
	4144	3394	3364	3226	3139	2940	2363
2016	Woman*	Animation*	Eat*	Man*	Black back	Sea	Character
	6320	5598	3975	2795	2768	2458	2152
2017	Animation*	Product*	Eat*	Cinema scope	Logo	Woman*	Dance
	4612	4569	3934	3824	3776	3573	2428
2018	Product*	Animation*	Eat*	Logo	Cinema scope	Woman*	Aerial view
	5369	4091	3828	3608	3546	3450	3286
2019	product*	Cinema scope	Black Back*	Logo	Man*	Eat*	Woman*
	4590	4471	4265	4143	3414	3087	2466
2020	Cinema scope	Product*	Man*	Logo	Animation*	White back	Woman*
	3147	2627	2576	2035	1832	1529	1418

Node with the highest strength in the community for each year and each community that has the first to the seventh largest size during the year. For the communities marked with a star, the other six nodes with large strengths are also shown in Tables 5 and 6

which two keywords co-occur in the same commercial. Therefore, a community can be regarded as a set of keywords that frequently co-occur in various commercials. In particular, the present study investigated how the features of such a co-occurrence network have temporally changed. Our analysis captured a temporal change of the image structure, in which the relationship between communities associated with entertainment and children and the category of PC and A/V gradually increases. By contrast, the

Table 5 Seven nodes with the largest strength in each community involved in Streams 1, 2 and 3

Stream	Year	Nodes
	2006	Woman, walk, eat, inside a store, sofa, stairs, office
	2007	Woman, indoor, photo, PC, sofa, bicycle, sit
	2008	Woman, product, drink, tell, inside a store, small screen, stairs
	2009	Woman, indoor, drink, tell, dog, mobile phone, PC
	2010	Woman, tell, indoor, walk, room, mobile phone, city
	2011	Woman, indoor, walk, room, city, PC, dog
Stream 1	2012	Woman, indoor, walk, room, inside a store, pose, stairs
'Woman'	2013	Woman, product, indoor, room, paint, cafe, stairs
'Product'	2014	Woman, product, indoor, walk, impression, stairs, paint
	2015	Woman, indoor, walk, smile, room, flower, fireworks
	2016	Woman, product, tell, indoor, walk, photo, smile
	2017	Product, tell, smile, studio, close-up a face, round frame, a word
	2018	Product, man and woman, drink, pose, surprise, tell, illustration
	2019	Product, indoor, a word, round frame, product in hand, description, studio
	2020	Product, product in hand, round frame, blue back, another product, impression, image
	2006	Animation, run, boy, C.G., fly, flame, fight
	2007	Animation, boy, run, C.G., fight, cry, flame
	2008	Animation, run, boy, illustration, fight, explosion, forest
	2009	Animation, girl, run, boy, flame, explosion, C.G.
	2010	Animation, girl, run, boy, character, card, flame
	2011	Animation, girl, boy, run, character, game screen, C.G.
Stream 2	2012	Animation, cinema scope, girl, run, boy, character, card
'Animation'	2013	Animation, cinema scope, girl, character, logo, boy, run
'Black back'	2014	Animation, cinema scope, character, logo, boy, run, card
	2015	Animation, logo, girl, boy, character, CG, game screen
	2016	Animation, run, girl, boy, game screen, character, illustration
	2017	Animation, girl, character, game screen, boy, illustration, pose
	2018	Animation, white back logo, girl, character, boy, shout, game screen
	2019	Black back, animation, white back logo, run, girl, character, boy
	2020	Animation, girl, character, boy, card, game screen, flame
	2006	Family, married couple, mother, child, daughter, father, wife
	2007	Family, eat, child, mother, daughter, father, mother and a child
	2008	Indoor, family, child, eat, photo, mother, married couple
	2009	Family, eat, child, married couple, mother, boy, girl
	2010	Eat, family, child, mother, married couple, photo, boy
	2011	Eat, family, child, married couple, mother, inside a store, girl
Stream 3	2012	Family, child, mother, girl, boy, dog, father
'Family'	2013	Family, married couple, child, mother, boy, conversation, photo
'Eat'	2014	Family, eat, married couple, mother, child, inside a store, taste
	2015	Eat, drink, family, surprise, married couple, inside a store, kitchen
	2016	Eat, family, married couple, red back, girl, child, kitchen
	2017	Eat, family, laugh, married couple, cooking, parent and child, inside a store
	2018	Eat, family, laugh, parent and child, cooking, child, kitchen
	2019	Eat, family, photo, parent and child, cooking, kitchen, inside a store
	2020	Eat, baby, field, white back,logo, cut, sleep, curry

Table 6 Seven nodes with the largest strength in each community involved in Streams 4 and 5

Stream	Year	Nodes
	2006	Man, drink, tell, train, rooftop, surprise, building
	2007	Man, tell, office, TV, look, bench, still image
	2008	Man, sea, bicycle, laugh, store, rooftop, desert
	2009	Man, TV, office, phone, class room, meeting, speak
	2010	Man, office, suited man, stairs, speak, window, question
	2011	Man, black and white, office, kimono woman, laugh, surprise, greeting
Stream 4	2012	Man, office, surprise, phone, greeting, another man, studio
'Man'	2013	Man, black and white, surprise, phone, laugh, another man, question
	2014	Man, office, conversation, phone, park, question, meeting
	2015	Man, office, conversation, laugh, phone, cafe, cry
	2016	Man, office, laugh, bicycle, greeting, phone, park
	2017	Man, another man, cat, izakaya, in a bus, Japanese room, call
	2018	Man, speak, another man, cafe, pointing, man, close-up, worried
	2019	Man, drink, laugh, surprise, speak, conversation, tell
	2020	Man, man and woman, a word, surprise, pose, speak, laugh
	2006	Woman, walk, eat, inside a store, sofa, stairs, office
	2007	Woman, indoor, photo, PC, sofa, bicycle, sit
	2008	Woman, product, drink, tell, inside a store, small screen, stairs
	2009	Woman, indoor, drink, tell, dog, mobile phone, PC
	2010	Woman, tell, indoor, walk, room, mobile phone, city
	2011	Woman, indoor, walk, room, city, PC, dog
Stream 5	2012	Woman, indoor, walk, room, inside a store, pose, stairs
'Woman'	2013	Woman, product, indoor, room, paint, cafe, stairs
	2014	Woman, product, indoor, walk, impression, stairs, paint
	2015	Woman, indoor, walk, smile, room, flower, fireworks
	2016	Woman, product, tell, indoor, walk, photo, smile
	2017	Woman, indoor, walk, office, dog, wave a hand, stairs
	2018	Woman, smile, walk, indoor, sky, dog, photograph
	2019	Woman, smile, pose, office, turn around, sit, cafe
	2020	Woman, walk, another woman, step, jump, turn around, sit

relationship between the categories and Stream 4 ('man') exhibited a different nature from that of Stream 5 ('woman'), and this feature was consistent during the period examined.

The power law in terms of degree or strength distribution has been found in networks representing various phenomena, such as social interaction and citation relationships, collaboration relationships or the co-occurrence of keywords in the academic literature (Newman 2003; Castellano et al. 2009; Karimi et al. 2019; Zhang et al. 2012). We also observed a power law or a strong heterogeneity in the co-occurrence relationship in the keywords from TV commercials. The strength of the keywords shown in Table 3, e.g. 'woman' and 'man', are large. These nodes are tightly connected to, that is, strongly associated with, other nodes through TV commercials. These nodes with high strength can be regarded as cores in the image structure produced by the TV commercials. In the later years of the examined period, the degree and strength of the keyword 'logo' became the largest, whereas they were not within the top five largest in earlier years. Studies investigating culture and TV commercials have pointed out that displaying the corporate identity logo is one of the features of Japanese commercials, where the establishment of trust between the company and consumers has a significant effect on consumer attitudes toward the products (De Mooij and Hofstede 2011). According to the data provider, the increase of the strength of 'logo' seems to be attributed to the change in their editorial policy needed considering the importance of this keyword. Temporal changes in the community structure in the co-occurrence network seem to capture the evolution of technology during the analysed period as well. Stream 2 is configured by communities represented by 'animation' (or by 'black back' in 2019) and is associated with keywords that evoke entertainment and aspects enjoyed by children. The relationship between this stream and the category of PC and A/V strengthens with each year. This presumably indicates the situation in which PC becomes a more common tool than ever before when enjoying hobbies or when children play.

Our analysis reveals not only such temporal changes, but also robust characteristics in Japanese culture. We found a significant inequality in Streams 4 and 5, configured with the communities represented by 'man' and 'woman', respectively, regarding their relationship to the various categories. Categories that had a strong relationship with the communities represented by 'man' had mostly a weak relationship to that represented by 'woman' and vice versa. The extent to which male and female roles are differentiated is an indicator characterising a culture in studies conducted on cultural differences in commercials. Our result is consistent with previous studies that showed the segregation of male and female roles in Japanese commercials, which are conducted around the year 2000 (Bresnahan et al. 2001; Milner and Collins 2000). Moreover, we found that the communities represented by 'woman' have a strong relationship to the image of 'product'. Considering the keywords included in Stream 1, e.g. 'woman', 'product', 'indoor', 'room', and 'stairs', and considering the strong relationship of the communities in this stream to the categories of medicine, detergent, household goods, cosmetics and so on, we can infer that a situation in which women actually use a product indoors is one of the significant images in TV commercials.

A strength of our study is the use of large data which records not only advertised products or aired time but also keywords representing the content in each TV commercial. The data covers all commercials aired in the Kanto area, including Tokyo, during the last 15 years. Many previous studies on cultural differences in TV commercials performed content analysis, which takes significant effort, e.g. coders rated the content of commercials according to various scales or checked the presence of items relating to the studies (Okazaki and Mueller 2007). Presumably because of this large effort, analysed commercials in the previous studies were limited to those aired in a single day or only during primetime for a week (Milner and Collins 2000; Bresnahan et al. 2001). Therefore, the findings in their analyses could be limited to the analysed seasons or period. In this study, we did not have to consider the effect of such selection bias on the results of the image structure produced by TV commercials and were able to observe an entire image underlying Japanese commercials. Although some of our findings on the characteristics of Japanese commercials are consistent with those in previous reports, the results of our study supported by such a large coverage of data should be more persuasive.

In that sense, gender role inequality, which is one of the implications of our analysis on TV commercials, should be a significant social issue. Analyses of large data have found gender inequality in various platforms and systems, such as in Wikipedia and crowdfunding, or in academic collaborations (Wagner et al. 2016; Horvát and Papamarkou 2017; Jadidi et al. 2018; Karimi et al. 2019). TV commercials are created to generate purchase intention for products or a positive attitude toward a brand; these motivations are unique to commercials and differ from those of other platforms, such as Wikipedia. That data with different reasons can reveal the same issue should be of interest to researchers. Our research should contribute to showing the diversity of large data uncovering social issues, as well as previous studies.

Here, we will discuss the technical aspects of our analysis. The resolution limit is an issue in community detection, which means that we cannot obtain communities whose size is relatively small when we conduct a graph partition through a modularity maximisation (Fortunato 2010). Therefore, the resolution limit can make us fail to find a group of nodes where the size is small but the nodes are mutually and tightly connected, which can indeed be called a 'community' in a social network. In our case, the resolution limit may not be a significant issue because what we attempted to observe is a rough image or an overview underlying the image structure produced by TV commercials, and even though we found a group consisting of a small number of nodes that are tightly connected, this might represent a trivial combination of keywords. However, as a future perspective, it would be interesting to compare the community structures obtained at various resolutions. By doing so, we may be able to observe in detail the streams of nodes among communities in years and help us better understand the detailed transition of our culture.

In addition, we may obtain further insight about the culture by analysing TV commercials in each season or area. Japanese seasons are distinct from each other, and there are many traditional events throughout the year. Moreover, the culture and value vary in different areas, and aired TV commercials are different in Japan. We may be able to find the effect of traditional or modern cultures on TV commercials by comparing commercials in various areas and in different seasons.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s41109-021-00393-4.

Additional file 1. Supplementary information.

Acknowledgements
Not applicable.

Authors' contributions
MII and TO contributed conception and design of the study; MII and TO performed the statistical analysis; MII wrote the first draft of the manuscript; MII and TO wrote sections of the manuscript. Both authors contributed to manuscript revision, read and approved the submitted version.

Funding
This work was supported by JSPS KAKENHI Grant Number JP20K19929 and JP19H01114.

Availability of data and materials
The data that support the findings of this study are available from M Data Co., Ltd. (https://mdata.tv/en/) but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of M Data Co., Ltd.

Declaration

Competing interests
The authors declare that they have no competing interests.

Received: 8 March 2021 Accepted: 29 June 2021
Published online: 15 July 2021

References

Aynaud T (2020) python-louvain 0.14: Louvain algorithm for community detection. https://github.com/taynaud/python-louvain

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech 10:10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Boyland EJ, Halford JCG (2013) Television advertising and branding. Effects on eating behaviour and food preferences in children. Appetite 62:236–241. https://doi.org/10.1016/j.appet.2012.01.032

Bresnahan MJ, Inoue Y, Liu WY, Nishida T (2001) Changing gender roles in prime-time commercials in Malaysia, Japan, Taiwan, and the United States. Sex Roles 45(1):117–131. https://doi.org/10.1023/A:1013068519583

Carreón ECA, Nonaka H, Hentona A, Yamashiro H (2019) Measuring the influence of mere exposure effect of TV commercial adverts on purchase behavior based on machine learning prediction models. Inf Process Manag 56(4):1339–1355. https://doi.org/10.1016/j.ipm.2019.03.007

Castellano C, Fortunato S, Loreto V (2009) Statistical physics of social dynamics. Rev Mod Phys 81(2):591. https://doi.org/10.1103/RevModPhys.81.591

Davtyan D, Cunningham I (2017) An investigation of brand placement effects on brand attitudes and purchase intentions: brand placements versus TV commercials. J Bus Res 70:160–167. https://doi.org/10.1016/j.jbusres.2016.08.023

De Mooij M, Hofstede G (2011) Cross-cultural consumer behavior: a review of research findings. J Int Consum Mark 23(3–4):181–192. https://doi.org/10.1080/08961530.2011.578057

Fortunato S (2010) Community detection in graphs. Phys Rep 486(3):75–174. https://doi.org/10.1016/j.physrep.2009.11.002

Gross T, Blasius B (2008) Adaptive coevolutionary networks: a review. J R Soc Interface 5(20):259–271. https://doi.org/10.1098/rsif.2007.1229

Hekkert P, Thurgood C, Whitfield TA (2013) The mere exposure effect for consumer products as a consequence of existing familiarity and controlled exposure. Acta Psychol 144(2):411–417. https://doi.org/10.1016/j.actpsy.2013.07.015

Horvát E-A, Papamarkou, T (2017) Gender differences in equity crowdfunding. In: Proceedings of the AAAI conference on human computation and crowdsourcing, vol 5. https://ojs.aaai.org/index.php/HCOMP/article/view/13319

Ito MI, Ohnishi T (2020) Uncovering the image structure of Japanese TV commercials through a co-occurrence network representation. In: Benito RM, Cherifi C, Cherifi H, Moro E, Rocha LM, Sales-Pardo M (eds) Complex networks and their applications IX. Ccomplex networks 2020. Studies in computational intelligence, vol 943. Springer, Berlin, pp 273–283. https://doi.org/10.1007/978-3-030-65347-7_23

Jadidi M, Karimi F, Lietz H, Wagner C (2018) Gender disparities in science? Dropout, productivity, collaborations and success of male and female computer scientists. Adv Complex Syst 21(03n04):1750011. https://doi.org/10.1142/S0219525917500114

Karimi F, Mayr P, Momeni F (2019) Analyzing the network structure and gender differences among the members of the networked knowledge organization systems (NKOS) community. Int J Digit Libr 20(3):231–239. https://doi.org/10.1007/s00799-018-0243-0

Liu F, Kanso A, Zhang Y, Olaru D (2019) Culture, perceived value, and advertising acceptance: a cross-cultural study on mobile advertising. J Promot Manag 25(7):1028–1058. https://doi.org/10.1080/10496491.2019.1612495

Mantegna RN (1999) Hierarchical structure in financial markets. Eur Phys J B 11(1):193–197. https://doi.org/10.1007/s100510050929

Milner LM, Collins JM (2000) Sex-role portrayals and the gender of nations. J Advert 29(1):67–79. https://doi.org/10.1080/00913367.2000.10673604

Mizokami C, Ohnishi T (2018) Revealing persistent structure of international trade by nonnegative matrix factorization. In: Cherifi C, Cherifi H, Karsai M, Musolesi M (eds) Complex networks and their applications VI. Complex networks 2017. Studies in computational intelligence, vol 689. Springer, Berlin, pp 1088–1099. https://doi.org/10.1007/978-3-319-72150-7_88

Moon YS, Chan K (2005) Advertising appeals and cultural values in television commercials A comparison of Hong Kong and Korea. Int Mark Rev 22(1):48–66. https://doi.org/10.1108/02651330510581172

Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45(2):167–256

Okazaki S, Mueller B (2007) Cross-cultural advertising research: where we have been and where we need to go. Int Mark Rev 24(5):499–518. https://doi.org/10.1108/02651330710827960

Özgür A, Cetin B, Bingol H (2008) Co-occurrence network of Reuters news. Int J Mod Phys C 19(05):689–702. https://doi.org/10.1142/S0129183108012431

Pan S (2011) The role of TV commercial visuals in forming memorable and impressive destination images. J Travel Res 50(2):171–185. https://doi.org/10.1177/0047287509355325

Pham MT, Geuens M, De Pelsmacker P (2013) The influence of ad-evoked feelings on brand evaluations: empirical generalizations from consumer responses to more than 1000 TV commercials. Int J Res Mark 30(4):383–394. https://doi.org/10.1016/j.ijresmar.2013.04.004

Radhakrishnan S, Erbis S, Isaacs JA, Kamarthi S (2017) Novel keyword co-occurrence network-based methods to foster systematic reviews of scientific literature. PLoS ONE 12(3):0172778. https://doi.org/10.1371/journal.pone.0172778

Socolar JES, Kauffman SA (2003) Scaling in ordered and critical random Boolean networks. Phys Rev Lett 90(6):068702. https://doi.org/10.1103/PhysRevLett.90.068702

Su H-N, Lee P-C (2010) Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in technology foresight. Scientometrics 85(1):65–79. https://doi.org/10.1007/s11192-010-0259-8

Wagner C, Graells-Garrido E, Garcia D, Menczer F (2016) Women through the glass ceiling: gender asymmetries in Wikipedia. EPJ Data Sci 5:1–24. https://doi.org/10.1140/epjds/s13688-016-0066-4

Watts DJ, Dodds PS (2007) Influentials, networks, and public opinion formation. J Consum Res 34(4):441–458. https://doi.org/10.1086/518527

Zhang J, Xie J, Hou W, Tu X, Xu J, Song F, Wang Z, Lu Z (2012) Mapping the knowledge structure of research on patient adherence: knowledge domain visualization based co-word analysis and social network analysis. PLoS ONE 7(4):34497. https://doi.org/10.1371/journal.pone.0034497

Publisher's Note