Journal of Computational Social Science (2019) 2:103–117
https://doi.org/10.1007/s42001-019-00039-7

RESEARCH ARTICLE

You are what you eat

A social media study of food identity

Kazutoshi Sasahara[1,2]

Received: 9 September 2018 / Accepted: 3 March 2019 / Published online: 20 March 2019

Abstract
Food preferences not only originate from a person's dietary habits, but also reflect personal values and consumer awareness. This study addresses "food identity" or the relationship between food preferences and personal attributes based on the concept of "food left-wing" (e.g., vegetarians) and "food right-wing" (e.g., fast-food lovers) by analyzing social data using information entropy and networks. The results show that food identity extends beyond the domain of food: The food left-wing has a strong interest in socio-environmental issues, while the food right-wing has a higher interest in large-scale shopping malls and politically conservative issues. Furthermore, the social interactions of food left-wing and right-wing factions show segregated structures, indicating different information consumption patterns. These findings suggest that food identity may be applicable as a proxy for personal attributes and offer insights into potential buying patterns.

Keywords Computational social science · Food identity · Marketing · Social media

Introduction

In today's age of gluttony, we are overwhelmed with information about food. Food is available at a moment's notice in supermarkets, and we continually see advertisements for food online. With almost unlimited options, the choice of what to eat and what not to eat depends less on biological aspects, such as individual survival or likes and dislikes, but rather it reflects the values of the individual [1]. Brillat-Savarin, a French gastronome famously stated, "Tell me what you eat and I will tell you what you are" [2]. This indicates that food was believed to be linked with

✉ Kazutoshi Sasahara
sasahara@nagoya-u.jp

1 Department of Complex Systems Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

2 JST, PRESTO, Kawaguchi, Japan

identity. Although several studies have examined this aspect (e.g., [3]), much remains understudied.

There are terms typifying the relationship between food and political ideologies that appear in sociological discourse in American culture, such as "Starbucks people" and "Coors beer people" [4]. These terms are meant as ironic signifiers of the metropolitan intelligentsia (liberal and Democratic Party-supporting people), who buy expensive coffee and read the New York Times at Starbucks versus rural people (conservative and Republican Party-supporting people), who drink inexpensive canned Coors beer while watching live broadcasts of American football. Other terms that similarly parody these two camps' lifestyles and political ideologies are "latte liberals" and "bird-hunting conservatives" [5].

In Japan, it remains to be seen whether such stark stereotypes would apply, but there is prior work describing the political and dietary sensibilities of Japanese people using the terms "food left-wing" and "food right-wing" [6]. Proponents of "food left-wing" are those who pursue natural food and are health-conscious, typically vegetarians and vegans. In contrast, the "food right-wing" group generally consumes available food products and enjoys eating fast food. Hayamizu (2013) describes how food preferences (food left-wing and right-wing) speak to political attitudes in Japan.

If food preferences reflect the values of people, then these preferences are also likely closely tied not only to political ideology but also to other personal attributes. If that were the case, food preferences could be a "mirror" reflecting latent consumption preferences and attitudes.

We study "food identity" based on the concept of the food left-wing and right-wing in order to determine whether this concept is useful in gaining insights into personal attributes. To this end, we analyze social data from Twitter, in which a large amount of food-related information is spontaneously posted and shared. Twitter is a better data source for our purse, because its use is better motivated in Japan given the high level of penetration in the country [7]. Although there are many previous researches about food-related social data analyses [8–10], little attention has been paid to food identity as a proxy for personal values and consumer awareness. This is the main focus of this study.

Data and methods

Data collection

The official Twitter Search application programming interface (API) [11] was used to create a crawler to harvest social data from the site and collected two datasets (Fig. 1).[1] We ran the crawler three times a day and thus obtained almost all of the tweets that contained keywords of interest described below.

[1] Alternative choice could be the Twitter Streaming API, but it has several issues: e.g., It often does not work properly for non-space separated languages (e.g., Japanese) and it is limited to 1% of the full Twitter Firehose. Thus, we decided to use the Twitter Search API.

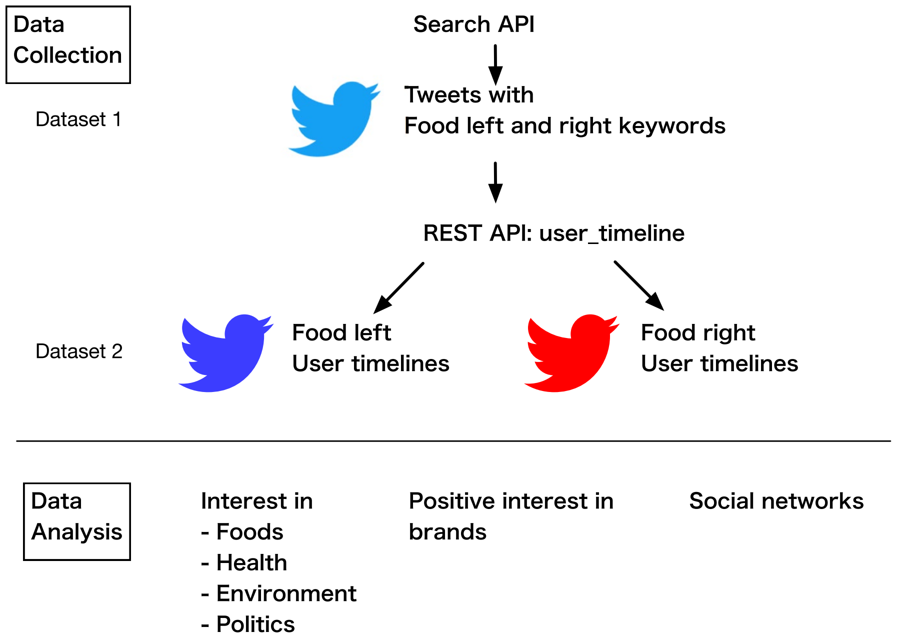

Fig. 1 Schematic illustration of data collection and analysis

Table 1 Keywords used in collecting Dataset 1

Food left	地産地消 (local production and consumption),スローフード (slow food), 道の駅 (roadside markets), ベジタリアン (vegetarian), 有機農法 (organic farming), オーガニック (organic),マクロビ (macrobiotic), ベジフェス (vegetable festivals),ファーマーズマーケット (farmers' markets)
Food right	ファストフード / ファーストフード (fast food), ジャンクフード (junk food), 冷凍食品 (frozen foods), メガ盛り (super-sized), デカ盛り (massive portions), 栄養ドリンク (energy drinks), コンビニ (convenience stores), 揚げ物 (fried foods)

We used 18 food-related keywords from Shimizu (2013) to identify the food left-wing and right-wing (see Table 1). Tweets containing these keywords were collected for a period of approximately one month beginning in December 2016. This process yielded 650,900 Japanese tweets containing keywords referencing food left-wing tendencies, and 3,141,527 Japanese tweets containing keywords referencing food right-wing tendencies (Dataset 1). Dataset 1 was then used to identify users making 30 or more total tweets containing any food left-wing keywords (using any of the keywords once or more per day, on average). The same process was performed for those making tweets containing food right-wing keywords. We excluded users making 30 or more total posts containing both food left-wing and right-wing keywords because they were deemed to not have a specific food preference. This process yielded 1233 users making food left-wing tweets and 5010 users making food right-wing tweets. This research treated the former group as food left-wing and the latter as food right-wing.

Timelines for all users identified with Dataset 1 were obtained: 3,655,936 tweets from the food left-wing and 15,091,255 tweets from the food right-wing (Dataset 2).

Data analysis

We computed the frequencies of keywords listed in Table 1 and later using texts from Dataset 2. To compare the food left-wing and right-wing groups, we constructed 10 bootstrapped samples in each group with 1000 resamplings, used for the measurements described below.

Measuring collective interests

The frequency of a keyword is often used as an indicator of the aggregate level of interest in a topic on social media, but this poses one major problem. It does not consider how many individuals actually used the keyword. For example, in an extreme case, the most frequently used keyword could be used by a single user. To avoid this, we measure keyword entropy (H_k) below as an indicator of collective interest in a keyword by using normalized entropy [12]:

$$H_k = -\sum_{x \in U} P_k(x) \log_2 P_k(x) / \log_2 N.$$

Here, U refers to a set of users, with $P_k(x)$ being the probability of a post with the keyword k made by user x, and N is the total number of users, with $0 \leq H_k \leq 1$. As is clear in the definition, when many users make posts including keyword k, then H_k increases.

To compare keyword entropy between the food left-wing and right-wing groups, we employ the following formula, which is often called the laterality index (LI):

$$\mathrm{LI} = \frac{\mathrm{H}_{\mathrm{k}}^{\mathrm{R}} - H_{\mathrm{k}}^{\mathrm{L}}}{\mathrm{H}_{\mathrm{k}}^{\mathrm{R}} + H_{\mathrm{k}}^{\mathrm{L}}}$$

where R represents the food right-wing, and L represents the food left-wing with $-1 \leq LI \leq 1$. For keyword k, a larger LI implies a greater degree of collective interest among food right-wing users; a smaller LI implies a greater degree of collective interest among food left-wing users.

Visualizing consumer awareness

We use association networks [13] to examine how the food left-wing and right-wing groups are aware of various keywords. First, the tweet texts from Dataset 2 are segmented into words using MeCab [14] and the mecab-ipadic-NEologd (a Japanese dictionary) [15], and then cleaned by removing stopwords (defined in SlothLib [16]), symbols (e.g., ! and @), and URLs. Next, the cleaned texts are used as the input to a word embedding method called word2vec [17, 18]. Word2vec can construct lower dimensional vectors that reflect word meanings based on word usage. Using the trained word2vec model, we can convert words used within a corpus into vectors, by which semantically similar words would become similar vectors. We used the Gensim library [19] for the word2vec modeling with the default parameter setting.

The resulting word vectors are visualized using association networks in the following manner. Given a seed word, we list words whose cosine similarity to the seed word vector is greater than the similarity threshold (0.4); we then selected the top 20 most similar words. Using the selected words as new seeds, we listed words in a same manner. Words obtained by such "association chains" are used as nodes. If term w_2 is selected when term w_1 is a seed, w_1 and w_2 are linked. If multiple words are selected when w_1 is a seed, then w_1 is connected to all these words. In this way, we visualize consumer awareness as word associations in tweets.

Social interactions in retweets

Twitter has the functionality of reposting or retweeting a friend's posts to own followers. This leaves a record of how information spreads among users. Given that user B retweets a post from user A, A and B can be treated as a node, with the directed link A → B. All of the retweets in Dataset 2 can be turned into directed links in this way, allowing for the reconstruction of social interactions in retweets among food left-wing and right-wing users. We refer to this as a retweet network, described as $G = (V, E)$ [20]. Here, V is a set of the users listed using the above method, and E is a set of links describing retweet transmissions. V includes users other than the left-wing and right-wing seed users in Dataset 2, but those never retweeted were not included. Analyzing the retweet network (G) allows us to examine structural patterns of information transmissions within and between food left-wing and right-wing groups.

Results

Collective interest in food and other keywords

First, we confirmed whether food left-wing and right-wing users had different food preferences and whether they had other preferences outside of food. We computed the frequencies of the keywords in Table 2 for the categories of food, health, socio-environmental issues, and politics that were featured in the Nikkei newspaper from 2015 to 2016 (except for meat, fish, and vegetables).[2] The resulting word frequencies were then used to compute keyword entropy to compare the degree of interest in each group.

[2] Note that a health freak is a person extremely enthusiastic about health.

Table 2 Keywords related to food, health, socio-environment, and politics

Food	肉 (Meat), 魚 (fish), 野菜 (vegetables), 遺伝子組み換え (GM (genetically modified)), トランス脂肪酸 (trans fatty acid), カップヌードル (instant noodles)
Health	高カロリー (High calorie), 低カロリー, 健康オタク (health freak), 無農薬 (Agrochemical-free), ジョギング (jogging), 低脂肪 (low fat)
Socio-environmental	温室効果ガス (Greenhouse gases), エコ (eco), フェアトレード (fair trade), 動物実験 (animal experiment), 環境保護 (environmental protection), 偽装食品 (food fraud)
Politics	安倍首相 (Prime Minister Abe), ネトウヨ (online right-wingers), リベラル (liberal), 保守 (conservative), ヒラリー (Hilary), トランプ (Trump)

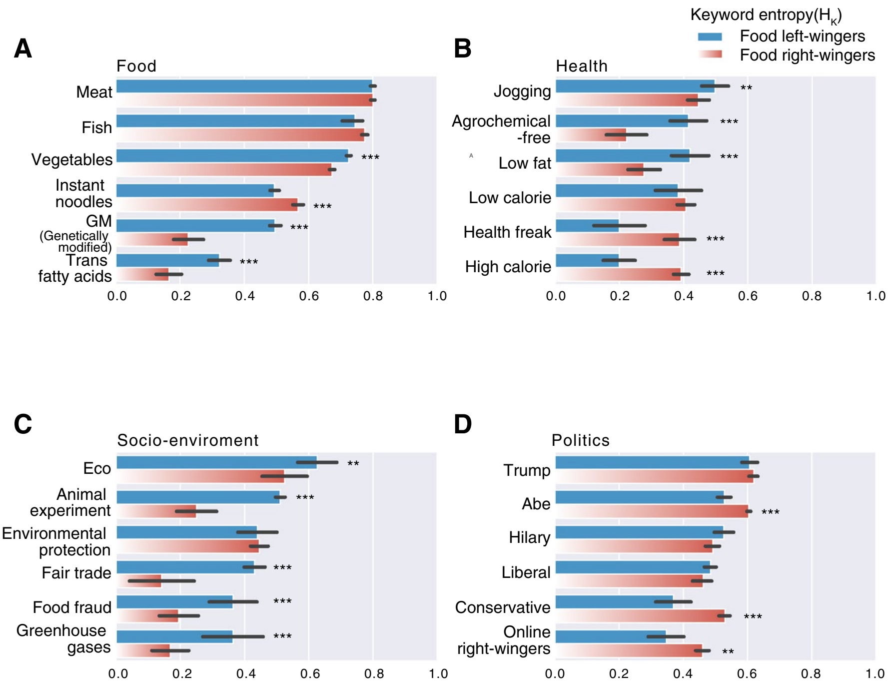

Fig. 2 Degree of collective interest in keywords related to food, health, socio-environment, and politics (means and SDs computed from the bootstrapped samples)

Figure 2 shows keyword entropy across different categories. In Fig. 2a, there is a high degree of interest in major foods, such as "meat," "fish," and "vegetables." Comparing the food left-wing and right-wing groups, there is no marked difference in meat and fish, but the right-wing group shows a higher degree of interest in vegetables. Of considerable interest here is that the food left-wing showed a greater interest in keywords like "trans fatty acid" that can lead to heart disease and "GM (genetically modified)" crops such as soybeans and corn in which the produce is manipulated by humans. In contrast, the food right-wing showed a

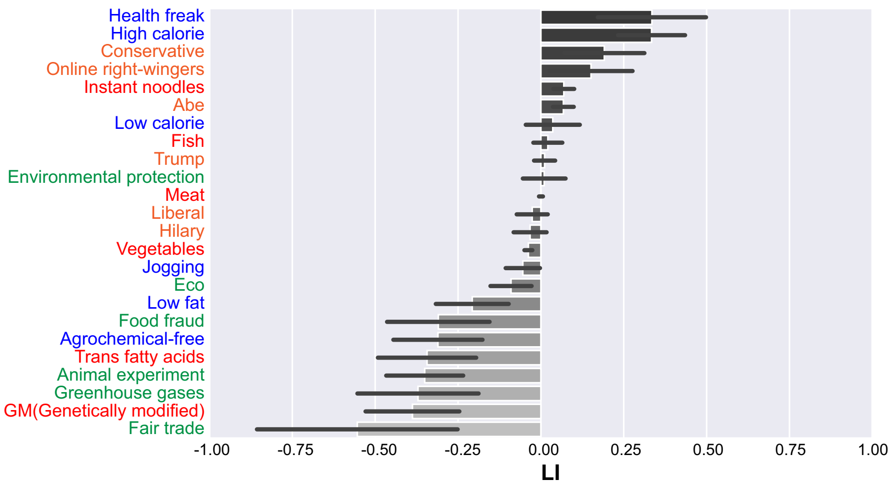

Fig. 3 Collective interest in all keywords (means and SDs computed from the bootstrapped samples). In the PDF version of this paper, colors correspond to categories (Food: red, Health: blue, Socio-environment: green, Politics: orange)

higher interest in "instant noodles," the standard-bearer for junk food. To confirm in what context these keywords were used, we randomly sampled 200 tweets that included each of these keywords, and then manually checked whether these keywords were used in a positive or negative context. It transpired that the food left-wing users mostly made negative statements about trans fatty acid (positive: 1, negative: 199) and genetically modified (GM) crops (positive: 4, negative: 176); instant noodles were mostly used in a positive context by the food-right (positive: 131, negative: 29). These results align with the image of the food left-wing as liking natural food and the right-wing as liking fast food.

Next, we examined the two groups to determine their preferences for categories other than food. This can be confirmed in Figs. 2b–d when looking at categories of health, socio-environment, and politics. In the health category, the food left-wing had a strong interest in "jogging," "agrochemical-free," and "low fat" while the food right-wing had a strong interest in "high calorie," and "health freak." This brings to mind the image of food right-wing users as eating fast food while worrying about the calories associated with it. Given that food left-wing users show a greater interest in most of the keywords in the socio-environmental category, we can see how this concurs with the image of food left-wing users as highly conscious about socio-environmental issues. Interestingly, the food right-wing showed a strong interest in keywords like "conservative" and "online right-wingers" and "Abe" (the current prime minister of Japan, seen as a right-wing politician). Thus, there is some degree of correlation between food and politics. In contrast, such correlation was not observed in the food left-wing.

Figure 3 summarizes the results of the keyword entropy when plotted by the laterality index (*LI*), showing specific differences between the two groups in categories

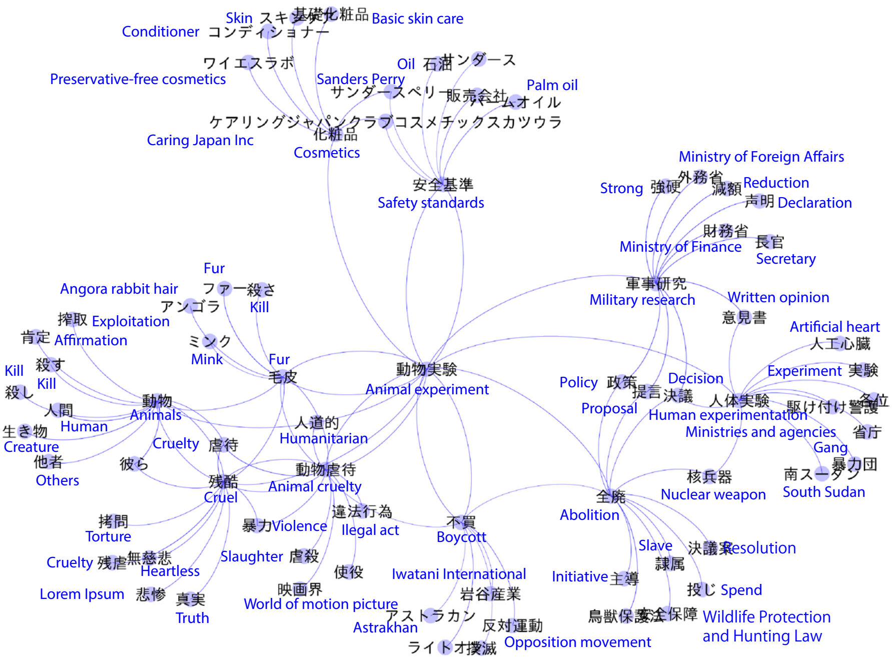

Fig. 4 Association network of the term "animal experiment" for food left-wing users

that go beyond food. Again, a greater *LI* implies a greater degree of interest among food right-wing users, while a smaller *LI* implies a greater degree of interest among food left-wing users. We will not repeat the same observations shown in Fig. 2. Rather, we emphasize that the food left-wing group had a marked interest in socio-environmental issues and the food-right showed some interest in political issues.

It is notable that the term "animal experiment" is of high interest among the food left-wing. Using association networks [13], we examined whether this is indeed the case. Figure 4 is an association network of the words obtained when using "animal experiment" as the seed word when mapping against food left-wing users. The links show word associations with a cosine similarity of 0.4 or higher. The term "animal experiment" directly ties to "animal cruelty" and then to word clusters comprising terms like "cruel," "torture," "fur," and "mink." Other notable connections seeded from "animal experiment" include "boycott," "cosmetics," and "military research." This association network reveals how food left-wing users harbor a strong negative image toward animal experiments. We conducted the same process on the food right-wing data, but there was no markedly negative association with the term "animal experiment."

These findings indicate that the food left-wing has strong feelings about animal welfare and strongly shirks from the idea of artificially manipulating animals. Therefore, we can plainly see the way in which consumer attitudes, such as refusing to buy cosmetics that have been tested on animals or refusing to buy mink coats that are the result of killing and skinning animals, are manifested in tweets.

Table 3 Industries and major brands used

Technology	Apple, Google, Microsoft, Facebook, Amazon, Sony
Retail	McDonalds, Starbucks, IKEA, Costco, KFC, UNIQLO
Beverage	Pepsi, Coca-Cola, Corona, Red Bull, Budweiser, Heineken
Automobile	Toyota, BMW, Mercedes, Honda, Volkswagen, Nissan

Table 4 Positive keywords

ほしい/欲しい (want), 買った/購入した (bought), 格好いい (cool), 良い (nice), すばらしい/素晴らしい (awesome), 好き (like), 美味い/うまい (tasty), 美味しい/おいしい (delicious), ほしい/欲しい (want), 素晴らしい/すばらしい (excellent), 素敵 (sweet), 最高 (best)

Consumer interests in brands

Next, we present an investigation into consumer interests in specific brands. We used a 2016 Forbes survey [21] and a 2016 Nikkei survey [22] to identify 24 global brands across four industries (Table 3). The keyword entropy (H_k) is measured based on the co-occurrence of brand names and positive words (Table 4) in Dataset 2. In this case, a larger H_k implies more positive interest in the brand.

Figure 5 shows the degree of positive interest in brands across the categories of technology, retail, beverage, and automobile, comparing between the food left-wing and food right-wing. Figure 6 summarizes these results in terms of laterality index (*LI*).

The food right-wing group preferred most of the technology firms except for Apple and Sony.[3] In the retail space, the food right-wing showed a markedly higher interest in IKEA and Costco. Incidentally, Starbucks—which we mentioned in the introduction—showed an expected high level of interest among the food left-wing. Further differences were seen in the beverage space where the food left-wing showed strong interest in overseas beer brands like Corona, while the food right-wing showed high interest in energy drink brands like Red Bull. A counter intuitive finding is that the food left-wing's positive interest in soda, especially in Pepsi. We observed positive tweets, such as "Pepsi is way tastier than Coke. As for Guarana, it's a bit in the middle..." The food-left wing showed a higher interest prefers in most of the automobile brands except Toyota.

The differences in positive interest between the food left-wing and right-wing groups are a useful metric when determining which products to promote through advertising based on the unique preferences of each group.

[3] Given that most posts about Amazon concern online shopping, Amazon might be more appropriately interpreted belonging to the retail category rather than technology.

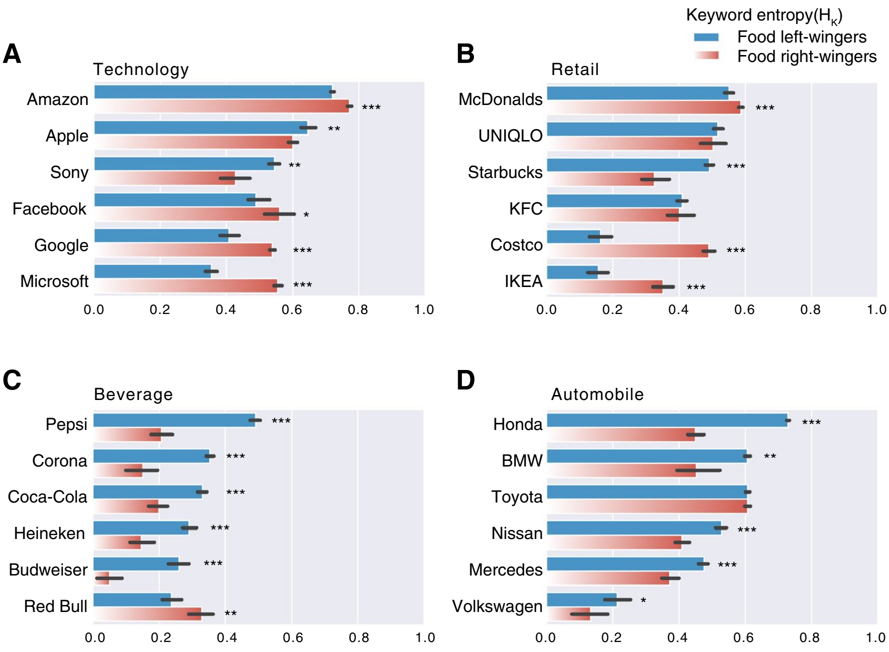

Fig. 5 Positive interest in brands related to technology, retail, beverage, and automobile (means and SDs computed from the bootstrapped samples)

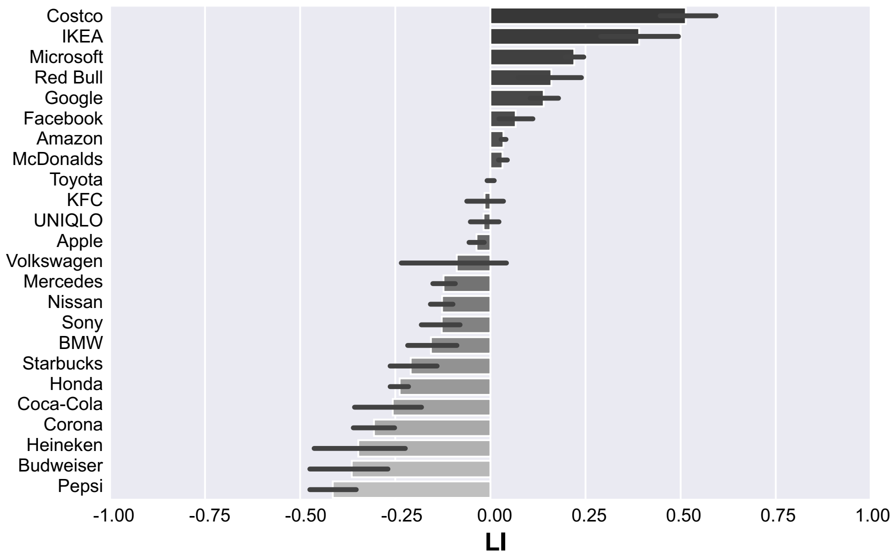

Fig. 6 Positive interest in brands (means and SDs computed from the bootstrapped samples)

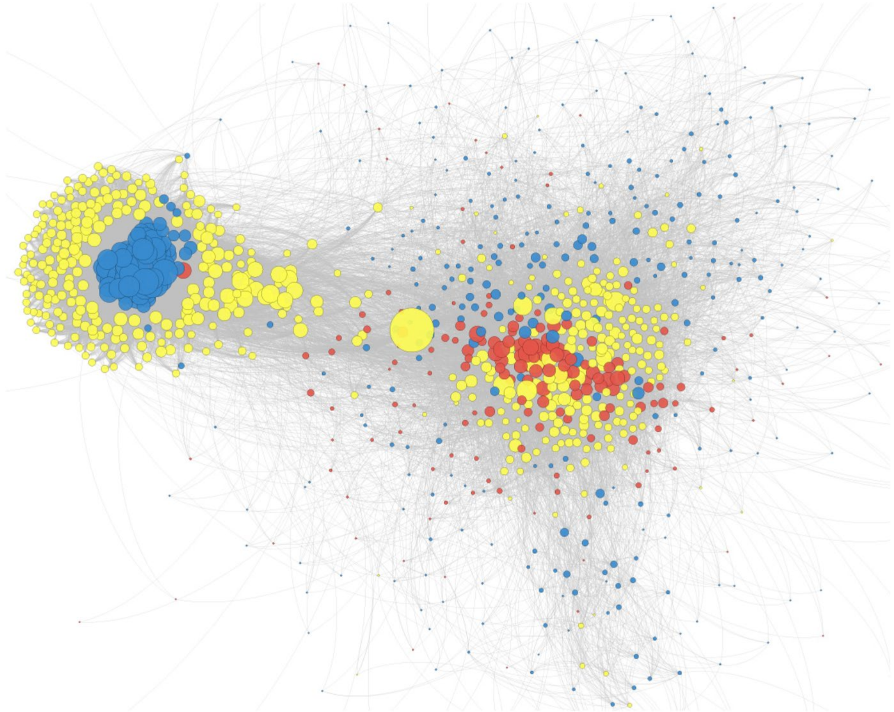

Fig. 7 Retweet networks of food left-wing and right-wing users. Nodes denote users. Blue corresponds to the food left-wing, and red corresponds to the food right-wing; yellow is unknown. The node size is proportional to overall degree. The links denote retweet transmissions

Network of social interactions in food left-wing and right-wing users

As described previously, we created a retweet network (G) based on the propagation of retweets ($|V| = 165,609$, $|E| = 382,899$). Figure 7 visualizes the largest connected components with a maximum of 50 or more orders of connection ($|V| = 1,113$, $|E| = 33,231$). Nodes represent users with blue being the food left-wing and red being the food right-wing. Yellow represents those whose orientation could not be ascertained (i.e., users other than the left-wing and right-wing users in Dataset 2). The links represent retweet transmissions. One salient feature of these networks is the formation of two distinct left and right clusters. Unlike the food left-wing users scattered on the right of Fig. 7, those on the left side of the graph suggest that the food-left users actively communicate about organic food-related information via retweets, thereby forming a cohesive cluster. In contrast, the food right-wing does not manifest as densely as the food left-wing, suggesting retweet communications are less cohesive.

Another notable feature in Fig. 7 is the existence of a "bridge" (the large yellow node) that links two separate clusters. The bridge has major potential influence because this class of users can convery information to both the food left-wing and right-wing users. This account belongs to livedoor News (@livedoornews) that manually curate and reposte articles posted to the livedoor News website. Since livedoor News has a food section, this account makes many posts about food and cuisine; some are preferred by the food left-wing and others by the food right-wing.

Discussion

We have demonstrated differences in personal values and consumer awareness for left-wing and right-wing users via social data analysis. Food preferences are not a binary of left or right but rather a continuum, as personal attributes are multidimensional in nature. This research, however, implies that mapping the plurality of personal attributes across the single dimension of left and right and looking at these two extremes does offer useful insights. Measuring the degree of collective interest in certain keywords and brand names revealed that beyond the domain of food, the food left-wing has a strong interest in socio-environmental issues (notably, animal experiments) and the food right-wing has a higher interest in large-scale shopping malls offering discounts and volume sales of goods, such as IKEA and Costco. In addition, we observed differences between the food left-wing and right-wing groups in word usage. Table 5 shows the top 100 popular keywords ranked by the average TF-IDF score [23]. Note that the word's TF-IDF score is its term frequency divided by its document (tweet) frequency. Many food left-wing related words are ranked in the table in the food left-wing group (e.g., beauty, health, and nature), while several food-right wing related words are ranked in the table in the food right-wing group (e.g., supermarket and ice cream).

The segregated network structures in retweet interactions revealed that food-related information is also consumed differently in the food left-wing and right-wing groups. One point of interest in these networks is that clusters of food left-wing and right-wing users are linked via the conduit of a news outlets (livedoor News). Given that online social networks are mechanisms for disseminating information, identifying users' food preferences might be efficient in the distribution of a specific piece of information, an advertisement, or a campaign.

Looking solely at these results, one could conjecture as to the differences in consumer attitudes among these groups—the food-left wing rejects industrialization and prefers a return to nature, and the food right-wing does the opposite. However, reality is not quite so simple. Indeed, the food left-wing had a higher degree of interest in Apple and Sony as technology brands, as well as in most automobile brands. Therefore, we cannot say conclusively that the food left-wing comprehensively rejects industrialization and prefers a return to nature in contexts other than food.

This study has implications for social sciences and applications. The quantification of personal attributes is an important procedure in many areas of social sciences. Personal attributes, however, are difficult to accurately measure by survey research alone. Given the fact that food identity could be a concise yet useful proxy for various personal attributes and that there are numerous reports about food on social media, mining food identities from online digital traces can contribute to

Table 5 Top 100 popular keywords ranked by the average TF-IFD socre

Rank	Food lef-wing		Food righ-wing		Rank	Food lef-wing		Food righ-wing	
1	プレゼント	present	お前	you	51	必要	necessity	先輩	elder
2	キャンペーン	campaign	好き	love	52	一番	first	ご飯	rice
3	フォロー	follow	くん	Mr	53	ラーメン	rahmen	人生	life
4	ブログ	blog	画像	picture	54	成分	component	学校	school
5	日本	Japan	仕事	work	55	iherb	iherb	人見知り	shyness
6	応募	application	店員	clerk	56	女性	female	男性	male
7	好き	love	最近	recently	57	実施中	in operation	ニュース	news
8	商品	goods	ダイエット	diet	58	バイク	motorcycle	ブログ	blog
9	ツイート	tweet	日本	Japan	59	ケア	care	お知らせ	information
10	情報	information	あるある	possible	60	ニュース	news	普通	normal
11	無料	no charge	意味	meaning	61	今年	this year	無理	impossible
12	詳細	details	女性	female	62	利用	use	ダメ	no good
13	美容	beauty	バイト	part-time job	63	あるある	possible	料理	food
14	ダイエット	diet	定期	period	64	限定	limited	一人	alone
15	リツイート	retweet	メール	mail	65	現在	now	写真	picture
16	写真	picture	一番	first	66	コスメ	cosmetics	後ろ	back
17	抽選	lottery	言葉	language	67	是非	right and wrong	結果	result
18	人気	popular	まとめ	summary	68	簡単	easy	人気	popular
19	お願い	wish	一緒	together	69	募集	recruitment	理由	reason
20	更新	renewal	先生	teacher	70	話題	topic	幸せ	happy
21	動物	animal	大丈夫	safe	71	意味	meaning	食事	meal
22	紹介	introduction	女の子	girl	72	入り	entering	ゲーム	game
23	セット	set	世界	world	73	食品	food	誰か	anyone
24	世界	world	無料	no charge	74	最高	great	注意	attention
25	野菜	vegetables	人間	human	75	香り	smell	紹介	introduction
26	健康	health	友達	friend	76	相互フォロー	mutual follow	お腹	belly
27	開催	opening	電話	telephone	77	ok	ok	更新	renewl
28	自然	nature	電車	train	78	通販	mail order	募集	recruitment
29	効果	effect	子供	child	79	有名	famous	東京	Tokyo
30	販売	selling	お願い	wish	80	登場	arrival	女子	girl
31	動画	moving image	問題	problem	81	twitter	twitter	パン	bread
32	北海道	Hokkaido	絶対	obligatory	82	シャンプー	shampoo	二人	pair
33	カフェ	coffee	情報	information	83	配合	combination	発売	sell
34	記念	memory	動画	movie	84	温泉	hot spring	会社	company
35	使用	use	オレ	me	85	予約	reservation	生活	life
36	参加	participation	必要	necessarily	86	食事	meal	彼氏	boyfriend
37	楽天	rakuten	部屋	room	87	東京都	Tokyo	じゃなくて	not
38	場所	location	商品	goods	88	試し	trial	大好き	love
39	おすすめ	recommended	相手	partner	89	言葉	language	最後	last
40	クリスマス	Christmas	お金	money	90	まとめ	summary	ラーメン	rahmen
41	yahoo	yahoo	気持ち	feeling	91	定期	period	携帯	mobile phone
42	料理	cooking	弁当	box lunch	92	大丈夫	safe	場所	place
43	送料	carriage	w w w	w w w	93	ワイン	wine	テレビ	television
44	東京	Tokyo	会話	conversation	94	画像	picture	チケット	ticket
45	投稿	contribution	風呂	bath	95	周年	anniversary	ネット	net
46	イベント	event	効果	effect	96	全国	national	スーパー	supermarket
47	応援	support	トイレ	toilet	97	気持ち	feeling	最高	great
48	人間	human	簡単	easy	98	問題	problem	コーヒー	coffee
49	仕事	work	さっき	some time ago	99	大阪	Osaka	アイス	ice cream
50	最近	recently	フォロー	follow	100	完了	completion	結婚	wedding

social science research. Furthermore, the idea of food identity could be applicable in social media marketing and other applications. For example, rather than relying on influencers [24], which represent a small and limited population, or on simply trafficking large quantities of advertisements at random, selectively running advertisements based on the underlying values around food is much more likely to hit the mark. However, one should remember that simply extrapolating the concepts of the food left-wing and right-wing without conducting prior research would prove detrimental when doing social sciences or engaging in marketing.

We recognize the limitations of social media analysis for studying food identity, because social media users are biased toward age, gender, geolocation, etc. To supplement social data analysis, we have to incorporate survey research, which is one of our future research directions. Moreover, the keyword-based food preference identification done here has potential concerns with misclassification. For example, food left-wing users who often post criticisms about fast food could be classified into food right-wing users; similar errors could occur for food right-wing users. In addition, this approach cannot capture criticisms and cynicisms, potentially conflating very different types of contexts. To resolve these issues, contextual information needs to be considered to improve the classification accuracy. Although several limitations remain, our social media analysis has demonstrated that food identity can be a useful concept to address personal attributes and consumer awareness.

Acknowledgements K.S. thanks to M. Karasawa and K. Hioki for discussions. This research was supported in part by Yoshida Hideo Memorial Foundation, JSPS/MEXT KAKENHI Grant Numbers JP16K16112 and JP17H06383 in #4903, JST PRESTO Grant Number JPMJPR16D6, and JST CREST Grant Number JPMJCR17A4.

References

1. Guptill, A., Copelton, D., & Lucal, B. (2017). *Food and society: Principles and paradoxes*. Hoboken: Wiley.
2. Brillat-Savarin, J., & Fisher, M. (2009). *The physiology of taste: or Meditations on transcendental gastronomy*. New York: Knopf Doubleday Publishing Group.
3. Laufer, P., Wagner, C., Flöck, F., & Strohmaier, M. (2015). In *Proceedings of the ACM Web Science Conference (WebSci '15)*, pp. 3:1–3:10.
4. Masahito, W. (2008). *Mienai America (Invisible America) (Koudansya [published in Japanese]*. Japan: Tokyo.
5. DellaPosta, D., Shi, Y., & Macy, M. (2015). Why do liberals drink lattes? *American Journal of Sociology*, *120*(5), 1473.
6. Hayamizu, K. (2013). Fudo Sayoku To Fudo Uyoku (Food Left, Food Right) (Asahi Shinbunsya [published in Japanese]. Japan: Tokyo.
7. Hale, S. A. (2014). In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems (CHI '14)*, pp. 833–842.
8. Fried, D., Surdeanu, M., Kobourov, S., Hingle, M., & Bell, D. (2014). In *2014 IEEE International Conference on Big Data*, pp. 778–783.
9. Abbar, S., Mejova, Y., & Weber, I. (2015). In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI'15)*, pp. 3197–3206.
10. Mejova, Y., Haddadi, H., Noulas, A., & Weber, I. (2015). In *Proceedings of the 5th International Conference on Digital Health 2015* (ACM, New York, NY, USA), DH '15, pp. 51–58.
11. https://dev.twitter.com/rest/public/.
12. Tria, F., Loreto, V., Servedio, V. D. P., & Strogatz, S. H. (2015). *Scientific Reports*, *4*, 1.
13. Sasahara, K. (2016). *New Genration Computing*, *34*, 323.

14. http://taku910.github.io/mecab/.
15. https://github.com/neologd/mecab-ipadic-neologd/.
16. http://www.dl.kuis.kyoto-u.ac.jp/slothlib/.
17. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). In *Proceedings of the International Conference on Learning Representations (ICLR 2013)*.
18. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Advances in Neural Information Processing Systems, pp. 3111–3119.
19. https://radimrehurek.com/gensim/.
20. Sasahara, K., Hirata, Y., Toyoda, M., Kitsuregawa, M., & Aihara, K. (2013). *PLoS One*, *8*(4), e61823.
21. Forbes. The Worlds Most Valuable Brands. http://www.forbes.com/powerful-brands/list/. Accessed Aug 2 2016.
22. Nikkei. Kakusa, Brandoryoku, ... Asia, Seicho De Mieta Genjitsu [published in Japanese]. https://vdata.nikkei.com/prj2/asia-economy2/. Accessed 2 Aug 2016.
23. Manning, C., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge: Cambridge University Press.
24. Keller, E., & Berry, J. (2003). *The influentials: One American in ten tells the other nine how to vote, where to eat, and what to buy*. New York: Free Press.