An Analysis of Japanese Consumer Consciousness on Mislabeling Food: Derived from a Consumer Questionnaire on Kuruma Prawns and Black Tiger Prawns

Masashi MOCHIZUKI
Technology Research Department; Kyoto Prefectural Agriculture, Forestry and Fisheries Technology Center, mocchy@mcn.ne.jp

Taro OISHI
Fukuoka Institute of Technology, t-oishi@fit.ac.jp

ABSTRACT

Since October 2013, consecutive cases of food mislabeling, such as Black Tiger Prawns being sold under the name of Kuruma Prawns in famous restaurants, department stores, etc., have begun to surface in Japan, and this has become a social problem. This study conducted a consumer questionnaire in free-answer format in November 2013 with the objective of investigating what kinds of images consumers currently held of Kuruma and Black Tiger Prawns directly after the revelation of mislabeling.

By applying text-mining analysis to the responses acquired from the questionnaire, the authors (1) extracted frequently occurring words using simple tabulation analysis, and (2) gained an understanding of the relationships between words using co-occurrence network analysis to illuminate consumer images of Kuruma and Black Tiger Prawns.

INTRODUCTION

Nowadays, consumers are stricter in their judgment of food safety and security. Nevertheless, since October 2013, consecutive mislabeling problems have begun to surface in Japan. For example, Black Tiger Prawns have been sold under the name of Kuruma Prawns, and Whiteleg Shrimp have been sold under the name of Shiba Shrimp, by famous restaurants, department stores, etc., and this has become a social problem. Gaining an understanding of consumer images of the aforementioned foods associated with mislabeling is important from the perspective of the impact of mislabeling on consumer awareness. However, in existing studies, although some analyses concern consumer awareness surveys regarding the purchase of beef, pork, and chicken (Yokota et al., 2010) and awareness surveys relating to the image of seafood (Ogawa et al., 2003), etc., to the best of our knowledge, there is no research that analyzes consumer images of mislabeled food.

In this research, the authors conducted a consumer questionnaire concerning consumer images of food that had been subject to mislabeling. In this questionnaire, which focused on the most recently-occurring case involving Kuruma Prawns and Black Tiger Prawns (hereafter “BT Prawns”), consumers were asked to respond in a free-answer format regarding what kinds of images the words Kuruma Prawns and BT Prawns evoked. In this research, the Authors applied text-mining analysis to these responses in the free-description field and analyzed consumer images of Kuruma and BT Prawns.

METHOD

The Data

In this research, text-mining analysis was applied to the response text in the free-description field of the "Questionnaire Survey on Images of Marine Products" conducted by the authors in November 2013. This Web questionnaire was administered to 448 general consumers (aged from their 20s to 60s) throughout Japan. Those 448 persons comprised 224 men and 224 women, constituting an extracted sample (stratified sampling method) that was proportional to the age and sex population ratios of the Ministry of Internal Affairs and Communications' "Population Census."

The questionnaire survey contained two questions: question 1, "What kind of images do you hold of the word 'Kuruma Prawn'? Please write freely as many images as come to mind." and question 2, "What kind of images do you hold of the word 'Black Tiger Prawn'? Please write freely as many images as come to mind." Respondents were asked to respond in a free-description format. The number of people who provided valid responses in the free-description field was 401 (200 men, 201 women) for question 1 and 386 (188 men, 198 women) for question 2. No large bias was observed between the entire sample population and the valid respondents in terms of composition ratios based on sex or age group.

Analytical Method

Text mining is "a method and system of sampling valid information by using a technique of natural language analysis on groups of non-formalized sentences to separate sentences into words and phrases, and then analyze these to ascertain frequency of occurrence and interrelationships. The word 'mining' means digging out information of value from mountains of text. Text mining is one kind of data mining."[1] The technique of natural language analysis here refers to computer analysis of morpheme and sentence structure. In other words, it is text-mining analysis conducted by a computer to quantify and statistically analyze sentence (text) data and extract valid information objectively. In the analysis of conventional questionnaire surveys, quantitative analysis has generally used the response results of multiple-choice format question items. Response statements in the free-description field (text data) have not been used for quantitative analysis. To date, as the free-description field text data contains important information that researchers are not expecting, the general method has been to read the text data, one by one, and manually organize and classify that data. However, the manual processing of large volumes of text data acquired from a questionnaire survey is extremely time-consuming, and issues have been raised about retaining objectivity and removing arbitrariness in analysis.

However, through the popularization of PCs and the development and practical application of text mining, it has become possible to mechanize the processing of large volumes of data and perform quantitative analysis to generally resolve the problem of labor reduction, as well as the problem of retaining objectivity and removing arbitrariness. Recently, it has become possible to use software that supports Japanese language such as KH Coder, quantitative text analysis software by Higuchi (2004), TinyTextMiner by Matsumura and Miura (2009), and MLTP by Jin (2009). Applied research using text mining is increasing rapidly in various research fields such as psychology, medicine, marketing, and quality control.

In this research, KH Coder was used as the analysis software, and text mining was performed according to the following procedure. First, using ChaSen[2], a morphological analysis tool provided as standard for KH Coder, text data was put through Cleaning[3] to divide it into morphemes[4], and word-category information, etc., was appended to the words extracted as morphemes. Specifically, the following word categories extracted by morphological analysis were made the target of analysis: nouns, Sahen nouns, adjective verb, tags, verbs, adjectives, B-type nouns, and C-type nouns[5]

Next, consumers' unconscious images of Kuruma and BT Prawns were revealed by (1) extracting frequently occurring words using simple tabulation analysis, and (2) gaining understanding of the relationships between words using co-occurrence network analysis.

RESULTS

(1) Simple Tabulation Analysis

Table I shows the frequently occurring words for Kuruma and BT Prawns (extraction of the top-20 words). As Table I shows, in responses relating to consumer images of Kuruma Prawns, the appearance frequency of *koukyuu* 'high class' (155 times) and *ebi* 'prawn' (131 times) were extremely high, followed by the words *oishii* 'taste delicious' (89 times), *gisou* 'camouflage' (47 times), and *takai* 'high' (47 times). In the responses relating to consumer images of BT Prawns, *ebi* 'prawn' (128 times) was extremely frequent, followed by *yasui* 'inexpensive' (69 times), *oishii* 'taste delicious' (46 times), *furai* 'fry' (38 times), and *yunyu* 'import' (32 times).

As shown here, because the word *gisou* 'camouflage' ranks high for the Kuruma Prawns, it is clear that consumers have a bad impression of Kuruma Prawns related to the mislabeling incident. On the other hand, the appearance frequency of the word *gisou* 'camouflage' in relation to BT Prawns is 14 times, which is lower compared with Kuruma Prawns; therefore, the mislabeling does not seem to have had as bad an influence on the image of BT as it has had on Kuruma Prawns.[6]

Another point that we should pay attention to is that consumers evaluate taste highly, with *oishii* 'taste delicious' used for not only Kuruma but also BT Prawns, revealing that consumers did not have a bad image of BT Prawn quality.

Table I: Frequently occurring words (extraction of top 20 words)

rank order	Kuruma Prawns			Black Tiger Prawns		
	extraction word (*Japanese*)	meaning in English	appearance frequency	extraction word (*Japanese*)	meaning in English	appearance frequency
1	高級(*koukyuu*)	high class	155	エビ(*ebi*)	prawn	128
2	エビ(*ebi*)	prawn	131	安い(*yasui*)	inexpensive	69
3	美味しい(*oishii*)	taste delicious	89	美味しい(*oishii*)	taste delicious	46
4	偽装(*gisou*)	camouflage	47	フライ(*furai*)	fry	38
5	高い(*takai*)	high	47	輸入(*yunyu*)	import	32
6	食べる(*taberu*)	eat	36	養殖(*yousyoku*)	culture fishery	31
7	大きい(*ookii*)	big	27	大きい(*ookii*)	big	26
8	食材(*shyokuzai*)	cooking ingredient	24	外国産(*gaikokusan*)	foreign-made	25
9	料理(*ryouri*)	cooking	20	安価(*annka*)	affordable price	23
10	イメージ(*imeiji*)	image	19	クルマエビ(*kuruma-ebi*)	kuruma prawn	22
11	食品(*syokuhin*)	food product	18	スーパー(*suupar*)	supermarket	22
12	フライ(*furai*)	fry	14	庶民(*syomin*)	common people	22
13	天ぷら(*tenpura*)	tempura	14	冷凍(*reitou*)	freezing	22
14	クルマエビ(*kuruma-ebi*)	kuruma prawn	13	黒い(*kuroi*)	black	20
15	養殖(*yousyoku*)	culture fishery	13	一般(*ippan*)	general	16
16	高価(*kouka*)	expensive price	12	食べる(*taberu*)	eat	16
17	多い(*ooi*)	frequent	11	使う(*tsukau*)	use	15
18	値段(*nedan*)	price	11	イメージ(*imeiji*)	image	14
19	入る(*hairu*)	put in	11	偽装(*gisou*)	camouflage	14
20	表示(*hyouji*)	labeling	11	高級(*koukyuu*)	high class	13

(2) Co-Occurrence Network Analysis

The next analysis that was conducted focused on the interrelationships between words. In text-mining analysis, one of the analysis techniques commonly used to visualize the interrelationships between words is co-occurrence network analysis. A co-occurrence network is a network that visualizes the relationships between words used in a text using lines (edges) to bind words with words that have appearance patterns in common (a strong degree of co-occurrence).[7]

The index adopted to express the degree of co-occurrence was the Jaccard coefficient. The Jaccard coefficient is widely used in text mining as the index that expresses the co-occurrence relations between words. When the Jaccard coefficient value is large, it indicates that the degree of co-occurrence is strong and the interrelationship between word and word is strong. In the co-occurrence networks shown in Fig. 1 below, the stronger the level of co-occurrence between two words, the thicker the line joining them will be. To prevent this analysis becoming complicated due to too many words, the minimum occurrence frequency for words was set to at least five for Kuruma Prawns and at least six for BT Prawns, and the condition for the Jaccard coefficient was that both had at least 0.09.[8]

Fig. 1 shows a co-occurrence network diagram created using the Girvan-Newman method. The Girvan-Newman method (Newman and Girvan, 2004) is a method for grouping networks using the betweenness centrality of lines (edges)[9], which enables extraction of groups of word pairs that are comparatively strongly bound to each other. The words shown in Fig. 1 in the same color belong to the same group. The diagram for Kuruma Prawns has been divided into seven groups (A1 to A7), and the diagram for BT Prawns has been divided into six groups (B1 to B6). For each group classified in Fig. 1, the features of the images held by consumers collating with the response details of the questionnaire have been organized to produce Table II.

From Fig. 1 and Table II, with regard to the question "On what kind of occasion are they eaten?," it can be reasoned that Kuruma Prawns are regarded as a seasonal food eaten at New Year, while BT Prawns are regarded as ordinary everyday food. Moreover, with regard to the question "Where are they eaten?," it can be inferred from the words *hoteru* 'hotel' and *sushi* 'sushi' that Kuruma Prawns are eaten when dining out and from the word *suupar* 'supermarket' that BT Prawns are eaten as home cooking. Furthermore, the cooking method for Kuruma Prawns appears to be labor-intensive, as suggested by the words *shioyaki* 'broil with salt,' *tempura*, and *sushi*, but simple for BT Prawns, as suggested by *reitou-syokuhin* 'frozen food' and *ebi-furai* 'fried prawns.'

In other words, it is suggested that consumers not only differentiate between high-class Kuruma Prawns and affordably priced BT Prawns, but also differentiate the two kinds of prawn in terms of intended use.

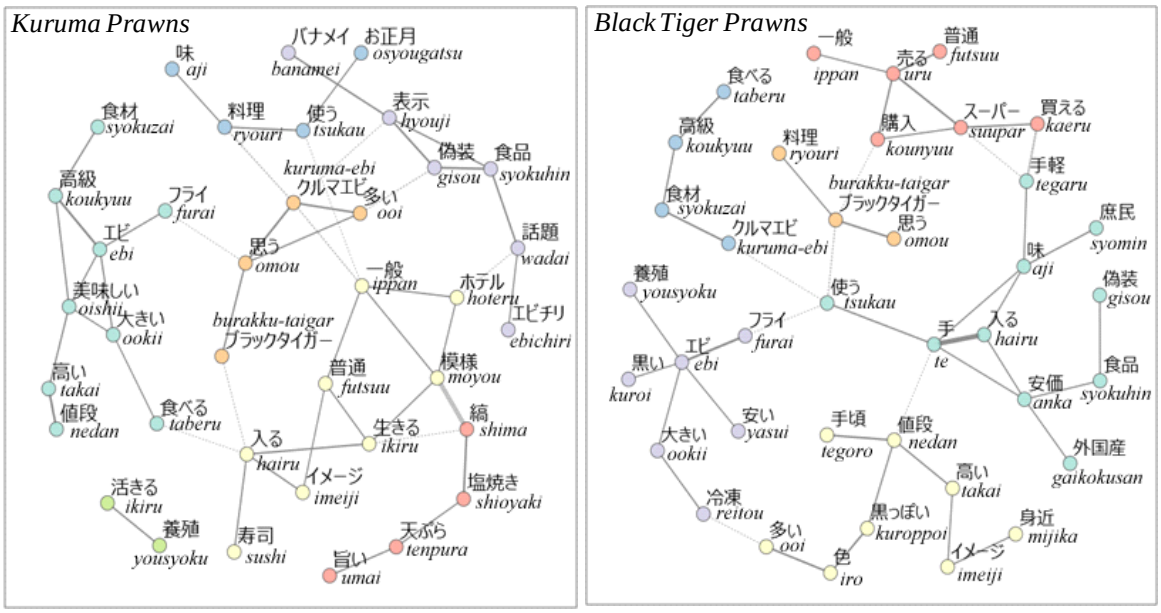

Fig. 1: Co-Occurrence Network Diagram of Kuruma and BT Prawns: Girvan-Newman Method

Table II: Features of Each Group

	Group	Word: *Japanese* 'phrase in English'	Feature
Kuruma Prawns	A1	*sushi* 'sushi,' *hairu* 'put in,' *imeiji* 'image,' *futsuu* 'ordinary,' *ippan* 'general,' *hoteru* 'hotel,' *moyou* 'pattern,' *ikiru* 'to live'	Image as a high-class prawn to eat in exclusive restaurant
	A2	*umai* 'tasty,' *tenpura* 'tempura,' *shioyaki* 'broil with salt,' *shima* 'stripes'	Image related to cooking technique
	A3	*nedan* 'price,' *takai* 'high,' *ebi* 'prawn,' *taberu* 'eat,' *ookii* 'big,' *oishii* 'taste delicious,' *koukyuu* 'high class,' *furai* 'fry,' *syokuzai* 'cooking ingredient'	Image as a high-priced, delicious-tasting cooking ingredient
	A4	*osyougatsu* 'New Year,' *tsukau* 'use,' *ryouri* 'cooking,' *aji* 'taste'	Image of prawn as traditionally used in New Year cooking
	A5	*ebichiri* 'chili prawns,' *wadai* 'topic of conversation,' *gisou* 'camouflage,' *syokuhin* 'food,' *hyouji* 'labeling,' *banamei* 'whiteleg shrimp'	Image related to mislabeling
	A6	*burakku-taigar* 'Black Tiger,' *omou* 'think,' *kuruma-ebi* 'Kuruma prawns,' *ooi* 'frequent,'	Image related to comparison with Black Tiger
	A7	*ikiru* 'to subsist,' *yousyoku* 'culture fishery'	Image related to production and sales format
Black Tiger Prawns (BT)	B1	*kounyuu* 'purchase,' *ippan* 'general,' *uru* 'sell,' *futsuu* 'ordinary,' *suupar* 'supermarket,' *kaeru* 'affordable'	Image of ordinary prawns purchased in a supermarket
	B2	*gaikokusan* 'foreign-made,' *anka* 'low price,' *te* 'hand,' *hairu* 'put in,' *tsukau* 'use,' *aji* 'taste,' *syomin* 'common people,' *tegaru* 'easy,' *syokuhin* 'food,' *gisou* 'camouflage'	Image of foreign-made prawn that can be put in the hand of the common people at a low price
	B3	*kuruma-ebi* 'Kuruma Prawns,' *syokuzai* 'cooking ingredient,' *koukyuu* 'high class,' *taberu* 'eat'	Image as a high-class cooking ingredient but not as high-class as Kuruma Prawns
	B4	*yousyoku* 'culture fishery,' *yasui* 'inexpensive,' *kuroi* 'black,' *ebi* 'prawn,' *furai* 'fry,' *ookii* 'big,' *reitou* 'frozen'	Image as a culture-fishery prawn that is a frozen food and an ingredient for fried prawn
	B5	*ooi* 'frequent,' *iro* 'color,' *kuroppoi* 'blackish,' *tegoro* 'reasonable,' *nedan* 'price,' *takai* 'high,' *imeiji* 'image,' *mijika* 'familiar'	Image as a reasonably priced prawn
	B6	*ryouri* 'cooking,' *burakku-taigar* 'Black Tiger,' *omou* 'think'	Image as a familiar cooking ingredient

CONCLUSION

This research has illuminated the following matters concerning BT prawns, mislabeled as Kuruma Prawns, by extracting frequently occurring words by simple tabulation using text mining and gaining an understanding of the relationships between words using co-occurrence network analysis.

Firstly, with respect to the extraction of frequently occurring words, the words *koukyuu* 'high-class' for Kuruma Prawns (ranked first place) and *yasui* 'inexpensive' for BT (ranked second place) were high-ranking frequently occurring words. These words were significantly more common than other words related to consumer concerns about food security and safety, such as *yunyu* 'import' and *yousyoku* 'culture fishery.' The results of analysis therefore suggest that the awareness that a product with an inexpensive image was sold at a high price worked more strongly on the consumer than concerns about health in this recent case of mislabeling.

Secondly, whereas the word *gisou* 'camouflage' (mislabeling) appears as a high-ranking frequently occurring word for Kuruma Prawns, it is not included amongst the high-ranked words for BT Prawns. Therefore, it seems that consumers have a bad image of Kuruma Prawns, but few people associate mislabeling with BT Prawns, on which the mislabeling has had little impact. From this, it would seem that it was mainly Kuruma Prawns that were damaged by the revelation of mislabeling.

Thirdly, as many consumers praised the quality of BT Prawns highly, using words such as *oishii* 'taste delicious,' the case is not a simple matter of an inexpensive, crude product being sold at a high price; rather, since BT Prawns were a product with two characteristics - affordable and high quality (a delicious tasting cooking ingredient) – then despite being the product that deceived the consumer, they identified it as a suitable cooking ingredient. In other words, this research shows that the mislabeling problem was a scandal that used a product's character and consumer awareness in an underhanded way.

Fourthly, the results suggest that Kuruma and BT Prawns, which are similar in taste and shape, are recognized by consumers as products with different intended purposes. This raises the question of what kind of conclusion would be brought about if these products, which are recognized as different in this way, were to be continuously traded within one market as the same product. In economic theory to date, a certain level of knowledge on this question has been accumulated. According to George Akerlof's analysis of the used car market, which he called a "Market for Lemons," where normally working cars and faulty cars are mixed together, a buyer who is unable to tell the faulty car apart attempts to purchase a normally working car cheaply on the consideration that it may possibly be faulty. Meanwhile, the seller of the normally working car does not put the car on the market as it only sells more cheaply than its intrinsic value. As a result, normally working cars stop appearing in the used car market, which causes a situation of adverse selection where only faulty cars appear in the market. In recent events, through the revelation of the mislabeling of BT Prawns as Kuruma Prawns, it became clear that once it has been prepared in cooking, it is difficult for consumers to tell Kuruma Prawns apart from BT, and that there is an asymmetry of information between producers and consumers, just as in the aforementioned Market of Lemons. For this reason, similar to the Market of Lemons, if mislabeling occurs frequently, and these incidents continue to be announced on the news, etc., the consumers who are the purchasers will give a great deal of consideration to the possibility of mislabeling, and the amount that they are willing to pay for Kuruma Prawns at restaurants will probably lower. If this occurs, then sales of Kuruma Prawn cuisine at restaurants and first-class hotels may slump, and the trade volume of Kuruma Prawns at restaurants may decline. A consequence of this is the possibility of a lowering of the trade price of high-priced Kuruma Prawns in the markets that the fishermen supply. In this way, the recent problem of food mislabeling is not simply a problem for the restaurant business alone; caution should be given to the fact that this is also a problem that could possibly impact the producers as well.

ACKNOWLEDGMENT

This paper is contribution No.173 from Fisheries Technology Research Department; Kyoto Prefectural Agriculture, Forestry and Fisheries Technology Center. This work was in part supported by JSPS Grants-in-Aid for Scientific Research (Grant No. 25870194).

REFERENCES

Akerlof, G. (1970) , "The market for lemons: quality uncertainty and the market mechanism," *Quarterly Journal of Economics* 84 (3), 488-500.

Higuchi, K. (2004) ”Quantitative Analysis of Textual Data : Differentiation and Coordination of Two Approaches,” *Sociological Theory and Methods*, 19(1), pp.101-115.

Jin, M. (2009) *Tekisuto-deeta no toukei-kagaku nyuumon [Introductory Statistical Science of Text Data]*. Tokyo: Iwanami shoten.(Japanese)

Matsumura, N. and A. Miura (2009) *Jinbun & shakai kagaku no tame no tekisuto mainingu [Text Mining for The Researchers of Humanistic and Social Sciences]*. Tokyo: Seishinshoboo. (Japanese)

Newman, M. E. J. and M. Girvan, (2004), "Finding and Evaluating Community Structure in Networks," *Physical Review*, E 69: 026113.

Ogawa, S., K. Usui, T. Ishii, S. Yamamoto, H. Ishii, K. Katou, T. Yamamoto and K. Egawa (2003) ”The marketing research about the fish image of the consumer in Kanagawa Prefecture,” *Bulletin of The Kanagawa Prefectural Fishery Experimental Station*, 8, pp.25-32.

“Sapporo no Chuukaten mo Ebi de Gisouhyouji [Chinese shop in Sapporo mislabeled shrimp],” Web-ban Nihon Keizai Shinbun [Online: Nikkei Newspaper in Japanese], 2013/10/30. URL：http://www.nikkei.com/article/DGXNASDG3001M_Q3A031C1CC0000/ (Accessed：2014/6/5)

Yokota, S., N. Ikeda, Y. Suzuki, S. Furusawa, S. Ogura and S. Sato (2010）”Consumer's trend of beef, pork and chicken in Japan,” *Bulletin of Integrated Field Science Center Tohoku University*, 26, pp.1-8.

NOTE

1. The section in brackets is a translation of Japanese quoted from IT Yougo Jiten e-Words. URL: http//sp.e-words.jp/ (original Japanese quoted from Web on December 7, 2013)
2. ChaSen was developed at the Computational Linguistics Laboratory, Graduate School of Information Science, Nara Institute of Science and Technology (at Yuji Matsumoto’s Laboratory) and it was publicly released on the same laboratory’s website (http://cl.naist.jp/).
3. Cleaning is the name of the operation of removing unnecessary symbols and strings from the text and amending mistaken strings.
4. A morpheme is the minimum unit of a string that has meaning.
5. These names of word categories are in accordance with the word category system of KH Coder. B-type nouns are nouns (words written entirely in hiragana) and C-type nouns are nouns (words with one kanji character). For details on the KH Coder word category system, see the “KH Coder 2.x reference manual” (in Japanese) (October 5, 2013). A tag is a KH Coder unique word category. For example, “Black Tiger” will be split by ChaSen into two morphemes of “black” (noun) and

"tiger" (proper noun). This situation is undesirable as the analysis will be problematic because "Black Tiger" is not handled as one word. Here, KH Coder is used to make a setting that forces the word "Black Tiger" to be extracted as one morpheme so that it can be analyzed. The morpheme word that is forcefully extracted is categorized as a word category called a tag. The words that contained tags in this research included *burakku-taigar* 'black tiger,' *banamei* 'whiteleg shrimp,' *gaikokusan* 'foreign made,' and *ebi-chiri* 'chili prawns.'

6. The Japanese words associated with food mislabeling included *gisou* 'camouflage,' *gisou syokuhin* 'mislabeled food,' *gisou hyouji* 'mislabeling,' and *syokuhin gisou* 'food mislabeling.' Moreover, even when the number of people who responded were counted instead of the appearance frequency of the words, 45 or approximately 11.2% of the number of valid respondents for Kuruma Prawns used the word *gisou* 'camouflage,' which was more than the 13 or approximately 3.4% of the number of valid respondents for BT Prawns.
7. In the co-occurrence network, it was important whether a line joined a word pair, and significance was not given to the relationship if the line did not join, even if the distance between words was close.
8. By doing this, the number of words that were included in the drawing was about the same: 37 for Kuruma Prawns and 38 for BT.
9. In the field of graph theory, "comparatively strong mutually bound parts" are called a "community." However, as the word "community" has a special meaning in the field of sociology, the extraction of these communities is called grouping so as not to be misleading. Newman and Girvan (2004) also use the expression "extract community structure."