Combining Behaviors and Demographics to Segment Online Audiences: Experiments with a YouTube Channel

Bernard J. Jansen[1], Soon-gyo Jung[1], Joni Salminen[1,2], Jisun An[1], Haewoon Kwak[1]

[1] Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
[2] University of Turku, Turku, Finland
bjansen@hbku.edu.qa, sjung@hbku.edu.qa, jsalminen@hbku.edu.qa, jan@hbku.edu.qa, hkwak@hbku.edu.qa

Abstract. Social media channels with audiences in the millions are increasingly common. Efforts at segmenting audiences for populations of these sizes can result in hundreds of audience segments, as the compositions of the overall audiences tend to be complex. Although understanding audience segments is important for strategic planning, tactical decision making, and content creation, it is unrealistic for human decision makers to effectively utilize hundreds of audience segments in these tasks. In this research, we present efforts at simplifying the segmentation of audience populations to increase their practical utility. Using millions of interactions with hundreds of thousands of viewers with an organization's online content collection, we first isolate the maximum number of audience segments, based on behavioral profiling, and then demonstrate a computational approach of using non-negative matrix factorization to reduce this number to 42 segments that are both impactful and representative segments of the overall population. Initial results are promising, and we present avenues for future research leveraging our approach.

Keywords: Audience segmentation; audience analytics; user profiling.

1 Introduction

Understanding the audience is a critical task in many domains, including marketing, advertising, system development, online content creation, and website design. In the wake of social media proliferation, companies and other organizations now have access to much more user data than ever before [1] However, dealing with "big data" has been found to be cumbersome [2], so that the cognitive limitations assign much greater constraints for utilizing user data than the availability of the data itself [3, 4]. One of the means to counter the overload of data is through efficient audience segmentation.

Although there has been considerable work in many domains focused on segmenting audiences by information consumption patterns [5], as this is often critical to successful online content creation, identifying audience segments in many situations is difficult due to lack of data, too much data, or privacy concerns [6]. It is also problematic to determine what a meaningful audience segment is, especially for large, diverse, and complex audiences, such as international social media followership. This difficulty

motivates our work, as the question of determining the right number of audience segments for efficient decision making [7] is constrained by finding meaningful segments. By meaningful, we mean an audience segment that is *behaviorally and demographically different* than one or more other segments from the same overall audience.

Given the increased interest in micro-targeting [8], i.e., focusing on extremely small audience segments, we do not aim at maximizing the size of the segment in our approach of combining behavioral and demographic information. In contrast, we utilize online videos from the content collection of a major news organization, along with associated behavioral and demographic attributes to identify *the maximum possible number* of distinct audience segments within the overall total audience. We then develop an approach for reducing this set of audience segments to the smallest number of meaningful audience segments. This reduction permits a narrower focus by organizations on the most impactful (i.e., distinct and actionable) audience segments for decision making in content creation, marketing, or other uses.

In this manuscript, we lead off with a short background section, introduce our data, and present our methods and results. We conclude with discussion, implications, and directions for the next stages of the on-going research effort.

2 Related Literature

Audience segmentation [9] is the process of separating a group of people into homogenous sub-segments, typically based on behaviors, demographics, or both, with the grouping most commonly grounded around some goal, product, system, or content [10]. Each audience segment can be defined as a group of people from the overall audience who are similar in specific ways but different from the other segments of the population. The identification of audience segments has long been central to marketing [11], and it is increasingly important in online content publishing. The purpose of segmentation is typically to increase understanding of users or customers, often using some key performance indicators capturing a strategic goal, such as increasing content views [12].

However, prior research has identified several challenges related to segmentation. Firstly, deriving actionable results from segmentation is not easy. For example, Tkaczynski, Rundle-Thiele, and Prebensen [13] analyze more than two thousand customers using a clustering, arriving at two segments, neither of which were actionable according to the researchers. In examine online news consumption, An and Kwak [14] attempt to segment audiences across 129 countries and by topic, and find it challenging to summarize the complex audience. Secondly, demographic differences are typically used for segmentation, but there is a need for more meaningful segmentation criteria, such as behavioral differences [15]. Prior work has shown that social media analytics can inform aspect of information behavior, such as information sharing [16]. Social media data can thus offer insights into understanding audience information preferences driving the online behavior [17]. This is extremely important, as social media actors, such as YouTube channels, often have complex audiences [18].

As an example of behavioral segmentation, online data has been used to segment audiences into various purchasing groups [12]. Behavioral segmentation using Hidden

Markov Models shows that there are consistent behaviors within searching stages [19]. Garcia et al. [20] report on a variety of behavioral attributes of a YouTube audience. Zhou and Zhang [21] use data from Weibo, one of the most popular social media platforms in China, to detect patterns of dietary preferences. In the broadcast domain, with the increased use of social media sites for news distribution, there is a concern of audience fragmentation. Fletcher and Nielsen [22] use audience segmentation to show that online audiences are no more fragmented than off-line audiences, although they found that cross-platform audiences vary by country. Lo, Chiong, and Cornforth [23] leverage large-scale social media data to identify the top-k audience members. Araújo et al. [24] explore YouTube audiences by country and age in online adolescence channels, finding large and diverse segments of audiences.

Overall, segmentation has been researched in a variety of contexts, including system design [25], health care [26], crisis response [27], journalism [28], and marketing [29]. However, prior work typically focuses on either demographic or behavioral segmentation, but they are rarely applied simultaneously. In particular, there is a lack of research into complex populations that may contain multiple behavioral audience segments *within a certain demographic group*, such as with popular social media channels that provide content ranging over a variety of topics. In these cases, the content preferences of individuals belonging to the same demographic group might drastically differ, requiring behavioral segmentation. However, even in these cases decision makers are interested in retaining some demographic information in the segments, as having this information facilitates audience insights and immersion [30]. Therefore, dealing with large and demographically and behaviorally diverse audience populations can be quite challenging. One approach is to simplify these complex populations by identifying the most impactful audience segments, which is the approach we investigate here.

3 Research Objectives

Our goal is to develop a methodology for reducing audience segments from large and diverse audience populations to the most meaningful segments that retain both demographic and behavioral information. In other words, we trim the audience segments to achieve a practical number of segments by focusing on reducing the audience segments to a minimal but still informative number. In practice, this number is somewhat arbitrary and requires further empirical work with decision makers of audience segments. Here, we simply measure how much we can *reduce* the number of segments in order to simplify the audience without losing its essential characteristics.

With this goal in mind, we define B as a behavioral audience segment, which is a segment based on interactions (behavior) with a set of content (C). D is a demographic audience segment, which is segment based on demographic characteristics. U is an integrated audience segment composed of both behavioral and demographic attributes.

In this research, we will:

1. Identify the number of behavioral segments (B) within a total audience.

2. Associate each of these behavioral segments with a set ($D_{1 \text{ to } X}$) of weighted demographic segments (D_i), resulting in a set of integrated segments U, defined as (B x $D_{1 \text{ to } X}$), where X is the maximum number of segments.
3. Develop an approach for reducing the set ($D_{1 \text{ to } X}$) while retaining the most descriptive demographic segments ($D_{1 \text{ to } i}$) resulting in a reduced set of U.

We present our data and methodology in the following sections.

4 Methodology

4.1 Data Collection

For our research, we collect actual audience data from a major online media and mobile channel based with millions of audience members with varying interests in the online content and geographically distributed worldwide. Our data source is the *online news channel AJ+*[1], which was designed from its founding to serve news in the medium of the viewer, with no redirect to a website or other platform. AJ+ is based on social platforms, meaning the digital content developed is specifically designed to be viewed by audiences on the Facebook, YouTube, Twitter, or Instagram, depending on the audience members who are most active on each platform.

For the data collection platform for the research reported in this manuscript, we use the AJ+ YouTube Channel, reserving analysis of the Twitter, Facebook, and other platforms for future work. However, the technique presented here is generalizable to any social media channel providing aggregated audience statistics [31]. As with many other social media platforms, the YouTube channel's analytics platform provides detailed statistics for every video. As an example of an AJ+ YouTube video, see Figure 1.

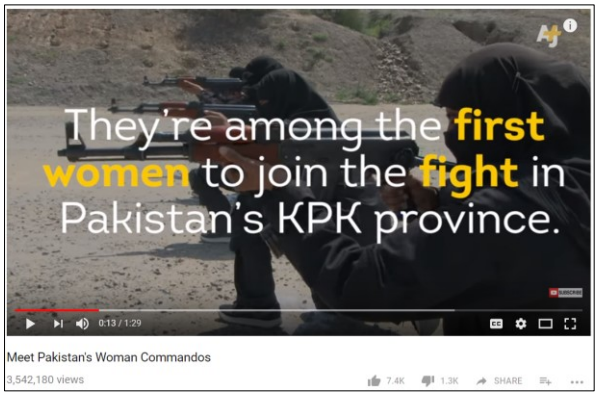

Fig. 1. Example of YouTube video from the AJ+ YouTube channel, with number of views.

[1] Part of the Al Jazeera Media Network.

For this research, we collect data on 4,320 videos produced from June 13, 2014 to July 27, 2016. Collectively, these videos have had more than 30 million views from people in 190 countries at the time of the data collection. The YouTube analytics platform provides, for each piece of the content collection, user attributes (e.g., gender, age, country location) at an aggregate level. One can access the data via the YouTube application programming interface (API), through which data can be collected automatically if permission is granted by the channel owner. We obtained the permission and collected the data. The attributes used for this research are listed in Table 1.

Table 1. Demographic and behavioral variables in the dataset.

Type	Description
Demographic attributes	**ageGroup:** YouTube viewers are classified into multiple age categories (13-17, 18-24, 25-34, 35-44, 45-54, 55-64, and 65 years and older); 7 possible age categories for a customer. **gender:** YouTube viewers are classified as either male or female, so there are 2 possible categories. **country:** YouTube uses the two-letter ISO-3166-1 country code index to classify where viewers are from, with 249 current officially assigned country codes at the time of this study.
Behavioral attribute	**viewCount:** YouTube provides the number of views per video for a given [country] by [gender, ageGroup].

We operationalize a demographic segment as a unique combination of (country, gender, age group). With 2 genders, 7 age groups, and 249 counties, this yields an upper limit of 3,486 audience segments. However, our data has 2,214 demographic segments, as not all countries and not all age groups for each country are represented. Even so, the dataset can be said to describe a large and diverse social media audience.

4.2 Segmentation Procedure

To isolate audience behavior patterns, the embedded structures in the aggregated data should be discovered. We employ a matrix decomposition technique, specifically non-negative matrix factorization (NMF), for this purpose. As this process is outlined in [32], we only conceptually present it here (see Table 2).

Table 2. Matrix decomposition process using Non-negative matrix factorization (NMF). The original matrix V is decomposed into two matrices, H and W.

Step	Description
Step 1	We first develop a matrix representing users' interaction with online content.
Step 2	The columns of the matrix are the online content pieces, e.g. videos (e.g., c contents (C_1, C_2, ..., C_c)).
Step 3	The rows of the matrix are the user groups or customer demographic segments (e.g., g user groups (G_1, G_2, ..., G_g)).
Step 4	Therefore, the matrix describing the association between user groups and contents is denoted by V the g×c matrix of g user groups or customer demographic segments and c contents.
Step 5	The element of the matrix V, V_{ij}, is any statistic that represents the one interaction or set of interactions of the user group G_i for content C_j.

Using this matrix approach as the basis, we can decompose (i.e., separate into simpler components) the overall matrix **V** into two matrices: **W** and **H**. In other words, one complex matrix **V** can be approximated as the product of smaller matrices **W** and **H**. The matrix **H** contains the audience behaviors and matrix **W** contains the user demographics. The matrix **H** encodes an association between the behavioral segments and the individual pieces of content. The resolution in finding user behavioral segments can be adjusted by the number of rows in **H**. The outcome of this step is the determination of *B*. Although one can present as many behavioral segments as the data contains, cognitive limits of the end users of the system pose a practical upper bound. It is not purposeful to show end users hundreds of customers segments.

Once we have the matrix **H**, we discover, via NMF, the underlying latent patterns, which describe the user segment interactions with the sets of individual content. These latent patterns become the basis of the user demographic segments, $D_{1 \text{ to } X}$, in the next step. The matrix **W** encodes an association between user demographic groups and behavioral user segments (i.e., latent content interaction patterns) with the set of corresponding user demographic segments each with an associated weight indicating how strongly each demographic segment is associated with the given behavioral segment. Each row in **W** represents how each demographic segment can be characterized by different behavioral patterns. The columns in **W** show how each latent behavioral pattern is associated with different demographic segments.

The resolution in finding demographic segments can be adjusted by the number of columns in **W**. We identify the most impactful demographic segments associated with the previously defined behavioral segments. A single behavioral segment can have multiple associated demographic segments, as shown in Table 3. This reflects the fact that users within the same demographic attributes can interact with different content.

Table 3. Demographic segments (D_1 to D_{10}) associated with behavioral Segment 1 (B_1). Shaded areas are below the elbow and with negative z-scores.

User Behavioral Segment #1 (B1)					
Seg.	Country	Age	Gender	Weight	z-score
D_1	US	25	male	415.73	2.330935
D_2	US	35	male	313.00	1.215947
D_3	US	45	male	221.56	0.223497
D_4	CA	25	male	216.42	0.167709
D_5	CA	35	male	176.54	-0.26513
D_6	US	55	male	165.83	-0.38137
D_7	GB	25	male	147.47	-0.58064
D_8	CA	45	male	124.21	-0.8331
D_9	US	65	male	114.93	-0.93382
D_{10}	GB	35	male	113.99	-0.94402

Given our dataset, we define a behavioral audience segment as a pattern of video viewing. With 4,320 videos, we have an extremely high number of combinations of videos viewed by different segments. However, prior research has shown that individuals struggle to work with large volumes of information [34]. For this research, we use 15 behavioral segments, each associated with 10 demographic segments. Therefore, we have a total of 150 potential segments, which is the number we aim to reduce.

5 Results of Simplification

We have four variables to utilize for simplifying our segmentation, which are: (a) weight, (b) country, (c) gender, and (d) age. Leveraging first weight, we want to determine a cut-off beyond which the demographic segments are less meaningful.

For this, we employ the *elbow approach*, in which one chooses the number of segments so that adding another segment does not add much information. For many phenomena, the marginal gain will drop, giving an angle, or elbow, in a graph. The number of segments is chosen at this point. The elbow can be verified using the z-score, which is a measure of how many standard deviations above or below a population's mean a raw score is. The z-score is calculated with $z = (X - \mu) / \sigma$ where z is the z-score, X is the value of the element, μ is the population mean, and σ is the standard deviation.

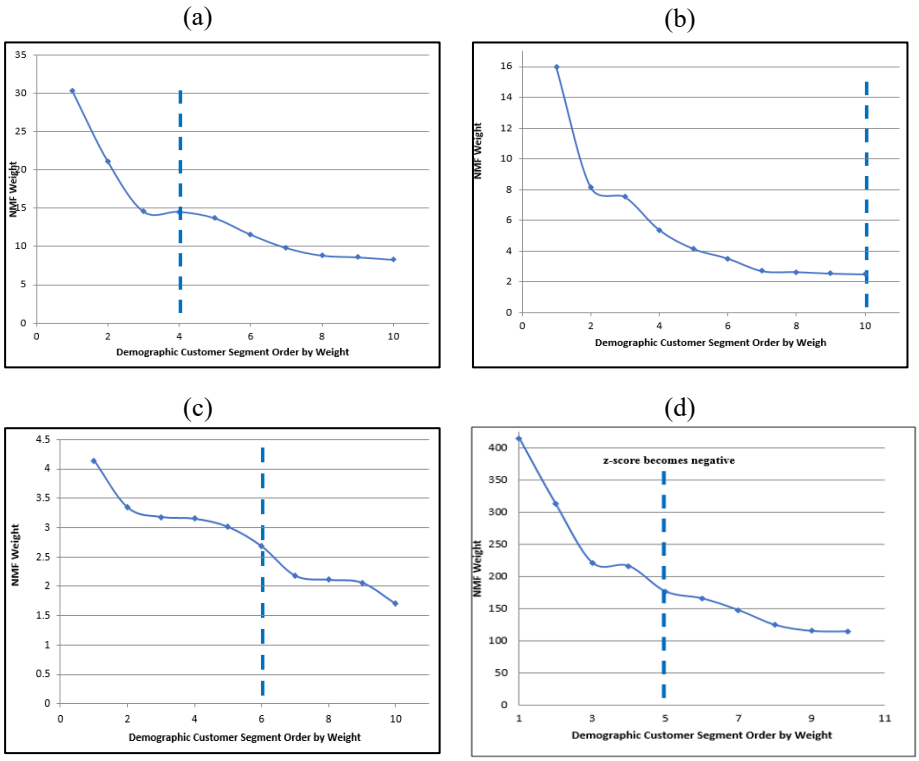

Fig. 2. Segment 2 with elbow at demographic segment 4 (a); segment 10 with no elbow in top ten demographic groups (note: elbow at segment 10) (b); segment 15 with elbow at demographic segment 6 (c); segment 1 with elbow at demographic segment 5.

We utilize both the graphical elbow graph and the z-score for all 15 behavioral segments (with three sample graphs shown in Figure 2, eliminating the demographic segments below the elbow (i.e., where the z-score becomes negative). Figure 2 displays the graphical display of the demographic weights of behavioral Segment 1 with the critical z-score shown as a vertical line. Using elbow and z-score approaches, we reduce our entire data set of 150 audience segments to 63 segments (a 58% reduction).

Table 3 shows the z-score calculations for the set of demographic segments (D_1 to D_{10}) for behavioral Segment 1 (B_1). For each behavioral segment, we group demographic segments based on country, specifically using the ISO 3166-1 country codes utilized by most of the major online platforms. Keeping the demographic segments distinct by country makes sense, as privacy, marketing, and other legal restrictions often vary by country. If two or more demographic segments are from the same country, then those audience segments are candidates for consolidation (see Table 4).

Table 4. Segment 10 grouped into 6 possible demographic segments from the original 10 based on country variable.

Country	Age	Gender	Weight
GB	25	male	15.97
GB	35	male	8.16
GB	18	male	7.53
DE	25	male	5.38
CA	25	male	4.15
GB	45	male	3.52
IT	25	male	2.7
AE	25	male	2.64
CA	18	male	2.54
CA	35	male	2.49

We group demographic segments based on gender, specifically in this research using a binary of male and female. If demographic segments are of the same gender from the same country, those segments are candidates for consolidation (see Table 5).

Table 5. Segment 4 grouped into two possible demographic segments of four based on gender variable.

Country	Age	Gender	Weight
PH	25	male	3.4
PH	18	male	1.59
PH	35	male	1.21
PH	25	female	1.18

We group demographic segments based on age category. As this is initial research, we take a heuristic approach: (a) if the age bracket is bounded by corresponding age brackets from the same country and gender, then the age brackets are aggregated, or (b) if the age bracket is adjacent to a corresponding age bracket from the same country and gender, then the two age brackets are aggregated (see Table 6).

Table 6. Segment 11 grouped into one demographic segment based on age variable.

Country	Age	Gender	Weight
MY	55	male	9.26
MY	25	male	7.13
MY	65	male	6.71
MY	45	male	6.42
MY	35	male	5.76

We then apply the country, then gender, and then age consolidations to each of the 15 behavioral segments. Table 7 presents an example of the overall simplification approach for behavioral segment 1, with the original 10 demographic segments reduced to 2. The *Weight* attribute reduced the set to 6. The *Country* and *Gender* attributes can reduce the segments to 2, which was the result when the *Age* heuristic was applied.

Table 7. Segmentation simplification process, reducing 10 demographic segments (D_1 to D_{10}) into 2 (D_{US} and D_{CA}) for behavioral segment 1.

Country	Age	Gender	Weight	z-score
US	25	male	415.73	2.330935
US	35	male	313.00	1.215947
US	45	male	221.56	0.223497
CA	25	male	216.42	0.167709
CA	35	male	176.54	-0.26513
US	55	male	165.83	-0.38137
GB	25	male	147.47	-0.58064
CA	45	male	124.21	-0.8331
US	65	male	114.93	-0.93382
GB	35	male	113.99	-0.94402

Table 8 shows the results for all 15 behavioral segments after applying the reductions discussed above. Our original 150 integrated audience segments were reduced to 42 segments, a 72% simplification, with each segment distinct in terms of behaviors. The average simplification was 2.8 segments, with a maximum of 9 and a minimum of 1.

Table 8. Set of 150 segments simplified to 42 by applying first weight and then country-gender-age reduction.

Behavioral Segment (*B*)	No. of Demographic Segments (*D*)	Applying Weight (*D*)	Applying Country-Gender-Age (*D*)
1	10	4	2
2	10	4	4
3	10	3	3
4	10	4	2
5	10	3	1
6	10	3	2
7	10	5	5
8	10	3	2
9	10	3	3
10	10	10	9
11	10	5	1
12	10	3	1
13	10	4	2
14	10	4	1
15	10	5	4
	150	63	42

6 Discussion and Conclusion

The abundance of social media data has been transformative for online content creation. However, at the same time it has resulted in an overwhelming number of potential audience segments, especially for channels with large international audience.

In this research, we show that meaningful audience segmentation that retains both demographic and behavioral information is achievable. Our research results show that audience segmentation can be accomplished rapidly and dynamically using a large-scale user data from major online social media platform, reflecting the content consumption behavior of real people forming the channel's online audience.

Furthermore, we show that one can employ consistent techniques to reduce the number of audience segments in complex user populations. With our approach on the example dataset, we achieved a 72% simplification of the audience segments, while considering both demographic and behavioral variation in the output personas.

Concerning limitations, a known shortcoming of the applied elbow method is that the elbow cannot always be unambiguously identified, even when using the z-score. In future research, we are investigating ways to solve this issue.

Moreover, while we limited our focus here to understanding the meaningful audience segments with social media data, it would be interesting to apply the approach to the long tail of audience segments to investigate possible micro-audience segments, cultural differences [35], and content topics [36].

From a practical point of view, although this manuscript is specifically focusing on digital content on YouTube, our approach can be applied in a wide range of contexts, given that the data structure remains similar. Increasingly, audience segmentation processes are leveraging aggregated audience data [37], having the advantage of retaining the privacy of individual users. Our approach, utilizing aggregated data from the YouTube API, has this benefit as the privacy of users is kept safe.

Overall, the strength of the research is that we use real content and user data from a major social media platform to investigate novel ways for audience segmentation, with promising areas for future research. We are actively investigating the use of other social media and data sources to provide richer audience segments for online content creators.

References

1. Nguyen, T., Zhou, L., Spiegler, V., Ieromonachou, P., Lin, Y.: Big data analytics in supply chain management: A state-of-the-art literature review. Computers & Operations Research. (2017).
2. Agarwal, R., Dhar, V.: Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. Information Systems Research. 25, 443–448 (2014).
3. Gandomi, A., Haider, M.: Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management. 35, 137–144 (2015).
4. Edwards, J.S., Taborda, E.R.: Using Knowledge Management to Give Context to Analytics and Big Data and Reduce Strategic Risk. Procedia Computer Science. 99, 36–49 (2016).
5. Hendahewa, C., Shah, C.: Evaluating user search trails in exploratory search tasks. Information Processing & Management. 53, 905–922 (2017).
6. Salminen, J., Şengün, S., Kwak, H., Jansen, B.J., An, J., Jung, S., Vieweg, S., Harrell, F.: From 2,772 segments to five personas: Summarizing a diverse online audience by generating culturally adapted personas. First Monday. 23, (2018).
7. Sweller, J.: Cognitive load during problem solving: Effects on learning. Cognitive science. 12, 257–285 (1988).
8. Cho, M., Auger, G.A.: Extrovert and engaged? Exploring the connection between personality and involvement of stakeholders and the perceived relationship investment of nonprofit organizations. Public Relations Review. 43, 729–737 (2017).
9. Shafto, A.: Mastering Audience Segmentation: How to Apply Segmentation Techniques to Improve Internal Communication. Melcrum (2006).
10. Stern, B.B.: A Revised Communication Model for Advertising: Multiple Dimensions of the Source, the Message, and the Recipient. Journal of Advertising. 23, 5–15 (1994).
11. Smith, W.R.: Product Differentiation and Market Segmentation as Alternative Marketing Strategies. Journal of Marketing. 21, 3–8 (1956).
12. Ortiz-Cordova, A., Jansen, B.J.: Classifying Web Search Queries to Identify High Revenue Generating Customers. J. Am. Soc. Inf. Sci. Technol. 63, 1426–1441 (2012).
13. Tkaczynski, A., Rundle-Thiele, S.R., Prebensen, N.K.: To segment or not? That is the question. Journal of Vacation Marketing. 24, 16–28 (2018).

14. An, J., Kwak, H.: Multidimensional Analysis of the News Consumption of Different Demographic Groups on a Nationwide Scale. In: Social Informatics. pp. 124–142. Springer, Cham (2017).
15. Jansen, B.J., Booth, D.: Classifying Web Queries by Topic and User Intent. In: CHI '10 Extended Abstracts on Human Factors in Computing Systems. pp. 4285–4290. ACM, New York, NY, USA (2010).
16. Liu, Z., Jansen, B.J.: Questioner or question: Predicting the response rate in social question and answering on Sina Weibo. Information Processing & Management. 54, 159–174 (2018).
17. Gonzalez Camacho, L.A., Alves-Souza, S.N.: Social network data to alleviate cold-start in recommender system: A systematic review. Information Processing & Management. 54, 529–544 (2018).
18. Nguyen, H.T., Le Nguyen, M.: Multilingual opinion mining on YouTube – A convolutional N-gram BiLSTM word embedding. Information Processing & Management. 54, 451–462 (2018).
19. Han, S., He, D., Chi, Y.: Understanding and modeling behavior patterns in cross-device web search. Proceedings of the Association for Information Science and Technology. 54, 150–158.
20. Garcia, D., Abisheva, A., Schweitzer, F.: Evaluative Patterns and Incentives in YouTube. In: Social Informatics. pp. 301–315. Springer, Cham (2017).
21. Zhou, Q., Zhang, C.: Detecting dietary preference of social media users in China via sentiment analysis. Proceedings of the Association for Information Science and Technology. 54, 523–527.
22. Fletcher, R., Nielsen, R.K.: Are News Audiences Increasingly Fragmented? A Cross-National Comparative Analysis of Cross-Platform News Audience Fragmentation and Duplication. Journal of Communication. 67, 476–498 (2017).
23. Lo, S.L., Chiong, R., Cornforth, D.: Ranking of high-value social audiences on Twitter. Decision Support Systems. 85, 34–48 (2016).
24. Araujo, C.S., Magno, G., Meira Jr, W., Almeida, V., Hartung, P., Doneda, D.: Characterizing videos, audience and advertising in Youtube channels for kids. arXiv:1707.00971 [cs]. (2017).
25. Salminen, J., Jung, S.-G., An, J., Kwak, H., Jansen, B.J.: Findings of a User Study of Automatically Generated Personas. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. p. LBW097:1–LBW097:6. ACM, New York, NY, USA (2018).
26. Burkell, J., Fortier, A.: Could we do better? Behavioural tracking on recommended consumer health websites. Health Info Libr J. 32, 182–194 (2015).
27. Kim, Y., Miller, A., Chon, M.-G.: Communicating with Key Publics in Crisis Communication: The Synthetic Approach to the Public Segmentation in CAPS (Communicative Action in Problem Solving). Journal of Contingencies and Crisis Management. 24, 82–94 (2016).
28. Nelson, J.L.: And Deliver Us to Segmentation. Journalism Practice. 12, 204–219 (2018).
29. Ashley, C., Tuten, T.: Creative Strategies in Social Media Marketing: An Exploratory Study of Branded Social Content and Consumer Engagement. Psychology & Marketing. 32, 15–27 (2015).

30. Nielsen, L., Storgaard Hansen, K.: Personas is applicable: a study on the use of personas in Denmark. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1665–1674. ACM (2014).
31. An, J., Kwak, H., Jansen, B.J.: Personas for Content Creators via Decomposed Aggregate Audience Statistics. In: Proceedings of Advances in Social Network Analysis and Mining (ASONAM 2017), Sydney, Australia (2017).
32. Jung, S.-G., An, J., Kwak, H., Ahmad, M., Nielsen, L., Jansen, B.J.: Persona Generation from Aggregated Social Media Data. In: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. pp. 1748–1755. ACM, New York, NY, USA (2017).
33. Jansen, B.J., An, J., Kwak, H., Salminen, J., Jung, S.-G.: Viewed by Too Many or Viewed Too Little: Using Information Dissemination for Audience Segmentation. Presented at the Association for Information Science and Technology Annual Meeting 2017 (ASIST2017) , Washington DC, USA November 27 (2017).
34. Miller, G.A.: The magical number seven plus or minus two: some limits on our capacity for processing information. Psychol Rev. 63, 81–97 (1956).
35. Salminen, J., Şengün, S., Kwak, H., Jansen, B.J., An, J., Jung, S., Vieweg, S., Harrell, F.: Generating Cultural Personas from Social Data: A Perspective of Middle Eastern Users. In: Proceedings of The Fourth International Symposium on Social Networks Analysis, Management and Security (SNAMS-2017). , Prague, Czech Republic (2017).
36. AL-Smadi, M., Jaradat, Z., AL-Ayyoub, M., Jararweh, Y.: Paraphrase identification and semantic text similarity analysis in Arabic news tweets using lexical, syntactic, and semantic features. Information Processing & Management. 53, 640–652 (2017).
37. Jansen, B.J., Sobel, K., Cook, G.: Classifying ecommerce information sharing behaviour by youths on social networking sites. Journal of Information Science. 37, 120–136 (2011).