

The Categorising Characteristics of Facebook Pages: Using the K-Means Grouping Method

Pin Luarn¹, Hong-Wen Lin¹, Yu-Ping Chiu¹, Yu-Liang Shyu² & Pei-Ching Lee²

¹ Dept. of Business Administration, National Taiwan University of Science and Technology, Taipei City, Taiwan

² Innovative DigiTech-Enabled Applications & Services Institute, Institute for Information Industry, Taipei City, Taiwan

Correspondence: Hong-Wen Lin, Ph.D. Candidate, Dept. of Business Administration, National Taiwan University of Science and Technology, No.43, Sec. 4, Keelung Rd., Da'an Dist., Taipei City, Taiwan. Tel: 886-2-2733-3141 ext. 7227. E-mail: woodylin34@hotmail.com

Received: December 10, 2015

Accepted: December 28, 2015

Online Published: January 23, 2016

doi:10.5539/ijbm.v11n2p60

URL: <http://dx.doi.org/10.5539/ijbm.v11n2p60>

Abstract

This study conducts the K-means grouping analysis on 1,373 Facebook pages in order to find the difference and characteristics between groups, and furthermore attempt to understand the behavioural characteristics of Facebook page users. The study produces four clusters with different characteristics, all of which are named and defined according to their qualities. The four types of pages are the “*functional video and audio informational pages*,” “*audio and video entertainment with low discussion pages*,” “*high-identifying celebrity pages*”, and “*food and travel with active discussion pages*.” The result of this study not only provides a clear understanding of the grouping structure of Facebook pages, but also provides reference for fan page managers when constructing management strategies.

Keywords: Facebook page, K-means clustering analysis, key characteristics

1. Introduction

The widespread use of social networking sites (SNSs) (i.e., Facebook, Twitter, Google+, LinkedIn, Flickr) have increased internet usage of social media users, with Facebook being the most popular (Liu, Hu, Mian, Tian, & Zhu, 2014). As of September 2015, the total number of monthly active Facebook users have reached 1.44 billion (a 12% increase compared to 2014), with individuals spending an average of 18 minutes on Facebook, and total of 74.2 million Facebook pages worldwide (Statistic Brain Research Institute, 2015).

More and more cooperation with social marketing companies has lead to the discovery that not only can SNSs connect corporations with new clients, but also expand the scale of the corporation, with Facebook being the highest received internet platform (Kang, Tang, & Fiore, 2014; Sarwar, Haque, & Yasmin, 2013). Facebook pages not only have business potential, but are also able to promote brand image and attract the attention of users (Lin & Lu, 2011). Strand (2011) also stated that Facebook pages benefit corporations or brands by connecting present and potential clients, which not only build good customer relations, but also brings entertainment to fans. Therefore, more and more corporations are investing in the management of Facebook pages.

However, the phenomenon of information overload in SNSs have made it difficult for users to find useful information according to their interests (Zhang, Chen, & Pan, 2013). Even when the message itself is potentially useful, information overload has become an obstacle for users (Eppler & Mengis, 2004). If corporations wish to satisfy the many users of Facebook with pages and post information that fits their interests, understanding the key characteristics of Facebook pages is crucial. Past research has arranged data and characteristic categorisation through cluster analysis (Broderick, Jordan, & Pitman, 2013), which have been applied in hydrology and management (Romesburg, 2004). Therefore, this study will base cluster foundations on the attributes and characteristics of Facebook pages, while taking the advice of Ramaswamy and Rose (2011), and include related variables of Facebook pages, calculate its distance from the median, and conduct the following research:

- 1) Categorize the attribute information of Facebook pages using cluster analysis, and deduce the key characteristics in order to construct a complete grouping structure of Facebook pages.
- 2) Name each cluster by its grouping characteristic and clarify the characteristics of each cluster.

2. Theoretical Background

2.1 Cluster Analysis

Cluster analysis is a powerful way of finding precise themes within large amounts of data (Larsen & Aone, 1999). Cluster analysis is a multivariate method that uses the measure between distances to view the connection between multivariate variables (Skarbinski et al., 2009). Its aim is to calculate the adjacent values of variables within spaces, and further group them with variables with the same characteristics in order to form a hierarchical cluster (Larsen & Aone, 1999). According to the different methods, cluster analysis can be classified into hierarchical cluster analysis and nonhierarchical cluster analysis, as well as two step cluster analysis, which is a combination of the first two.

Hierarchical cluster analysis is a widely used method which groups similar objects according to the distance between observations and variables, and then groups the observations through agglomerative clustering (Yüksel, 2003). It is usually applied when there are fewer observations, and the number of clusters is chosen through dendrograms. It is divided into the coacervation method and the disintegrating method (Moore et al., 2010). Nonhierarchical cluster analysis selects various groups first and find the most adequate grouping through the iterative method, and is not only suitable when there are large numbers of observations, but also when analysed with the K-means analysis method (Wagstaff, Cardie, Rogers, & Schrödl, 2001). The unique convergence method of the nonhierarchical cluster analysis can provide few insights; therefore, it is often applied repeatedly with hopes of receiving a meaningful result (Davidson & Ravi, 2007). However, if there are too many observations and the judgement is based on dendrograms, the data will be overly dispersed and difficult to be read and explained (Singh & Dubey, 2010). The two step cluster analysis, however, first uses the hierarchical cluster analysis to decide the number of clusters, and then uses the nonhierarchical cluster analysis for grouping. Two step cluster analysis can conduct mixed analysis on continuous and related data, and present clusters with diverse quantity (Haase, 2014).

Hierarchical cluster analysis, nonhierarchical cluster analysis and two step cluster analysis are used in different groupings of data. If the amount of data is small and when conducting simple viewings of solutions of increasing clusters, the nonhierarchical cluster analysis can be applied. Those with larger amounts of data can use the K-means cluster analysis within nonhierarchical cluster analysis. Lastly, when confronting large amounts of data, or those mixed with continuous and categorical variables, the two step cluster analysis is the best choice (Haase, 2014). This study applies the K-means clustering for the grouping analysis of Facebook pages.

2.2 Facebook Page

Facebook pages are an important tool for constructing the relationship between corporations and fans (Maeda & Kinoshita, 2011). The reason for this is that Facebook pages moves the interaction between corporations and customers into the virtual social groups that are not restricted to the real world (Ellison, Vitak, Gray, & Lampe, 2014). Once a user joins a Facebook page, the user is able to receive posts, images, and videos from Facebook pages on their walls (Jahn & Kunz, 2012), and also to interact with other members of the same Facebook page through posts, comments, likes and sharing (Kim, 2013). Therefore, for the user, Facebook pages are virtual groups that exist with the premise of common interests. Users are connected to other members of the page through sharing interests, and to construct mutual values (Seraj, 2012).

Facebook pages also provide various benefits for corporations. Past research indicates that when fans join the Facebook page of a particular brand, they form a sense of loyalty towards the brand (Sahin, Zehir, & Kitapçı, 2011), and become more acceptable towards information that the brand provides (De Vries, Gensler, & Leeflang, 2012). In addition, fans that join Facebook pages are more likely to visit stores than those who do not (Dholakia & Durham, 2010), and are more willing to leave positive reviews on the internet (Gupta & Harris, 2010). Sysomos (2009) analysed up to 600 thousand Facebook pages and discovered that most Facebook pages have 10 to 1,000 fans, while 4% have more than 10 thousand fans, and 0.76% have over 100 thousand fans. The above shows the undeniable business value of Facebook pages.

However, with the increase in number of Facebook pages, the amount of information that users encounter has also increased (Lin & Lu, 2011). Excessive and overly diverse information creates difficulty for users when trying to find Facebook pages that interests them (Zhang, Medo, Ren, Zhou, Li, & Yang, 2007). Therefore, the managers of Facebook pages are contemplating methods to increase exposure and attract current and potential fans (De Vries et al., 2012). However, increasing exposure is not enough to enable present and potential fans to find Facebook pages that interests them, and is also more costly for Facebook page managers. Therefore, Zhou, Xu, Li, Josang and Cox (2012) has taken a step further and pointed out the more effective method of SNS recommendations.

3. Methodological issues

3.1 Data Collection

This study is performed through the Institute for Information Industry of Taiwan with the database method. There were 1,373 pieces of data regarding of 50 types of Facebook pages between February 15th and March 15th in 2014 are accumulated, which are furthermore analysed using the Waikato Environment for Knowledge Analysis (WEKA) 3.6 version. This tool for exploring information is only easy for use (Sharma, Bajpai, & Litoriya, 2012), but also excels at categorising and clustering, also at processing statistics on association rules.

This study uses categorical and continuous variables as evaluation items for the grouping of Facebook pages. The age of Facebook page fans are applied as grouping categorical variables. As for continuous variables, it includes the percentage of male fans, percentage of discussion, the growth of fan numbers, percentage of types of posts (Photo, Video, Status, Link), and the average monthly amounts of comments, shares, and likes of each post (variables definition as shown in Table 1).

Table 1. Element definition

Variable	Item	Content
Categorized variables	Age of Facebook page fans	The percentage of fans aged 18-24, 25-34, 35-44, 45-54, 55 and above
Continuous variables	Percentage of male fans	The percentage of men in the total number of fans
	Percentage of discussion	The percentage of fans that are talking about the Facebook page among the total number of fans
	Growth of fan numbers	The percentage of growth among total Facebook page fans
	Average monthly amounts of comments of each post	The percentage of comments among the total amount of posts on Facebook page
	Average monthly amounts of shares of each post	The percentage of shares among the total amount of posts within Facebook page
	Average monthly amounts of likes of each post	The percentage of likes among the total amount of posts in Facebook page
Percentage of types of posts		The individual percentage of 4 types of posts among total amount of posts
		*Posts with photos
		*Posts with videos
		*Post with status only
		*Posts with links

3.2 K-Means Clustering

K-means clustering is widely applied among fields such as business, biology, categorising internet documents, and images processing, all of which have received effective results (Kanungo et al., 2002). Compared with hierarchical cluster analysis, K-means clustering is less effected by outliers, similarity measure, and improper grouping variables (Liu, 2007). It also has the characteristic of being faster and simpler (Xu & Wunsch, 2005).

This study applies the following steps of conducting the K-means clustering analysis. (1) Find the most suitable amount for clustering (Ghosh & Dubey, 2013). (2) Set a starting point for each cluster as the initial estimate for each cluster seed (Sharma et al., 2012). (3) Calculate the distance between each piece of information and cluster centre, and distribute data to the cluster with the shortest distance (Bhatia & Khurana, 2013). (4) Re-calculate the distance to the new centre (Ghosh & Dubey, 2013). (5) Repeat step 3 and 4 until no information can change the categorisation of clusters, or the numerical of the centroid ceases to change (Ghosh & Dubey, 2013).

4. Results

4.1 Clustering Results

This study uses the Weka 3.6 version and the simple K-means cluster algorithm to categories 1,373 pieces of Facebook page information. In order to obtain the best amount of clusters, this study applies the sum of squared errors as determining criterion. The smallest numerical indicates the closeness of the same cluster, and better clustering results. After numerous algorithms and adjustment of the seed, the results indicate 4 best clusters with the seed value of 100 and sum of squared errors as 4283.35.

As for age groups and gender percentages, the highest amount of Facebook pages are those in cluster 3, with

545 pages. Within these pages, female fans aged 18-24 are with highest percentage, with female fan percentage reaching 72.05%. Cluster 1 is with least amount of Facebook pages, with only 145 pages and consists mostly of males fans aged 25-34, reaching 69.74%. The gender percentage in cluster 2 is more even, with fans most aged between 18-24. Cluster 4 mostly consists of female fans aged 25-34. **As for discussion percentage and growth of fan number**, the discussion percentage of the 4 clusters are 8.84~12.71%, with fan growth between 0.94~8.55%. Both variable indexed point to cluster 4 as the highest and cluster 2 as the lowest. **Regarding average monthly number of posts and comments, shares and likes**, the average monthly amount of comments of each post is around 15.63~28.03, with cluster 4 distinctively higher, and cluster 2 with the least comments. In addition, the average monthly amount of shares are around 12.93~46.09 posts, with cluster 4 the highest and cluster 1 the lowest. The average monthly amount of likes are 846.73~1,773.57, distinctively higher than the amount of comments and shares, and among them cluster 3 is the highest, while cluster 1 is the lowest. **As for the types of posts**, apart from cluster 1, the posts of all other clusters are mainly videos, with cluster 4 exceptionally passionate about photo posts. Cluster 1, with 26.16%, has the highest percentage of video posts compared with other clusters. Video post types lie between 15.33~21.72%, with cluster 1 being the highest and cluster 4 the lowest. Link posts are with smallest numbers among the four clusters, with cluster 2 at 6.28% begins the highest and cluster 4 at 2.42% being the lowest (as shown in Table 2).

In addition, among the analysis of 50 Facebook pages with different management qualities (as shown in Appendix A), cluster 1 mostly consists of “Digital and electronics,” “Reporters and anchors,” and “Books and magazines.” Clusters 2 and 3 are mainly Facebook pages of “Singers and bands,” “Internet celebrities,” and “Actors.” Cluster 4 mostly consists of Facebook pages regarding “Food and cuisine,” “Travel,” and “Internet celebrities.”

Table 2. Cluster data analysis

Measure	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Number of Facebook page	145(10.56%)	176(12.82%)	545(39.69%)	507(36.93%)
Percentage of male fans	63.39%	49.56%	27.95%	35.45%
Percentage of discussion	10.33%	8.84%	9.05%	12.71%
Growth of fan numbers	1.00%	0.94%	3.26%	8.55%
Average monthly amounts of comments of each post	18.89	15.63	24.79	28.03
Average monthly amounts of shares of each post	12.93	16.75	25.00	46.09
Average monthly amounts of likes of each post	846.73	1,466.98	1,773.57	1,185.45
Percentage of types of posts				
*Posts with photos	25.43%	36.65%	40.22%	47.78%
*Posts with videos	26.16%	13.75%	10.65%	11.48%
*Post with status only	21.72%	19.46%	20.30%	15.33%
*Posts with links	4.57%	6.28%	3.71%	2.42%

4.2 Naming of Clusters

The result of this study induces the characteristic of each cluster, and defines the content as well as names each cluster according to the quality of the Facebook page (as shown in Table 3).

Cluster 1 is comprised mostly of male fans aged 25-34 with the percentage up to 63.39%, and is the cluster with most male percentage. The Facebook pages are mostly informational pages such as “Digital and electronics,” “Reporters and anchors,” and “Books and magazines.” The interaction between fans are lower key, and mostly regard functional information. The average posts specify statuses, photos, and videos. Therefore, this study names cluster 1 as “functional video and audio informational pages.”

Cluster 2 mostly includes Facebook pages such as “Singers and bands,” “Internet celebrities,” and “Actors.” Most members are fans aged 18-24, with male and female fans each occupying half the amount. Cluster 2 favours posts with links, and is the highest among all clusters. There is lower discussion between fans, also lower growth in the number of fans. Therefore, cluster 2 is named as the “audio and video entertainment with low discussion pages.”

Cluster 3 has more Facebook pages, and is also composed mostly of pages regarding “Singers and bands,” “Internet celebrities,” and “Actors.” The fans are mostly women aged 18-24, and usually express interests through liking posts. This study names cluster 3 as the “high-identifying celebrity pages.”

Cluster 4 mostly consists of women aged 25-34 who like to express opinions through commenting and sharing posts. Discussion rate and fan amount in these Facebook pages are with high percentage, and consists mostly of “Food and cuisine” and “Travel.” This study names cluster 4 as “food and travel with active discussion pages.”

Table 3. Induction of cluster characteristics

Cluster	Key characteristics	Named
1	Facebook page: mainly about “Digital and electronics,” “Reporters and anchors,” and “Books and magazines” Fans: Men aged 25-34 Behaviour: dislike likes, sharing, and commenting on posts Other: post types are mostly images, words and videos, more like platforms that provide information	Functional video and audio informational pages
2	Facebook page: mainly about “Singers and bands,” “Internet celebrities”, and “Actors” Fans: both men and women aged 18-24 Behaviour: then to like posts but rarely comments Other: low in discussion percentages and growth of fan amounts, mostly includes posts with links.	Audio and video entertainment with low discussion pages
3	Facebook page: mainly about “Singers and bands”, “Internet celebrities”, and “Actors” Fans: female aged 18-24 Behaviour: tends to like posts and have high identification Other: with the most amount of Facebook pages	High-identifying celebrity pages
4	Facebook page: mainly about “Food and cuisine”, “Travel” Fans: female aged 25-34 Behaviour: tends to comment and share posts, and has more interaction Other: high in both discussion rates and fan growth	Food and travel with active discussion pages

5. Conclusions and Implications

5.1 Conclusions

This study has acquired 1,373 pieces of Facebook page data, which is analysed through the K-means cluster analysis methods to inspect and induce the key characteristics and furthermore name the different clusters. This study uses the age of fans as categorical variables, and percentage of male fans, discussion percentages, fan growth, types of posts (photo, video, status, link), as well as average monthly comments, share and likes of posts as continuous variables, while also taking the management nature into account. The result show 4 clusters with different characteristics, and are named the “functional video and audio informational pages,” “audio and video entertainment with low discussion pages,” “high-identifying celebrity pages,” and “food and travel with active discussion pages.”

The results indicate that cluster 1, the “functional video and audio informational pages,” stresses the dissemination of knowledge and messages, and mostly consists of male fans aged 25-34 (with male fans up to 63.39%) who visit Facebook pages regarding “digital and electronics,” “reporters and anchors,” and “books and magazines.” The overall activity of the Facebook pages are lower key, with equal posts on statuses, photos and videos. This study infers the formation of the characteristic of this cluster as fans who use Facebook page as a platform for information, therefore they mostly leave after reading complete information, and rarely interact with others.

Cluster 2, “audio and video entertainment with low discussion pages,” mostly consist of Facebook pages regarding “singers and bands,” “internet celebrities,” and “actors.” The fans of these Facebook pages are mostly young fans aged 18-24, with male and female each taking up half of the population. The percentage of posts with links are the highest amongst all clusters, but are lower on discussions. This study infers that the reason for the formation of the characteristic of this cluster is mainly the distribution of idol-related news, therefore posts with links make it easier to guide users to related webpages. However, once these young users click the links and leave Facebook, they rarely return to Facebook for discussions. The management method of stressing external links result in the relatively low percentage in discussion and fan growth.

Cluster 3, the “high-identifying celebrity pages,” is the cluster with most Facebook pages, and mostly consists of female fans aged 18-24 (female percentage up to 72.05%). Similar to cluster 2 “audio and video entertainment

with low discussion pages,” cluster 3 is also composed of Facebook pages regarding “singers and bands,” “internet celebrities,” and “actors.” Apart from gender percentage and age interval, the fans of cluster 3 like to express concept through likes. Furthermore, the posts of these Facebook pages are mostly photos and statuses. This study infers that these pages gain identification from fans through their posts, which consists of more photos and statuses, and less links. These kinds of posts helps fans to concentrate on the content of posts.

Cluster 4 “food and travel with active discussion pages” are pages with contents regarding “food and cuisine” and “travel,” with fans mostly consisting of females aged 25-34 (up to 64.55%) who like to comment and share posts. The percentage of “travel” is distinctively higher than other pages within this cluster. Furthermore, the rate of discussion and fan growth in this cluster is the highest amongst others. This study infers the reason that comments and shares of this cluster to be higher than others is due to its quality, since food and travel interests most people.

5.2 Managerial Implications

This study is based on the grouping of Facebook pages and provide suggestions regarding actual management methods according to the characteristics of each cluster, with hopes of assisting Facebook page managers to understand the characteristics of each cluster, and develop adequate marketing strategies in order to gain profit.

After observing 30 pieces of data of “functional video and audio informational pages,” this study has found that its posts are mainly status updates plus photos/videos. This type of information display results in fans merely liking posts and treating Facebook pages as platforms for information, while rarely interacting with each other. However, some posts are popularly received, which attract fans to engage in discussions that involve gifts. The study has also found that if the gifts are not related to the page itself, then the result of the gift will also be limited. Therefore, this study suggests that the managers of “functional video and audio informational pages” apply gifting activities more frequently and strengthen the relation with Facebook pages, in order to increase the engagement of Facebook pages.

On the other hand, this study has observed that the posts of “food and travel with active discussion pages” mainly consists of statuses plus photos/video, but receive high amounts of comments and discussions. This study has found that apart from the differences between the qualities of different clusters, the managers of “food and travel with active discussion pages” mostly treat their pages as a brand. This study suggests that Facebook page managers make use of the high discussion and sharing rates of this cluster and organise limited offers and activities, such as free accommodation coupon for a groups of ten, or meal offers, in order to effectively attract potential customers. Not only does this promotes reputation, but also increases profit for the cooperation.

This study hopes to not only provide reference in research, but also methods that can be widely used in corporations through the extension of cluster analysis of Facebook pages, in order to promote the matching accuracy between business owners and users, and furthermore create a win-win situation for business owners and users.

References

- Bhatia, M., & Khurana, D. (2013). Experimental study of data clustering using k-means and modified algorithms. *International Journal of Data Mining and Knowledge Management Process*, 3(3), 17-30. <http://dx.doi.org/10.5121/ijdkp.2013.3302>
- Broderick, T., Jordan, M. I., & Pitman, J. (2013). Cluster and feature modeling from combinatorial stochastic processes. *Statistical Science*, 28(3), 289-312. <http://dx.doi.org/10.1214/13-STS434>
- Davidson, I., & Ravi, S. S. (2007). The complexity of non-hierarchical clustering with instance and cluster level constraints. *Data Mining and Knowledge Discovery*, 14(1), 25-61. <http://dx.doi.org/10.1007/s10618-006-0053-7>
- De Vries, L., Gensler, S., & Leeflang, P. S. (2012). Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. *Journal of Interactive Marketing*, 26(2), 83-91. <http://dx.doi.org/10.1016/j.intmar.2012.01.003>
- Dholakia, U. M., & Durham, E. (2010). One café chain’s Facebook experiment. *Harvard Business Review*, 88(3), 26. Retrieved from <https://hbr.org/2010/03/one-cafe-chains-facebook-experiment>
- Ellison, N. B., Vitak, J., Gray, R., & Lampe, C. (2014). Cultivating social resources on social network sites: Facebook relationship maintenance behaviors and their role in social capital processes. *Journal of Computer-Mediated Communication*, 19(4), 855-870. <http://dx.doi.org/10.1111/jcc4.12078>
- Eppler, M. J., & Mengis, J. (2004). The concept of information overload: A review of literature from

- organization science, accounting, marketing, MIS, and related disciplines. *The Information Society: An International Journal*, 20(5), 325-344. http://dx.doi.org/10.1007/978-3-8349-9772-2_15
- Ghosh, S., & Dubey, S. K. (2013). Comparative analysis of k-means and fuzzy c-means algorithms. *International Journal of Advanced Computer Science and Applications*, 4(4), 35-39. <http://dx.doi.org/10.1371/journal.pone.0085884>
- Gupta, P., & Harris, J. (2010). How e-WOM recommendations influence product consideration and quality of choice: A motivation to process information perspective. *Journal of Business Research*, 63(9-10), 1041-1049. <http://dx.doi.org/10.1016/j.jbusres.2009.01.015>
- Haase, S. (2014). Engineering students' sustainability approaches. *European Journal of Engineering Education*, 39(3), 247-271. <http://dx.doi.org/10.1080/03043797.2013.858103>
- Jahn, B., & Kunz, W. (2012). How to transform consumers into fans of your brand. *Journal of Service Management*, 23(3), 344-361. <http://dx.doi.org/10.1108/09564231211248444>
- Kang, J., Tang, L., & Fiore, A. M. (2014). Enhancing consumer-brand relationships on restaurant Facebook fan pages: Maximizing consumer benefits and increasing active participation. *International Journal of Hospitality Management*, 36, 145-155. <http://dx.doi.org/10.1016/j.ijhm.2013.08.015>
- Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 881-892. <http://dx.doi.org/10.1109/TPAMI.2002.1017616>
- Kim, J. (2013). Relationship between Facebook usage and self-efficacy among collegiate athletes. *Media Watch*, 4(3), 364-374.
- Larsen, B., & Aone, C. (1999). Fast and effective text mining using linear-time document clustering. *Paper presented at the Proceedings of the 5th ACM SIGKDD international conference on knowledge discovery and data mining*. <http://dx.doi.org/10.1145/312129.312186>
- Lin, K. Y., & Lu, H. P. (2011). Intention to continue using Facebook fan pages from the perspective of social capital theory. *Cyber Psychology, Behavior, and Social Networking*, 14(10), 565-570. <http://dx.doi.org/10.1089/cyber.2010.0472>
- Liu, B. (2007). *Web data mining: Exploring hyperlinks, contents, and usage data*. Heidelberg: Springer.
- Liu, H., Hu, Z., Mian, A., Tian, H., & Zhu, X. (2014). A new user similarity model to improve the accuracy of collaborative filtering. *Knowledge-Based Systems*, 56, 156-166. <http://dx.doi.org/10.1016/j.knosys.2013.11.006>
- Maeda, Y., & Kinoshita, T. (2011). Structural remodeling, trafficking and functions of glycosylphosphatidylinositol-anchored proteins. *Progress in Lipid Research*, 50(4), 411-424. <http://dx.doi.org/10.1016/j.plipres.2011.05.002>
- Moore, W. C., Meyers, D. A., Wenzel, S. E., Teague, W. G., Li, H., Li, X., ...Bleecker, E. R. (2010). Identification of asthma phenotypes using cluster analysis in the severe asthma research program. *American Journal of Respiratory and Critical Care Medicine*, 181(4), 315-323. <http://dx.doi.org/10.1164/rccm.200906-0896OC>
- Ramaswamy, S., & Rose, K. (2011). Adaptive cluster distance bounding for high-dimensional indexing. *IEEE Transactions on Knowledge and Data Engineering*, 23(6), 815-830. <http://dx.doi.org/10.1109/TKDE.2010.59>
- Romesburg, H. C. (2004). *Cluster analysis for researchers*. Morrisville, NC: Lulu. com.
- Sahin, A., Zehir, C., & Kitapçı, H. (2011). The effects of brand experiences, trust and satisfaction on building brand loyalty; An empirical research on global brands. *Procedia-Social and Behavioral Sciences*, 24, 1288-1301. <http://dx.doi.org/10.1016/j.sbspro.2011.09.143>
- Sarwar, A., Haque, A., & Yasmin, F. (2013). The usage of social network as a marketing tool: Malaysian Muslim consumers' perspective. *International Journal of Academic Research in Economics and Management Sciences*, 2(1), 93-102. Retrieved from <http://www.hrmar.com/admin/pics/1491.pdf>
- Seraj, M. (2012). We create, we connect, we respect, therefore we are: Intellectual, social, and cultural value in online communities. *Journal of Interactive Marketing*, 26(4), 209-222. <http://dx.doi.org/10.1016/j.intmar.2012.03.002>

- Sharma, N., Bajpai, A., & Litoriya, R. M. (2012). Comparison the various clustering algorithms of WEKA tools. *International Journal of Emerging Technology and Advanced Engineering*, 2(5), 73-80. Retrieved from http://www.ijetae.com/files/Volume2Issue5/IJETAE_0512_13.pdf
- Singh, L., Singh, S., & Dubey, P. K. (2010). Applications of clustering algorithms and self organizing maps as data mining and business intelligence tools on real world data sets. *Methods and Models in Computer Science (ICM2CS)*, 2010 International Conference on, IEEE. <http://dx.doi.org/10.1109/ICM2CS.2010.5706714>
- Skarbinski, J., Ouma, P. O., Causer, L. M., Kariuki, S. K., Barnwell, J. W., Alaii, J. A., ...Hamel, M. J. (2009). Effect of malaria rapid diagnostic tests on the management of uncomplicated malaria with artemether-lumefantrine in Kenya: A cluster randomized trial. *The American Journal of Tropical Medicine and Hygiene*, 80(6), 919-926.
- Statistic Brain Research Institute. (2015). *Facebook statistics*. Retrieved from <http://www.statisticbrain.com/facebook-statistics/>
- Strand, J. L. (2011). Facebook: Trademarks, fan pages, and community pages. *Intellectual Property & Technology Law Journal*, 23(1), 10-13.
- Sysomos (2009). *Inside Facebook pages*. Retrieved from <http://www.sysomos.com/insidefacebook/>
- Wagstaff, K., Cardie, C., Rogers, S., & Schrödl, S. (2001). Constrained k-means clustering with background knowledge. Paper presented at the ICML.
- Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3), 645-678. <http://dx.doi.org/10.1109/TNN.2005.845141>
- Yüksel, A., & Yüksel, F. (2003). Measurement of tourist satisfaction with restaurant services: A segment-based approach. *Journal of Vacation Marketing*, 9(1), 52-68. <http://dx.doi.org/10.1177/135676670200900104>
- Zhang, W. Y., Zhang, S., Chen, Y. G., & Pan, X. W. (2013). Combining social network and collaborative filtering for personalised manufacturing service recommendation. *International Journal of Production Research*, 51(22), 6702-6719. <http://dx.doi.org/10.1080/00207543.2013.832839>
- Zhang, Y. C., Medo, M., Ren, J., Zhou, T., Li, T., & Yang, F. (2007). Recommendation model based on opinion diffusion. *Europhysics Letters*, 80(6), 68003. <http://dx.doi.org/10.1209/0295-5075/80/68003>
- Zhou, X., Xu, Y., Li, Y., Josang, A., & Cox, C. (2012). The state-of-the-art in personalized recommender systems for social networking. *Artificial Intelligence Review*, 37(2), 119-132. <http://dx.doi.org/10.1007/s10462-011-9222-1>

Appendix A.

The percentage of Facebook pages within each clusters

Type of Facebook page	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Life and living	1.38%	0.00%	0.92%	4.54%
Transportation	6.90%	0.57%	0.18%	1.38%
Merchandise and wholesale	0.00%	0.00%	1.47%	3.35%
Composite clothing	0.00%	1.14%	2.20%	1.78%
Finance and insurance	4.14%	0.00%	0.00%	2.56%
Beauty and fashion	1.38%	0.00%	4.95%	4.93%
Food and cuisine	1.38%	1.70%	4.40%	9.07%
Women's fashion	0.00%	0.00%	4.40%	1.38%
Men's fashion	0.00%	1.14%	1.10%	0.59%
Education and learning	0.00%	0.57%	0.37%	0.99%
Retails and chains	0.00%	0.00%	0.37%	1.18%
Transport services	0.00%	0.00%	0.00%	0.99%
Telecommunications	3.45%	0.00%	0.00%	0.20%
Health and medicine	0.69%	0.57%	0.55%	2.56%
Art and design	1.38%	1.70%	3.85%	2.17%
Professional sports	2.07%	2.27%	0.55%	0.20%

Shoe wear	0.00%	0.57%	1.65%	0.00%
Accessories	0.00%	0.00%	1.83%	0.39%
Travel	0.00%	1.14%	0.18%	6.71%
Sporting goods	2.07%	0.57%	0.92%	0.20%
Digital and electronics	11.03%	2.27%	0.00%	2.56%
Parenting/babies	0.00%	0.00%	0.18%	1.97%
Housing and real estate	1.38%	0.00%	0.00%	0.39%
Consumer/living information	0.00%	0.00%	0.18%	0.00%
Shopping/e-commerce	0.00%	0.57%	2.02%	4.93%
Other brands	3.45%	3.98%	1.83%	4.14%
Books and magazines	8.28%	5.68%	3.30%	5.72%
Internet society	2.07%	0.57%	0.55%	0.99%
Shows	2.07%	1.70%	2.39%	3.75%
TV channels	2.07%	2.27%	1.10%	1.38%
Movies	0.69%	1.14%	1.10%	0.20%
Motion picture companies	0.69%	1.14%	0.55%	0.79%
Animation and cartoons	0.69%	0.57%	0.92%	0.00%
TV series	0.00%	2.27%	1.65%	0.39%
News	5.52%	1.70%	0.37%	2.96%
Broadcast	0.69%	1.70%	0.37%	0.00%
Movie-related activities	0.00%	0.00%	0.37%	0.00%
Music	0.00%	0.00%	0.18%	0.00%
Other media	0.00%	1.14%	0.73%	0.99%
Singers and bands	4.14%	28.98%	22.94%	4.73%
Actors	3.45%	7.95%	8.26%	3.75%
Hosts	2.07%	3.98%	3.85%	1.78%
Reporters and anchors	9.66%	2.27%	0.00%	0.59%
Directors	0.00%	0.00%	0.00%	0.20%
Athletes	1.38%	1.70%	0.37%	0.59%
Internet celebrities	5.52%	8.52%	10.83%	6.71%
Writers	2.07%	2.27%	1.83%	1.38%
Political figures	4.83%	0.00%	0.18%	0.59%
Share mood	0.00%	0.00%	0.00%	0.20%
Other figures	3.45%	5.68%	4.04%	3.16%

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).