Future Generation Computer Systems 93 (2019) 121–129

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

# An approach to identify user preferences based on social network analysis

Stevan Milovanović, Zorica Bogdanović, Aleksandra Labus, Dušan Barać *, Marijana Despotović-Zrakić

*Faculty of Organizational Sciences, University of Belgrade, Jove Ilića 154, Serbia*

## HIGHLIGHTS

- An approach to identify user preferences in the context of cultural events based on the social network analysis (SNA).
- SNA was used for exploring data from two different sources: mobile application and questionnaire.
- The results revealed three clusters of cultural events, as well as the most influential users and their preferences towards certain types of cultural events.

## ARTICLE INFO

*Article history:*
Received 10 May 2018
Received in revised form 15 October 2018
Accepted 17 October 2018
Available online xxxx

*Keywords:*
Social computing
Social network analysis
User preferences
Mobile application event

## ABSTRACT

This paper introduces an approach for identifying user preferences using social network analysis (SNA). Main idea was to reduce complexity and enable effective and affordable social network analysis harnessing particular tools and techniques. As a proof of concept, we performed the research that included two sources: (1) the control data source — analytical data collected from mobile application FilterApp for cultural events and (2) the experimental data source — data based on survey for users of the mobile application. The results revealed three clusters of cultural events based on user preferences towards certain types of cultural events, the frequency of visits to cultural events and the size of groups when visiting these events. The obtained conclusions were used to develop system of recommendations and for customization of offer and marketing strategies according to the identified users' preferences. The main value of this paper is reflected in the clearly defined research process with all the phases from data collection to validation of results.



## 1. Introduction

Social network services include a plethora of personal, business and educational interactions among users. Data generated within social networks represents an input for social computing. The term social computing refers to an emerging concept that enables analysis of users' profiles, as well as their interactions and behaviors [1]. Discovered information could be used in various contexts. For example, results of social computing research can lead to new targets and facilitate the process of introducing new products or services [2]. By introducing social computing online services, sharing of knowledge, experience, equipment, and resources can be increased among individuals, companies and research groups [3].

Social computing facilitates collective actions and social interactions by using various applications and services for multimedia content exchange and knowledge aggregation [4–6]. Social networks are the most explored part of social computing. Social Network Analysis (hereinafter: SNA) is the result of harnessing social computing in the analysis of social networks. SNA represents an interdisciplinary methodology for mapping and measuring of relationships and flows between people, groups, organizations, computers, URLs, and other connected information entities [7]. SNA collects the data important for market research, making business decisions, analysis of marketing activities, identification of influential users of social networks, determining the interconnection of users of a particular social network, etc.

This paper shows an approach to identifying user preferences based on social network analysis. The main aim of the paper is to investigate possibilities of harnessing SNA for determining users' preferences in the context of cultural events. The paper should contribute to the literature by introducing an approach that enables SNA that is not dependent only on data from well known social networks, i.e could be performed on data from domain specific social network. As a proof of concept, SNA was used for exploring data from two different sources: (1) control data set represents analytics data from mobile application FilterApp that presents the offer of

* Corresponding author.
*E-mail addresses:* stevan@elab.rs (S. Milovanović), zorica@elab.rs (Z. Bogdanović), aleksandra@elab.rs (A. Labus), dusan@elab.rs (D. Barać), maja@elab.rs (M. Despotović-Zrakić).

https://doi.org/10.1016/j.future.2018.10.028
0167-739X/

**Table 1**
Social computing systems compared to traditional systems.
*Source:* Adapted from [9].

| Traditional systems | Social computing systems |
|---|---|
| Standardized procedure for quality assurance | Quality is based on customer feedback |
| Top-down structure | Bottom-up structure |
| Strictly defined | Flexible |
| Focused on a specific area | Focused on different interest areas |
| Point out the organizational boundaries | Expand borders and develop a community |
| Centralized | Decentralized |

cultural events in the city of Belgrade, Serbia and (2) experimental data set created with data collected using survey for users of this mobile application. The result of the research should describe different groups of cultural events separated based on their impact on users. These results should help us target marketing strategies to reach best results in terms of the number of application users and consequently the results of promotion of events through the application.

## 2. Theoretical background

### 2.1. Social computing

Social computing presents the use of computational devices to facilitate or augment the social interactions of their users, or to evaluate those interactions in an effort to obtain new information [6]. Main characteristics of social computing are [8]: the need for real-time communication, sharing experience, improving business and customer relationship management, online learning and statistical analysis of data collected from the Internet. A comparative overview of the characteristics of the traditional business system and social computing systems is presented in Table 1.

Social computing simply refers to the concept of the availability of social functions, such as communicating people over the Internet, creating communities on the web or information sharing. In a broader sense, social computing implies the realization of functions that involve the decision making, creation of a public opinion, knowledge aggregation, marketing forecast and reputation system. Numerous users' profiles are grouped according to different criteria and create a social network. In the scope of the social network, social computing can be used for economic purposes. The network is analyzed in such a way that conclusions about the characteristics of users or structures enable improving economic parameters [10].

Social networks are online environments where people can present themselves through their individual profiles, make links to other users and communicate with them [11]. User engagement within social networks can be described through following [12]: (1) downloading, viewing or listening to the content, (2) sorting, filtering, evaluating and commenting content created by other users, (3) the creation of content by other users and establishment of cooperation and communication, (4) communication frequency among users.

### 2.2. Social network analysis

The main elements of each social network are their users. User profiles have a threefold role: they contain information about the user (interests, activities, education, etc.), links to other members, and information about the users' affinities according to different stakeholders [10]. Users of social networks can be classified into several categories: [13,14]

- *Influencers*. Users or groups of users with a high level of knowledge, authority or position that can influence the decision-making of individuals on social networks.
- *Advocates*. Sub-groups of influential users who support or promote a brand or goal.
- *Socializers*. Active users on social networks, with a stable network of contacts, own their groups or pages, with great popularity and visitor numbers.
- *Observers*. They are quite inactive in generating and distributing any content, but they use information and services.

SNA aims to understand networks and their participants and has two main focuses: the actors and their relationships in a specific social context [12]. SNA portrays mapping and measurement of relationships and flows between people, groups, organizations or communities. In paper [15], social professional networks analysis was performed using SNA tools. Networks such as LinkedIn, GitHub and StackOverflow were investigated. The authors claim that different research topics address social professional networks and are divided into issues and tasks. The issues emerge from the need for crawling, storing, managing and treating the data from the networks. On the other hand, the tasks represent the ways that such networks can be analyzed, used, improved and applied in different contexts.

The methodological procedure of SNA includes determining cooperation among users, institutions and organizations, and analysis of hidden relationships, roles, and communities [10,16,17]. Furthermore, application of SNA methodology enables identification of direct correlation of network performance and cooperation in many different areas [17,18].

### 2.3. Social network analysis for detecting user preferences

With the merging of the cyber world and physical world, event-based social networks have been playing an important role in promoting the spread of offline social events through online channels [19]. The most common social networks used for this purpose are Facebook, Twitter, Instagram, and Meetup.

The process of collecting data about users is based on monitoring their geographic location, posts, likes, hashtags, favored events, frequency of activities on the social network and users' influence on their connections. Majority of papers who deal with the detection of user preferences, bases their research on various algorithms, such as content-based filtering, K-Nearest Neighbor, collaborative filtering, matrix factorization and many others to process data.

In [19] heterogeneous graph of interactions among different types of entities was constructed and event scoring algorithm called reverse random walk with restart was proposed in order to obtain the user-event recommendation matrix. The results of this paper are two heuristic algorithms for event participant scale control to balance the arrangements of users among upcoming events. Chen and Sun [20] propose a social event recommendation method that exploits a user's social interaction relations and collaborative friendships to recommend events of interest. This paper, through research carried out on data collected from social networks Facebook and Meetup, shows that proposed method is able to identify representative acquaintances of users to recommend events relevant to the preferences of the acquaintances. Chonghuan in [21] proposes a novel recommendation method based on the social network analysis using matrix factorization technique. The author clusters users and considers a variety of complex factors in order to increase accuracy and diversity of recommendation results and alleviate the sparsity and cold-start problems. Koren, Bell, and Volinsky in the paper [22] claim that matrix factorization models are superior to classic nearest neighbor techniques for producing product recommendations, allowing the incorporation of additional information such as implicit feedback, temporal effects, and confidence levels. In [23] grid-group cultural theory was incorporated in social network analysis to provide significant and alternative insights about the complexity of water pollution management. This paper states that the combination of qualitative and quantitative methods provides the breadth and depth of understanding and corroboration of complex problems. Jhamb and Fang propose a dual-perspective latent factor model for group-aware event recommendation by using two kinds of latent factors to model the dual effect of groups [24]: one from the user-oriented perspective (e.g. topics of interest) and another from the event-oriented perspective (e.g. event planning and organization). Using data from Meetup for events organized in four cities in the U.S.: New York, San Francisco, Washington DC and Chicago this article proves that latent factor model can be applied to other recommendation tasks where certain factors may have a dual view which is precisely the case in our research. Based on previous research on the subject of the recommendation of cultural events [25], this paper combines latent factor model, multi-relational matrix factorization, and probabilistic latent preference analysis in order to determine target groups of users for each event type.

Regarding data collection methods, in paper [26] a survey is recognized as one of the network data collection techniques along with sequenced and census. Furthermore, the authors of the paper [27] developed a close-ended survey to collect data for SNA as a process evaluation measurement instrument to gauge specific domains of potential partner interactions. The SNA survey measured material benefits, such as the number of referrals, collaboration, and sharing, and ideational benefits, represented by overall trust in the network. Following the success of these works, we have adopted the survey as a valid method of collecting data necessary for further research.

## 3. Identifying user preferences using SNA

The main idea of the research is to identify cultural events that are most visited as well as user preferences to different types of cultural events. Cultural events themselves were used as a suitable context where we could investigate if our approach was good enough or not.

The research should answer the following research questions:

1. Which types of cultural events are most preferred?
2. Which types of cultural events are most often visited?
3. Which types of cultural events are visited by the largest groups of visitors?

For the purposes of this research, mobile application FilterApp and online survey were used as data sources. Users' preferences determined according to information about them collected during application use should be compared with the expected users' preferences collected through the survey. Comparison and further analysis should be performed using SNA tools: Ucinet and Vosviewer. After data processing, a comparative analysis of the results of different clustering methods is required in order to compare identified and expected user preferences. The results should allow better market segmentation, planning social network marketing and personalized marketing campaigns through push notification in the FilterApp.

### 3.1. Research procedure

Research procedure for identification of user preferences for cultural events involves four different phases: data collection, data preparation, data processing and data visualization. In the first phase of the research, data was collected from two different sources, mobile application FilterApp, and an online survey. Two different data sources will be used to compare the expected and actual user preferences. Analytical data collected through the mobile application is a presentation of real user preferences, based on monitoring application usage patterns. The data collected through the survey represent the expected user preferences, which can be influenced by various factors, such as the desire of the respondents to present themselves in a socially acceptable way.

After data collection, it is necessary to prepare data for further research. Preparation involves summarizing survey data and scaling and normalizing obtained results. Prepared data will be presented with Excel Matrix Editor. Adjusted data is imported into network analysis software tool Ucinet and processed with two different clustering algorithms – tabu search and Johnson's hierarchical clustering method. The resulting clusters are validated and further analyzed through data visualization tools Netdraw (part of Ucinet tool) and Vosviewer. With these tools, clusters will be presented as areas within a network of event types. Alongside event type nodes, nodes for differentiation criteria will be presented in order to more easily detect cluster areas. Also, clustering density map will be presented in order to determine the validity of the resulting clusters. Characteristics of the network as distance and number of connections between nodes, their size and color intensity will be defined with survey data. After data visualization, it is possible to compare clusters identified on the created network with clusters obtained with clustering algorithms.

### 3.2. Data collection

Data for the research were collected from two sources: (1) mobile application usage analytics and (2) online survey distributed through a mobile application.

**Mobile application FliterApp**

FilterApp is a mobile application that serves to make the search for cultural events and institutions in the city of Belgrade easy, fast and simple. The app is a part of a startup project with support from the Ministry of Culture, Republic of Serbia. The main idea of the application is to bring closer the cultural offer of the city of Belgrade to visitors and to show the offer of various cultural institutions in one place. This application unifies the offer of cultural events such as movies, concerts, theatrical plays, festivals, public discussions and exhibitions of the city of Belgrade, Serbia. The main functionalities of the application are focused on filtering the offer of cultural events. Available filters are:

- event type — includes festivals, movies, public discussions, exhibitions, concerts and theatrical plays.
- institution type — contains libraries, cinemas, galleries, concerts, cultural centers, museums, theaters and cultural institutions.
- event time — filter implies any time period in the future and price range filter makes the difference between free and paid events.
- price range — common price range filter.

Further, it is possible to look at the offer and information of a particular institution. The application provides the functionality that enables event comparison. Events can be marked as favorites so that users can receive notifications about selected events.

Within the application, services for tracking usage analytics are integrated. For this purpose, Firebase Analytics [28] and Fabric Answers are used [29]. For a better understanding of the collected data and personalized experience during application usage, user authentication is enabled using the Firebase Authentication Service [30]. The application is intended for anyone interested in the cultural offer of the city of Belgrade, regardless of the socio-demographic characteristics or any other criterion.

The first data source includes information collected during the application usage. Analytical data is calculated on a user base of

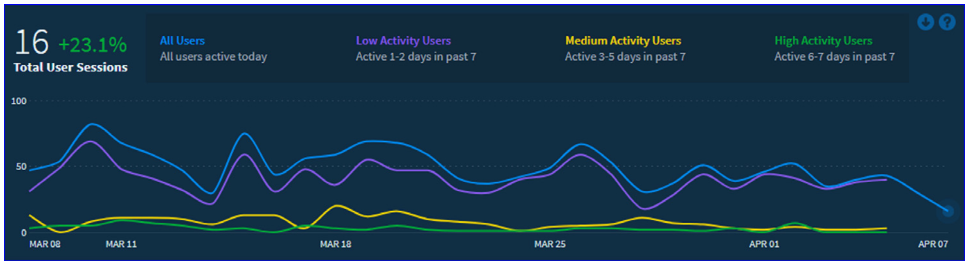



**Fig. 1.** Number of sessions by activity group (Data collected with fabric answers).

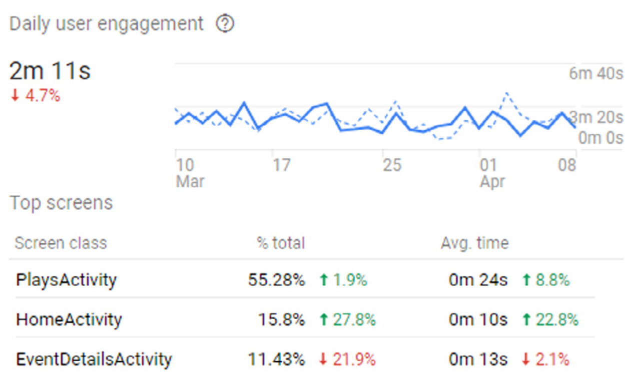



**Fig. 2.** Daily user engagement (Data collected with firebase analytics).

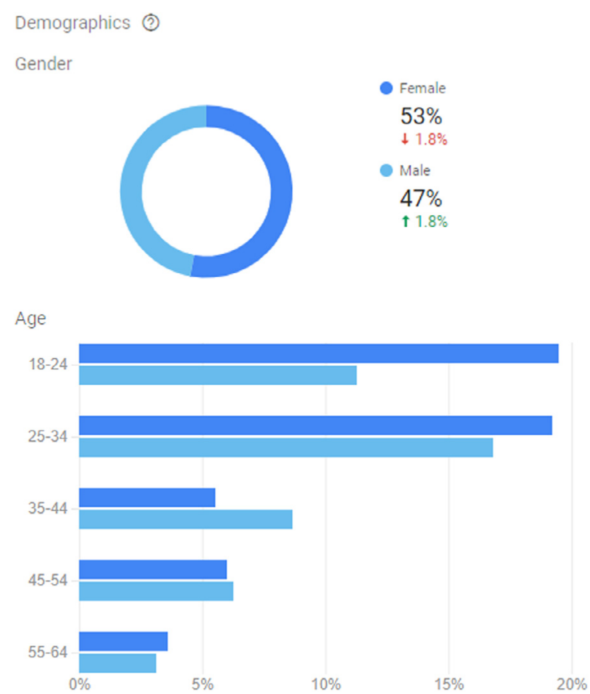



**Fig. 3.** Percentage of male and female users, by age group (Collected with firebase analytics).

over 600 members. Since the application shows cultural events in Belgrade, over 90 percent of users are located in that city. These data include information about the type and combination of filters that users use in order to find the events that interest them, most frequently reviewed institutions and events, events that are most often favored, the most frequently selected types of events and institutions, as well as plenty of other information.

In order to determine whether the sample was representative, we used two researches related to Cultural practices of citizens of Republic of Serbia [31,32]. As for the structure of our sample, most respondents are between 18 and 35 years old, which is understandable because our sample was collected by submitting a survey through a mobile application. Regarding the gender structure of the sample, as in the previous research, there is a slightly higher number of women than the number of men.

Data that influence the validity of the results is the number of user sessions during the day. Fig. 1 shows the total number of sessions during the day, with reference to the number of sessions of each of the identified group relative to the intensity of the application use (low, medium and high activity).

Fig. 2 shows average daily engagement with a graph that displays trends for the last 30 days, the name of the screen class and the average amount of time that screen was used for the time period selected [33]. The average session duration is 2 m 11 s and the most visited screen is PlaysActivity which shows filtered search results.

Percentage of male and female users by age group is presented in Fig. 3. The diagram shows that the users of the application are predominantly young (about 40% younger than 34 years). In younger age groups, the dominant gender is female while in the older age groups the case is reversed.

Data collected from the application provide information about users' preferences. Users are divided into two groups that contain different types of events. The first group represents users who prefer theaters, concerts, and festivals, while another group of users prefers public discussions, movies, and exhibits. The first group of users makes 52.4% of the total number of users while the rest of the group makes 47.6%.

**Table 2**
Questions in the online survey.

| Question | Purpose |
|---|---|
| Rank event types in relation to your preferences? | Compare the expected or pre-determined preference with the actual user preference. |
| How often do you attend events that you rank as the most important? | Determine the coefficient of the importance of a particular user on the attendance of cultural events. |
| When you visit the events that you rank as the most important, how many people do you go with? | Examine the influence of the group size on the attendance of a particular type of event. |
| How much money do you spend to attend these events on a monthly basis? | Determine the coefficient of the importance of a particular user on the attendance of cultural events. |

**Survey**

The second source of data is the survey created using Google Forms [34]. The survey consisted of the following questions (Table 2):

Using these questions, we wanted to extract the clustering criteria that we will use on the collected data. Total of 50 respondents answered the survey questions. Users were motivated to respond to the questions by getting the right to participate in the prize

**Table 3**
Data collected using a survey.

| | Score | Score (%) | Frequency (mth) | Connections |
|---|---|---|---|---|
| Movie | 53 | 25.24 | 3 | 5 |
| Concert | 46 | 21.90 | 2 | 3 |
| Theatrical play | 38 | 18.10 | 2 | 2 |
| Festival | 31 | 14.76 | 2 | 2 |
| Public discussion | 24 | 11.43 | 2 | 2 |
| Exibition | 18 | 8.57 | 1 | 1 |

```
NORMALIZE
-----------------------------------------------------

Dimension:                                Columns
Method:                                   Marginal
Diagonal valid?                           YES
Input dataset:                            raw dataset


                                   1     2     3
                               Score Frequ Conne
                               ----- ----- -----
1                    Movie     0.252 0.273 0.333
2                  Concert     0.219 0.182 0.200
3          Threatrical play    0.181 0.182 0.133
4                 Festival     0.148 0.182 0.133
5        Public discussion     0.114 0.091 0.133
6               Exhibition     0.086 0.091 0.067
```

**Fig. 4a.** Normalized data.

```
IMPORT EXCEL FILES
---------------------------------------------------------------

Input network dataset:                   Scores dataset (C:\
Excel file has row labels:               YES
Excel file has column labels:            YES
How to save output data:                 a single 3D dataset
Output dataset:                          Scores dataset (C:\
                          1 2 3 4 5 6
                          - - - - - -
   1            Movie     0 1 0 0 3 6
   2          Concert     0 0 1 4 3 2
   3    Theatrical Play   0 1 3 4 1 1
   4         Festival     0 3 3 2 1 0
   5 Public discussion    4 3 1 0 1 1
   6       Exhibition     6 2 1 0 1 0
```

**Fig. 4b.** Score matrix of event types.

```
TABU SEARCH OPTIMIZATION (TSO)
-----------------------------------------------------------------

Diagonal valid?                     YES
Number of clusters:                 2
Type of data:                       Similarities/Strengths/Cohesion
Method:                             correlation
Input dataset:                      Scores dataset (C:\Users\steva\

Starting fit: 1.062
Starting fit: 0.290
Fit: 0.290
Fit: 0.290
Fit: 0.290
Fit: 0.290          (smaller values indicate better fit.
r-square = 0.504

Clusters:

    1:  Movie Public discussion Exhibition
    2:  Concert Theatrical Play Festival

                       1 5 6   4 2 3
                       1 5 6   4 2 3
                      ---------------
 1              Movie | 6 3 6 |   1   |
 5 Public discussion  | 4 6 1 |   3 1 |
 6         Exhibition | 6 1 6 |   2 1 |
                      ---------------
 4           Festival |   1   | 6 3 3 |
 2            Concert |   3 2 | 4 6 1 |
 3    Theatrical Play |   1 1 | 4 1 6 |
```

**Fig. 5.** Tabu search results.

game for free tickets for the selected event after completing the survey. Also, participants were aware of the fact that their data is captured in purpose of this research. During the survey, users ranked preferences according to defined types of events on the scale from 1 to 6. Sum of ranks of an event type represents the event's score. After calculating the scores as a sum of the obtained ranks, all scores were divided with greatest common divisor (total sum of the scores was $50 \times 6 + 5 + 4 + 3 + 2 + 1 = 1.050$ and GCD was 5 for collected data). The frequency criterion represents the average number of visits to a particular type of event within a month by one user. In order to determine the association between people visiting cultural events, it is necessary to determine the number of persons in the group when visiting cultural events of a particular type (this criterion is named Connections in the further analysis). Table 3 presents data obtained from the survey.

In this dataset, rows represent different types of events and columns different criteria for analysis. Each data record represents the value of a particular analysis criterion for a certain type of event. Due to the scope and complexity of the data, various software tools, such as Ucinet and Vosviewer, are used for further analysis and visualization of research results [35,36].

### 3.3. Data preparation

Based on the data obtained from the survey, a dataset was created using the Excel Matrix Editor's Ucinet tool. The next step was processing the data in order to prepare them adequately for further analysis. As the data was obtained on different scales, it was necessary to normalize [37].

After the data was normalized, the results are shown in Fig. 4a. For Cluster Analysis Tabu search clustering method [38] is used in order to obtain preliminary results of event types clusters. For the purpose of this analysis, a matrix of scores for the types of events was created based on the results obtained by the survey (Fig. 4b).

### 3.4. Data processing

Tabu search clustering was done on the basis of mutual similarities and differences of network nodes. Since the matrix is not square and rows and columns do not represent the same data (rows indicate the type of event, and columns indicate the number of specific score occurrences), the diagonal elements are considered as valid in the analysis (Fig. 5).

After the analysis, event types were divided into two clusters. The first cluster contains movies as the most visited type of event, and public discussions and exhibits as the least visited types of events. This cluster represents outliers of the network. The second cluster contains festivals, concerts, and plays that are ranked among the extremes from the first cluster. This cluster represents a group of events that should be targeted if the goal of the marketing campaign is to include diverse user structure. Also, the density of the table created by clustering pointed out that difference between clusters was large. The density of the first cluster is 4.33, the density of the second cluster is 3.78, while the density between the clusters is 0.89.

The next cluster method used is Johnson's hierarchical clustering method (Fig. 6) [39].

For the purposes of this analysis, the normalized dataset from Fig. 6a was used. During the analysis, the similarities of the types of events were compared and as the comparison method, the average value was used among all couples. At the level of similarity of 6 units, the first cluster of movies and exhibitions was created. The second cluster was created at the level of similarity of 3.5 units and was made by festivals and concerts. In the last step, at the level of

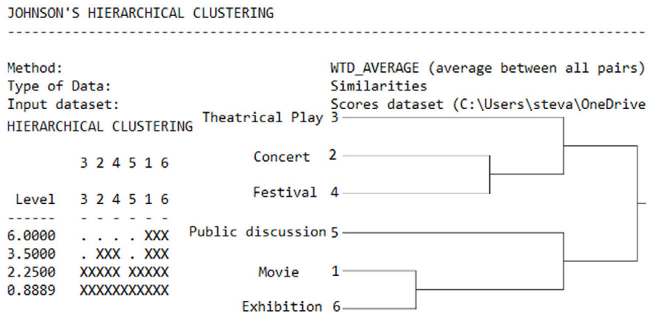



**Fig. 6.** Johnson's hierarchical clustering method results.

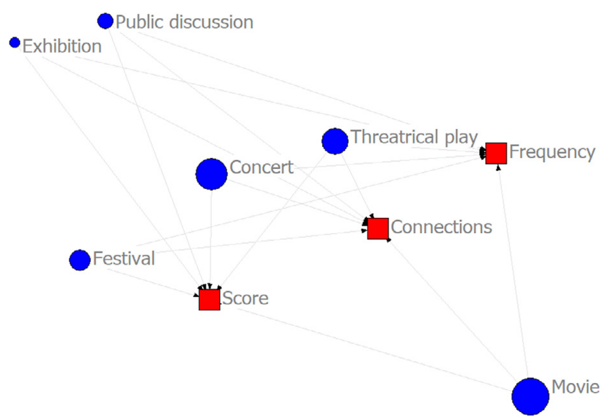



**Fig. 7.** Clusters obtained by Ucinet Netdraw data analysis.

```
*Vertices 9
1 "Movie"              0.252    0.273    0.333
2 "Concert"            0.219    0.182    0.200
3 "Theatrical play"    0.181    0.182    0.133
4 "Festival"           0.148    0.182    0.133
5 "Public discussion"  0.114    0.091    0.133
6 "Exhibition"         0.086    0.091    0.067
7 "Score"              0.240    0.230    0.000
8 "Frequency"          0.150    0.230    0.000
9 "Connections"        0.200    0.220    0.000
*arcs
1 7 0.252
1 8 0.273
1 9 0.333
2 7 0.219
2 8 0.182
2 9 0.200
3 7 0.181
3 8 0.182
3 9 0.133
4 7 0.148
4 8 0.182
4 9 0.133
5 7 0.114
5 8 0.091
5 9 0.133
6 7 0.086
6 8 0.091
6 9 0.067
*Edges
```

**Fig. 8.** Pajek network file.

similarity of 2.25 units, public discussions were added to the first cluster and theatrical plays were added to the second cluster.

## 4. Results

### 4.1. Data visualization

The results obtained by hierarchical clustering method match the results obtained by the Tabu search method. The result of the analysis is represented by two clusters in which networks outlier were clearly separated from the central nodes. Validation of previously identified clusters is performed through the graph that presents all types of events, as well as all the criteria used for clustering. Normalized data is used as the coordinates of the nodes within the graph. Nodes representing the types of events in the graph are represented by blue circles while the clustering criteria are represented by the squares of the red color (Fig. 7).

In the chart, criteria are represented with red squares while events types are represented with blue circles. It can be noticed that the movie as the most visited type of event with the largest groups of visitors and the most frequent visits distinguishes from the rest of event types. On the other hand, exhibits and public discussions are also distinguished from others as the least visited types of events. Festivals, plays, and concerts have more similarities with each other than with the other types, so it is obvious that they represent a separate cluster. The graph confirms results from tabu search method and hierarchical clustering method.

With collected data and documentation provided in [40], a Pajek network file is created (.net) using Vosviewer tool. The created dataset is defined by analogy to the example from [41]. An explanation of how to create a dataset in the Pajek network file format is shown in Fig. 8.

Within research, we used normalized scores of the corresponding criterion as the weighted coefficients of the relationships. After map loading, before analyzing the data, the map looks as shown in Fig. 9a, and after normalization as in Fig. 9b. First of all, we use the data normalization with association strength method [42]. It is necessary to change the additional parameters, that is, to increase the sampling precision in order to get the best possible results in terms of the sample size.

The next step is data analysis, i.e. the application of clustering over defined groups. Before using this functionality, it is necessary to define Resolution and Min. cluster size parameters. The first parameter affects the total number of clusters on the map, while the second parameter determines the minimum size of each cluster. In our research, since the types of events are clustered in relation to the three criteria, the resolution parameter is set to 3. Since 6 predefined types of events are included in the survey, the minimum cluster size is obtained when the total number of types of events is divided by the total number of dividing criteria – each cluster contains at least 2 nodes. The final map, after normalization and clustering, is presented in Fig. 10.

The map in Fig. 11 shows three clusters, each in a different color. Cluster one includes exhibits, concerts, and score. Cluster two contains festivals, plays, and frequency. Cluster three includes movies, public discussions, and connections. Display of density of elements and clusters density is shown in Fig. 11.

Density map shows that the concert and exhibition cluster is highly incoherent, as opposed to cluster consisted of theatrical play and festival. Cluster with movies and public discussions has high density because it contains the movie as the most dominant type of event.

## 5. Discussion and conclusion

This paper portrays an approach for social network analysis related to user preferences for cultural events. Data were collected from two different sources: mobile application FilterApp and an online survey. Main conclusions from the analysis are:

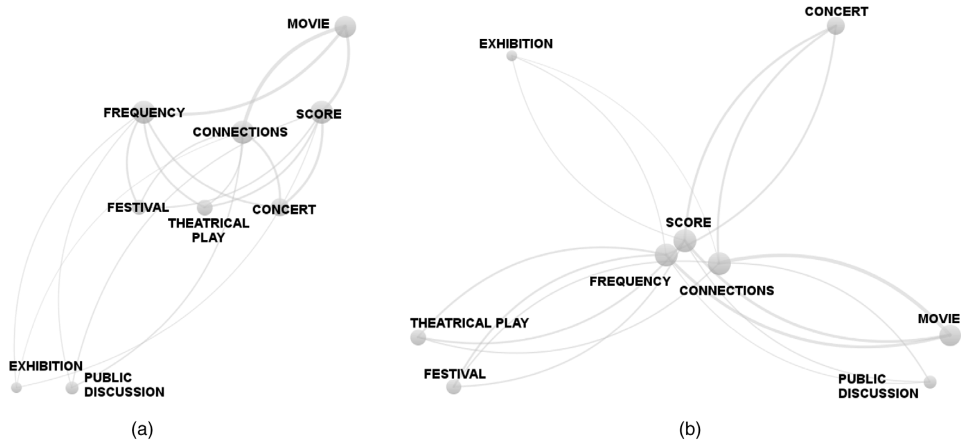



**Fig. 9a and b.** Vosviewer map before and after normalization.

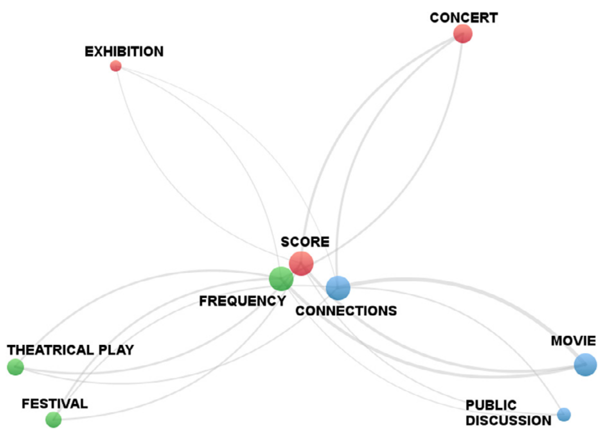



**Fig. 10.** Vosviewer map after cluster analysis.

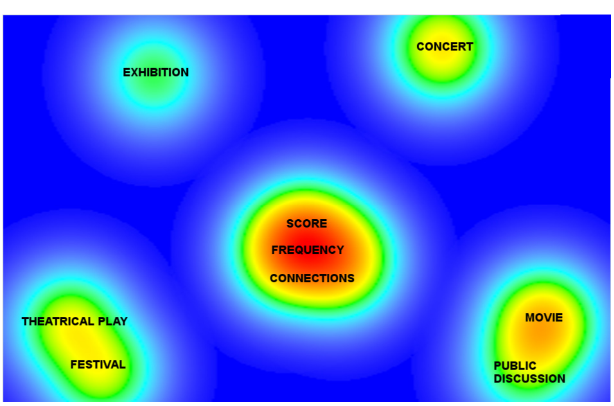



**Fig. 11.** Clustering density map.

- *Cluster in which the dominant segmentation is based on the size of a group of users* includes movies and public discussions. Their score for group size criterion is higher than the scores related to preferences or frequency. However, a score of public discussions is significantly lower in relation to the score of a movie, so cluster incoherence is clearly noticed.
- *Cluster in which the dominant segmentation is based on the frequency of the visit* includes plays and festivals as event types whose highest scores refer to this criterion. Also, the scores of these two event types according to the criterion of the frequency of visiting cultural events coincide (0.182), so this cluster is the most coherent among all.
- *Cluster in which the dominant segmentation is based on user preferences* includes concerts and exhibits. User preferences score for a concert is dominant and hence the clustering of the concert in this cluster is justified. On the other hand, exhibit, as the least visited type of event and the worst rated in all categories, appears to fit into this cluster due to the previously defined limit for the minimum number of elements in each cluster.

The main idea was to prove that our methodology for SNA can be used in area of culture and art. Specific conclusions related to the cultural events themselves were not of biggest importance for the research. Key contribution of the paper is reflected in the fact that the proposed approach could be used in different contexts and environments, particularly when data from common social network websites are not available or not adequate for the analysis. Further, steps, tools, and methods described within the paper could be useful for marketing activities within small and medium enterprises, startups, new product development departments where it is extremely demanding to get data related to customers behavior and preferences. The paper aims to make an impact on practice related to usage of SNA.

Numerous researchers and practitioners connect SNA with data from well-known social networks [20,24]. Furthermore, this paper allows immediate validation of research findings thanks to the use of information collected through the mobile application analytics service. Other papers, such as [19,20,23], require confirmation of results through further research and validation. The results obtained with research based on an online survey about users' preferences coincide with the results of the analytics analysis of the mobile application.

Limitations of this work in relation to the reviewed literature refer to the inability to identify the influential users in the network. The results of this paper related to the cultural events are not fully generalizable as they depend on several factors such as the economic standard of the population, the marketing activities of the cultural institutions, the habits of the users and many others. However, cultural events themselves were used as a proof of concept where we could test our approach.test. Accordingly, we did not focus on other researches in cultural industry. The main value of this paper is reflected in the clearly defined research process with all the phases from data collection to validation of results. The obtained conclusions are used to develop system of recommendations and for customization of offers and marketing strategies to the identified users' preferences. Promotions for group visits are created for movies and public discussions, monthly and season tickets are introduced for plays and festivals and groups of tickets for thematic units for concerts and exhibits. In future work, we will endeavor to examine the correlation between users' preferences in the social network of mobile app and user preferences in global social networks.

## Acknowledgment

Authors are thankful to Ministry of education, science and technological development, Republic of Serbia, grant 174031.

## References


[1] S. Chen, G. Wang, W. Jia, Cluster-group based trusted computing for mobile social networks using implicit social behavioral graph, Future Gener. Comput. Syst. 55 (2016) 391–400.

[2] M.D. Fernando, A. Ginige, A. Hol, Structural and behavioural models for social computing applications, in: 27th Australasian Conference on Information Systems, No. 53, University of Wollongong, Wollongong, NSW, 2016.

[3] D. Selloni, CoDesign for Public-Interest Services, Milan, 2017.

[4] A. Abraham, A. Hassanien, Computational Social Networks: Tools, Perspectives and Applications, Springer Science & Business Media, Auburn, 2012.

[5] S. Dasgupta, Social Computing: Concepts, Methodologies, Tools, and Applications, IGI Global, New York, 2009.

[6] M. Parameswaran, A.B. Whinston, Social computing: An overview, J. CAIS 19 (2007) 37.

[7] I.H. Osman, A.L. Anouze, A. Emrouznejad, Handbook of Research on Strategic Performance Management and Measurement Using Data Envelopment Analysis, IGI Global, 2014.

[8] M. Khosrow-Pour, Social Computing in Encyclopedia of Information Science and Technology, third ed, USA, 2014, p. 6754.

[9] M. Fernando, A. Ginige, A. Hol, Enhancing business outcomes through social computing, J. IADIS Int. J. 14 (2016) 91–108.

[10] V. Podobnik, I. Lovrek, Implicit social networking: Discovery of hidden relationships, roles and communities among consumers, J. Procedia Comput. Sci. 60 (2015) 583–592.

[11] R. Gross, A. Acquisti, Information revelation and privacy in online social networks (The Facebook case), in: ACM Workshop on Privacy in the Electronic Society, 2005, pp. 71–80.

[12] R. Cachia, Social Computing, Study on the Use and Impact of Online Social Networking, Luxembourg, 2008.

[13] L. Weber, Marketing to the Social Web, How Digital Customer Communities Build Your Business, second ed, John Wiley & Sons, Hoboken, NJ, USA, 2009.

[14] B. Radenković, M. Despotović-Zrakić, Z. Bogdanović, D. Barać, A. Labus, Elektronsko poslovanje, Fakultet organizacionih nauka, Beograd, 2015, p. 186.

[15] M.A. Brandão, M.M. Moro, Social professional networks: A survey and taxonomy, J. Comput. Commun. 100 (2017) 20–31.

[16] S. Schlattmann, Capturing the collaboration intensity of research institutions using social network analysis, Procedia Comput. Sci. 106 (2017) 25–31.

[17] M. Erfanmanesh, E. Hosseini, Using social network analysis method to visualize library & information science research, J. J. Adv. Inf. Technol. 7 (3) (2016) 177–182.

[18] L. de Marcos, E. García-López, A. García-Cabot, J. Medina-Merodio, A. Domínguez, J. Martínez-Herráiz, T. Diez-Folledo, Social network analysis of a gamified e-learning course: Small-world phenomenon and network metrics as predictors of academic performance, J. Comput. Hum. Behav. 60 (2016) 312–321.

[19] Y. Mo, B. Li, B. Wang, L.T. Yang, M. Xu, Event recommendation in social networks based on reverse random walk and participant scale control, Future Gener. Comput. Syst. 79 (2018) 383–395.

[20] C.C. Chen, Y.-C. Sun, Exploring acquaintances of social network site users for effective social event recommendations, Inform. Process. Lett. 16 (2016) 227–236.

[21] C. Xu, A novel recommendation method based on social network using matrix factorization technique, Inf. Process. Manage. 54 (2018) 463–474.

[22] Y. Koren, R. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, Computer 42 (2009) 30–37.

[23] C. Ruzol, D. Banzon-Cabanilla, R. Ancog, E. Peralta, Understanding water pollution management: Evidence and insights from incorporating cultural theory in social network analysis, Global Environ. Change 45 (2017) 183–193.

[24] Y. Jhamb, Y. Fang, A dual-perspective latent factor model for group-aware social event recommendation, Inf. Process. Manage. 53 (2017) 559–576.

[25] S. Milovanović, Z. Bogdanović, A. Labus, D. Barać, M. Despotović-Zrakić, Using social network analysis to identify user preferences for cultural events, in: WorldCIST'18 2018, in: Advances in Intelligent Systems and Computing, vol. 745, 2018, pp. 644–653.

[26] T.W. Valente, Social Networks and Health: Models, Methods, and Applications, Oxford University Press, 2010.

[27] J. Luque, D. Martinez Tyson, J. Lee, C. Gwede, S. Vadaparampil, S. Noel-Thomas, C. Meade, Using social network analysis to evaluate community capacity building of a regional Community Cancer Network, J. Community Psychol. 38 (2010) 656–668, http://dx.doi.org/10.1002/jcop.

[28] Log Events. https://firebase.google.com/docs/analytics/android/events, 2018 (accessed 09.05.18).

[29] Answers Events. https://docs.fabric.io/android/answers/answers-events.html, 2018 (accessed 09.05.18).

[30] Manage Users in Firebase https://firebase.google.com/docs/auth/android/manage-users, 2018 (accessed 09.05.18).

[31] P. Cvetičanin, M. Milankov, Kulturne prakse građana Srbije, Zavod za proučavanje kulturnog razvitka, 2011.

[32] B. Opačić, B. Subašić, Kulturne potrebe i navike građana Srbije, Zavod za proučavanje kulturnog razvitka, 2016.

[33] Firebase Answer–Daily user engagement. https://support.google.com/firebase/answer/6317517#engagement, 2018 (accessed 05.05.18).

[34] FilterApp - Preferencije korisnika - test grupa Beta #2. https://goo.gl/forms/lDn7Lnvh6nHmghcq2, 2018 (accessed 05.05.18).

[35] S.P. Borgatti, M.G. Everett, L.C. Freeman, Ucinet for windows: Software for social network analysis, J. Anal. Technol. (2002) Harvard, MA.

[36] N.J. Van Eck, L. Waltman, VOSviewer Manual, 2018, http://www.vosviewer.com/documentation/Manual_VOSviewer_1.5.4.pdf (accessed 05.05.18).

[37] UCINET 6 for Windows Help–Normalization, 2018 http://www.analytictech.com/ucinet/help/1mmzzrm.htm (accessed 05.05.18).

[38] UCINET 6 for Windows Help–Cluster optimisation, 2018 http://www.analytictech.com/ucinet/help/2cvtid.htm (accessed 05.05.18).

[39] UCINET 6 for Windows Help–Hierarchical cluster analysis, http://www.analytictech.com/ucinet/help/3j.x0e.htm, 2018 (accessed 05.05.18).

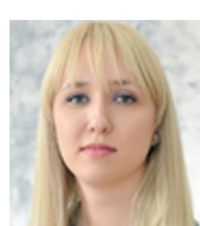
[40] V. Batagelj, A. Mrvar, Pajek - Program for analysis and visualization of large networks. Timeshift - the World in Twenty-Five Years: Ars Electronica, 2004, pp. 242–251.

[41] Pajek NET Format. https://gephi.org/users/supported-graph-formats/pajek-net-format/, 2018 (accessed 05.05.18).

[42] N.J. Van Eck, L. Waltman, How to normalize cooccurrence data? An analysis of some well-known similarity measures, J. Amer. Soc. Inf. Sci. Technol. (2009) 1635–1651.

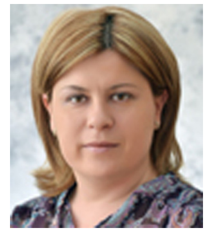

**Stevan Milovanović** is PhD student at Faculty of organizational sciences, University of Belgrade. His areas of interest are: Software engineering, mobile application development, social network analysis and web solutions.

**Zorica Bogdanović** is assistant professor at Faculty of organizational sciences, University of Belgrade . Her areas of interest are: data science, digital marketing, web development and mobile application development.



**Aleksandra Labus** is assistant professor at Faculty of organizational sciences, University of Belgrade . Her areas of interest are: social networks, data science, social network analysis, IoT, mobile business.

**Dušan Barać** is assistant professor at Faculty of organizational sciences, University of Belgrade . His areas of interest are: web programming, e-commerce, mobile business and digital marketing.



**Marijana Despotović-Zrakić** is full professor at Faculty of organizational sciences, University of Belgrade . Her areas of interest are: web programming, mobile business, digital marketing and social networks.