The Analysis of Teenagers' Vlogging Preferences in Educational Research

Smaranda CIOBAN¹,
Adrian HATOS²

¹ Doctoral School of Sociology, University of Oradea, smaranda.cioban@gmail.com

² Doctoral School of Sociology, University of Oradea, ahatos@gmail.com

Abstract: Theories on media consumption show that online content influences users' attitudes and behavior. We employed open-source instruments of data visualization in order to build a network of vloggers based on teenagers' preferences. The constructed network displays communities of users based on the YouTube content they follow. Used together with traditional research techniques, the method constitutes an effective tool in identifying consumption patterns of various groups, including students found at risk of engaging in deviant behavior or of school dropout. The revealed patterns may be explored afterwards with content analysis tools. Grounded in network science, the method also proposes ways of reducing the diffusion of problematic content inside networks and eventually of ameliorating the engagement in deviant behavior and of reducing school dropout.

Keywords: *Gephi; YouTube vloggers; online content; education; students at risk; consumption patterns;*

How to cite: Cioban, S., & Hatos, A. (2019). The Analysis of Teenagers' Vlogging Preferences in Educational Research. *Revista Romaneasca pentru Educatie Multidimensionala*, 11(2), 38-58. doi:10.18662/rrem/116

Introduction

The study presents a practical example of how to use open-source instruments of data visualization and network science for classifying teenagers in relation to their YouTube preferences. Developed on the assumptions of network studies and employed in market research for segmenting different types of consumers, the method we are presenting here may serve as an effective tool in identifying groups of units that have particular characteristics. As it allows a rapid observation of specific traits, this type of classification may be of great help in identifying types of students, for example those at *educational* risk or types of content, for example negative/deviant contents. Its focus on teenagers' YouTube consumption in terms of vloggers is motivated by the platform's popularity in Romania. As early as 2014, in Romania, 86% of children used YouTube (Mascheroni and Ólafsson 2014).

Throughout the research, we employed only open-source instruments. The data was collected by students from the University of Oradea coordinated by prof. dr. Adrian Hatos and lect. dr. Raluca Buhaș using a Google Form from a convenience sample of teenagers from Bihor County. After data collection, the database was downloaded as a CSV file and the results were imported in Gephi 0.9.2 version for analysis. With Gephi, we constructed a bipartite graph (out of respondents and vloggers - each child chose his top 5 vloggers) and we established connections between vloggers (if they were voted by the same person) by using the modularity function available in Gephi. As we will further develop in the paper, the modularity function is a classification tool that facilitates the identification of hidden structures inside a network (Newman, 2016).

With the increased use of ICT, the applied research from various fields began to incorporate software technology in order to obtain more in-depth analysis. The open-source software Gephi (Bastian, Heymann, and Jacomy 2009) constitutes such a tool. Gephi is a user-friendly software, which provides powerful visualizations and statistical network analysis for large sets of data available in various formats, including .xlsx, .csv, .gexf, .gdf, GraphML, Pajek NET, gml.

Its applications range from biology and medicine to social studies and sociology, as it brings benefits both in the study of biological networks and of social networks, micro and macro groups. Its layout and community detection algorithms facilitate the representation and understanding of all kinds of networks, providing tools for community mapping and entities'

classification. It also allows the identification of hidden patterns inside a network (Modularity Algorithm) and a visual representation of the influence of each entity within a network (scaling the size of entities in relation to their degree, connections, neighborhood, ranking, filtering the main entities in relation to specified algorithms).

This paper provides an example of how to use Gephi for studying social networks and identify hidden communities inside the network. We decided to employ Gephi because it provides the most used statistics in network analysis (centrality, degree, betweenness, closeness, path length, ranking), elaborate graphical representations of entities and implements an efficient algorithm for identifying the structures inside the network, namely the Louvain Algorithm (Blondel et al. 2008) which is available by running the Modularity Function. As Blondel et al. (2008) explain, the Louvain algorithm relies on the neighborhood relations and similarity within entities, out of which the algorithm decides to place or not to place them inside a community (in the beginning, each entity is taken as a community). The next step is building a new network composed of the formed communities. The two steps repeat until modularity within entities is maximized and the smallest number of communities that include similar entities is achieved. Afterwards, communities may be graphically represented by selecting in the node panel the modularity function. For a clearer visualization, entities may be filtered based on the modularity classes they belong to. We applied this method in order to classify teenagers in relation to their consumption habits in matter of vlogs.

Vlogging and The Youtube

YouTube is the most popular video sharing application (Figueiredo, Benevenuto, & Almeida 2011). Since its founding in 2005, YouTube quickly became a platform followed by billions of people. At the same time, YouTube is perceived as a social-network, allowing users to communicate through its facility of writing comments to the posted videos, (Bou-Franch et al. 2012), community platform and the instant messaging option.

Concerning its popularity in Romania, the study concerning the use of Internet shows that 90% of Romanian teenagers use YouTube (Salvați Copiii, 2019) The latest report EU Kids Online also witnesses the rise of instant messaging applications and media sharing platforms (Velicu, Balea & Barbovschi, 2019). The presented findings may serve as a point of departure for analyzing Internet use and YouTube content in particular.

Vlogging is the content that gains most of YouTube users' attention, encompassing 40% of the most popular videos in terms of visualizations,

rank as favorites and challenging discussions (Burgess & Green 2018). It (conversational video blogging) constitutes a way through which people share experiences, opinions and ideas. Becoming a vlogger is accessible for people, as it requires Internet connection, a digital camera and basic editing skills, which are even integrated in Social Media Platforms.

As early as 2007, Molyneaux defined vlogs as *user-generated forms of online communication that serve as media for social commentary, alternative newscasts, creative outlets or personal online diaries* (Molyneaux et al. 2007). As Biel emphasized (Biel and Gatica-Perez 2009), vlogging is usually associated with YouTube, representing a characteristic that distinguishes Youtube among other distribution systems. Its popularity ranges from not only the presented text, but also from non-verbal cues. In this respect, Biel and Gatica-Perez (2013) point out that vlogs provide an opportunity of analyzing users in social media, as their creators share information about themselves that text cannot express only by itself.

Acknowledging its increased importance in understanding what happens online, scholars focused on YouTube vlogs in their attempt to characterize both producers' characteristics (Ferchaud et al. 2018; Biel and Gatica-Perez 2010.; 2013; Mogallapu 2011) and users' specific features (Bou-Franch et al. 2012b; Yoganarasimhan 2011; Molyneaux, O'donnell, & Gibsonanice 2008; Molyneaux et al. 2007; Biel and Gatica-Perez 2010; Haridakis, Haridakis, & Hanson 2009). Put together, this data creates a profile of the YouTube user in terms of wants and needs. In this respect, a question asked by Chiang and Hsiao (2015) is what makes users stick to YouTube, respectively the main motivations of using and sharing content.

A variety of factors, including interactivity, reputation, altruism, reciprocal benefits, social norms, community identification, video creation ability and video sharing self-efficacy thus explain the increased use of video sharing platforms (Chiang and Hsiao 2015). Grounded on the uses and gratification theory, which holds that viewers actively select the content they watch (Rubin 2009, Ruggiero 2000, So 2012), studies reflect the key dimensions related to consumers' motivation to use Internet (Stafford et al 2004), social networking sites (Nelson 2016; Whiting & Williams 2013; Sundar & Limperos 2013), YouTube (Shao 2009; Khan 2017). In this direction, Shao (2009) assumes that user-generated media are appealing because they are "easy to use" and "let user control", conferring motivations related to entertainment, interaction and self-expression. Khan (2017) provides a more detailed classification of motivations to use YouTube, namely: seeking information, giving information, self-status seeking, social interaction, relaxing entertainment. Still, the content that is watched shapes

users' attitudes and behaviors, especially for intensive users (Gerbner 1998; Stefanone, Lackaff, & Rosen 2010; Morgan, Shanahan & Signorielli 2014, 2017). This is conceptualized in the scientific literature as the process of cultivation and Gerbner and Larry Gross framed it in relation to TV viewing (Gerbner & Gross 1976). Adding more data to this description, especially with relation to the types of content that are accessed, the types of practices, the effect of content consumption brings benefits in analyzing groups and communities that included Internet in their daily lives. In the following sections, we attempt to build a profile of the Romanian teenagers that use Internet, based on the existent literature in the field.

Profiling in Social Sciences

Profiling is a method largely used in social sciences and in social intervention, especially for distinguishing specific traits that signal a potential risky position or a protective factor. Thus, scholars aim towards creating profiles, which may largely explain the behavior of people with specific traits. Used in a variety of fields, ranging from biology, medicine, criminology, psychology, sociology, market research and educational sciences, the method has its origins in the theory of market segmentation. As early as 1956, Wendell introduced the term market segmentation, defining it as *viewing a heterogeneous market (one characterized by divergent demand) as a number of smaller homogenous markets in response to differing product preferences among important market segments* (Smith 1956). Market segmentation theorists followed two directions, grounding their research either on microeconomic theory or behavioral sciences (Wedel & Kamakura 2012 p 6).

Starting with the 1990's, marketing researchers envisioned the potential of network structures in companies' development, segmenting markets and internationalization (Coviello & Munro 1995). In this sense, the assumptions and methods of the newly emerged science of networks become incorporated into sophisticated procedures of accurately distinguishing between different categories of consumers, along to algorithms of classification, clustering, dynamic segmentation, etc. But, improving accuracy in distinguishing particular traits is not of great interest only to market specialists. It may be applied to a variety of fields, including the broader domain of social sciences. Hence, identifying specific features brings benefits in signaling at risk or deviant behaviors, especially when it deals with the online. These type of classifications add value to education research, which in the last years attempted to apply algorithms of network science and data visualization tools in educational studies in order to achieve

a better description of social phenomena (Badge, Saunders & Cann, 2012; Grunspan, Wiggins & Goodreau, 2014; Paluck, Shepherd & Aronow, 2016).

Romanian Teenagers and Youtube. A Profile of the Internet User

As Simona Ștefănescu (2008) shows, Romanian teenagers quickly incorporated Internet in their daily schedule researchers employing the term “the ubiquity of the internet” (Velicu, Balea & Barbovschi, 2019). Thus, the online represents for them the main or even the unique way of spending free time (Ștefănescu 2008), as almost half of them – 48% - spend more than 6 hours online everyday (Salvați Copiii, 2019). In Romania, teenagers mostly use Internet for entertainment and socialization, rather than for acquiring information, the main mentioned motivation among Internet users of this age being socialization (Diana Dămean in Maria Diaconescu, Monica Barbovschi 2008, pp 11-20; Velicu, Balea & Barbovschi, 2019). Compared to their European counterparts, Romanian youngsters and teenagers have higher scores on what concerns the daily time spent online, but significantly lower ones with regard to the number of activities that they do online (Mocanu 2018; Ragnedda & Kreitem 2018; Balea 2016; Tufă 2010; Mascheroni et al. 2013; Ragnedda & Muschert 2013). Both the use of Internet mostly for entertainment and socialization and the low number of types of activities realized when online reveal the obstacles that Romanian children encounter in benefitting from the use of Internet, as compared to their European counterparts, term coined in literature as digital divide (Ragnedda & Kreitem 2018; Scheerder, van Deursen, and van Dijk 2017; van Deursen & van Dijk 2014; Hargittai 2010; Van Dijk & Hacker 2003). Moreover, inequality is even more visible in vulnerable groups and communities, considering that the existent inequalities are deepened by networks’ mechanisms (DiMaggio & Garip 2012).

The differences in usage leave teenagers with lower online skills vulnerable to accessing harmful content and encountering online risks, such as: cyberbullying, the non-voluntary and voluntary exposure to pornographic content, the exposure to negative user generated content (NUGC), online victimization, meeting people known through social networks, etc. These contents are easier to be searched and opened because they do not require a high degree of sophistication in using Internet and they are found on the first pages (Hargittai 2010; Hargittai et al. 2000; Hargittai and Hinnant 2008). In addition, the research of Colletto, Aiello, Lucchese and Silvestri (2017) reveals that deviant contents propagates in a larger network than the one of diffusion. Focusing on a adult content consumption network as a deviant

community, the above mentioned authors explore the interaction between network structures and external factors, pointing that the deviant content gets to a high number of users that are unintentionally exposed to (Coletto et al. 2017). Therefore, research should focus on deviant content and understanding its impact on one hand and analyzing its distribution network on the other, facts that considerably increase the chance of identifying users that got exposed to harmful content or identify risky practices.

Our study follows this line of thinking, studying a population of vloggers, more exactly the aggregate pool of vloggers ranked as top 5 favorites by almost 200 teenagers from Bihor. In network terms, a connection between vloggers means that those vloggers were chosen by the same user. A practical use of this network is in observing how a content from a vlogger propagates in a network. What is more, this method serves in accurately distinguishing the contents which have the highest probability to determine the unwanted exposure to a deviant content.

Research Methodology

The data used in this research were collected during November – December 2017 by the students from the department of Sociology and Social Work, Faculty of Social Sciences, University of Oradea using an online questionnaire. The topic of the research was the exploration of Internet use and vlogging preferences among teenagers of 11-19 years from Bihor County. The 196 respondents were chosen by the students using convenience sampling. The recorded responses were downloaded as a CSV file and analyzed with Gephi 0.9.2 version.

The study aims to describe the use of Internet among teenagers focusing on their vlogging habits and to classify respondents based on their favorite vloggers. At the same time, the study relies on the assumptions of the theory of uses and gratification (Rubin 2009; Ruggiero 2000; So 2012), which postulates that users select the content in accordance to their preferences. Even if we point out that consumers' are choosing what they watch, we have to acknowledge that content plays a role in shaping people's attitudes and behaviours, as cultivation theorists argue (Michael Morgan, James Shanahan 2014; Morgan, Shanahan, and Signorielli 2017; Stefanone, Lackaff, and Rosen 2010).

Therefore, it is highly probable that the choice of vloggers shapes children's attitudes and behavior. In this context, the challenge of this study was to identify groups of influencers on the basis of children's options. The respondents were asked to provide data about Internet access and to name

and rank their top five vloggers. As a result, 306 vloggers were identified, out of which 56 were considered relevant and were included in the second stage of analysis. The relevance was established based on the number of mentions received by a vlogger. Hence, 56 of them got at least four mentions out of the 980 rankings expressed by the 196 questioned teenagers.

At first, the obtained data were analyzed through descriptive statistics, such as Cross tabulations and frequencies. The second stage was constructing a network, which includes the respondents and the relevant vloggers. Each mention of a vlogger corresponded to a connection between the vlogger and a teenager and an edge in the network while the 56 selected vloggers represented nodes. Also, a connection was established between two vloggers if they were voted by the same person. The structure of the communities was visualized with the use of Gephi and then the modularity was calculated with the same software (Gephi). The modularity function is used for the identification of hidden patterns within the network (Danon et al. 2005), which constituted an objective of this research. Modularity measures the links inside a community as compared to the links between communities. The function assumes the existence of some groups within which the entities are more connected than outside them.

Population and Sample

The research inquires the way teenagers from Romania access digital content and YouTube content in particular. The study was conducted on a sample of 196 teenagers that watch vlogs, aged 11-19, from Bihor county. The study used opportunity sampling, as students selected the respondents from their acquaintances (persons that they easily have access to). Most of the respondents (64.3%) are from Oradea. There are slightly more girls than boys (55.1% girls) and more respondents aged 15 to 19 (58.2%) than 10 to 14. More than 70% (77.1%) of the children come from urban areas.

Results and Analysis

General Findings

The first set of questions deal with Internet access. The first one shows that teenagers have access to Internet from a range of devices, such as mobile phones (including smartphones), tablets, PC's and laptops. Almost all participants (98%-192 out of 196) have access to Internet from their mobile phones. This is in accordance to recent findings, as the spread of mobile use and smartphones among children and teenagers was witnessed by researchers from 2014 (Livingstone, Haddon & Olafsson) and after 2017 it

became a certitude in EU countries (Mascheroni & Holloway 2017) and in Romania (Velicu, Balea & Barbovschi 2019; Salvați Copiii, 2019). The only respondents that answered negatively to this question are three girls and one boy from Oradea. The majority of respondents (73%-143 out of 196) access the Internet from their mobile phones. The second preference is the laptop at a higher distance (11.2%- 22 out of 196), while the third one is the tablet (9.2%- 18 out of 196). The last in the top of preferences comes the personal computer, with a 6.1 % (12 out of 196 respondents). Although the sample is non-representative, the results of our study illustrate the mobility and privatization of Internet access, thus coinciding to the data provided by Net Children Go Mobile (Mascheroni and Ólafsson 2014)

With regard to the location of access, most of the respondents (97.4%) access their favorite vlogs from home. At the same time, a significant percent of the persons access their favorite vlogs from more than one location, considering that 49% affirmed that they access YouTube when they meet with friends and 41.8% that they access digital content from school. In addition, 44.4% access their favorite vlogs in other situations. The data reveal the diversification of ways and locations where teenagers go online, showing the increased mobility of Internet use, results also shown by Net Children Go Mobile project and by the recent data of EU Kids Online (Velicu, Balea & Barbovschi 2019).

The table below reveals the most frequently used devices to go online in relation to the locations of accessing one of the most popular online category of content, namely YouTube vlogs. The mostly used devices are mobile phones (including smartphones) in each location. At the opposite corner, we witness a significant decrease in the use of personal computers in different places.

Table 1. Devices frequently used to go online and Locations of accessing favorite vlogs. Personal work

Devices Internet Access	Locations of accessing favorite vlogs			
	At home	Meeting with friends	At school	Other places
Other devices	0.50%	0%	0%	0.50%
Laptop	11.20%	7.70%	5.10%	2.60%
PC (personal computer)	5.60%	4.60%	2.60%	2.60%
Tablet	8.70%	4.60%	4.60%	3.10%
Mobile phone	71.40%	32.10%	29.60%	35.70%

Total	97.40%	49%	41.80%	44.40%
-------	--------	-----	--------	--------

Along to Internet access and location of access, the participants answered a series of questions about their YouTube experience. One of the questions referred to the manner in which they interact with the content they follow. The interaction with a specific content is defined as engagement (Shao 2009) and it varies from a passive approach (just for information purposes-rate the content, save to their favorites) to ones that are more active (the highest level of engagement is creating a new content in response to the video). A large share of respondents (63.3%) access vlogs for getting information (rate, save to their favorites), which is the lowest level of engagement with a content. Still, most of the respondents engage with their favorite content in some manner. Thus, 158 out of 196 (80.6%) hold that they “Like” the content, which is a second level of engagement. Concerning active forms of engagement, a significantly lower percent of respondents choose these alternatives. Thus, 60 out of 196, which corresponds to 30.6% “Share” a content and only 46 (23.5%) leave a comment or a new post.

The other questions focused on YouTube experience asked the teenagers to rank their top 5 favorite vloggers. Based on their answers there were identified 306 vlogs, out of which only 56 were mentioned more than four times. The top 12 vloggers were the following: CODRIN BRADEA, BIANCA ADAM, ZMENTA, ILIE’S VLOGS, SELLY, MIKEY HASH, LIKE ONE, BROMANIA, NOAPTEA TÂRZIU, IONUT RUSU, VLAD MUNTEANU, MAXINFİNITE. (See Figure 1: Top 12 Vloggers).

Figure 1 Top 12 vloggers. Personal Work

Moreover, we checked if there is a possibility that the vlogging' preferences were associated with certain users' characteristics. The data showed a gender pattern, considering that some vlogs were chosen only by girls, especially vlogs produced by female vloggers; while others with male authorship got mentions predominantly from boys. Other features related to content are associated obviously with this gendered pattern of preferences.

In addition, vlogging choice seems to be influenced by age. Vloggers such as: IRINA'S VLOGS, ANDRA GOGAN, ALEX ALVAREZ, DIANA C, VLAD MUNTEANU, DAVID DOBRIK, ITSANCA, IOANA114 are exclusively or predominantly chosen by respondents aged 10-14, which represent 41.8% of the total sample. Other vloggers, such as CARMEN GREBENISAN, DOI DEGEABA, DORIAN POPA, MARILU, PEWDIEPIE, CREATIVE MONKEYZ ARMY, ALINA CEUSAN, CRISTINA ALMĂȘAN, MIKEY HASH, ADRIAN POV, FLORENTIN HRISCU, JAKE PAUL, LECTURA DELAA-LAZ, MAXSIALTELE, ZAIAFET, are clearly preferred by respondents older than 14.

Constructing and Exploring the Vlogging Network

We constructed a bipartite network with vloggers and respondents in Gephi by using the Fruchterman-Reingold layout, Area 5000.0, gravity 5.0. The network is composed of 241 nodes, out of which 56 are the relevant vloggers and 621 edges, as all the edges that led to other vloggers than the relevant ones were removed. The resulted network has an average degree of 2.583, an average weighted degree of 8.154 and a graph density of 0.011 . By looking at the network, we observed that the nodes which coincide with the most influential vlogs have the highest weighted Degree (CODRIN BRADEA- 176; BIANCA ADAM - 132, ZMENTA: 130, MIKEY HASH: 110, ILIE'S VLOGS:103, SELLY: 95) and the highest eigenvector centrality (CODRIN BRADEA -0.66; BIANCA ADAM and ZMENTA -0.52; ILIE'S VLOGS- 0.47; SELLY -0.44; MIKEY HASH 0.43). These measures show the connection between nodes, respectively their importance. We also realized graphs that show the preferences for the most popular vloggers by gender, through the gender attribute (Figure 2a. to 2c.). These graphs reveal differences in the ranking behavior between boys and girls. While the majority of boys ranked male vloggers as top 5, girls were more reserved in ranking girls as the most popular vlogger (a significant percent of girls ranked male vloggers as the most popular ones, fact shown in the Figure 2c.).

Figure 2 a. Network structure by gender. b. Men's network (boys that chose men influencers). c. Women's network (girls that chose women's influencers).

The next step was checking for the hidden patterns within the total network. The results were obtained by calculating the Modularity function in Gephi, a function that allows the identification of hidden patterns. First, we calculated the Modularity Score, which had a significant value: 0.517. As the value was greater than 0.5, the score indicates that there is a sense of hidden communities within the network and there can be distinguished smaller entities (Blondel, et al. 2008). Starting from respondents' vlogging preferences, the software looked for similarity between their options. Based on this algorithm, the data was classified in 10 classes (Figures 3a. to 3c.) with highly skewed distribution. Nearly 75% of the participants belonged to the first five classes, out of which 40% from the most representative three classes (Figure 3b).

Figure 3 a. The whole network. b. The biggest three communities. c. The smallest communities

The method allows for the identification of each community and also of the relations established between communities. In the figure illustrated below (Figure 4 to 10) we present the structure of the main communities, highlighting the most influential nodes.

Figure 4 Vloggers from abroad **Figure 5** Codrin Bradea's community
Figure 6 Bianca Adam's and Zmenta's community

Figure 7 BRomania's community **Figure 8** Mikey Hash's community
Figure 9 Selly's community **Figure 10** Ionut Rusu's community

The identified communities resulted from students' preferences may be compared to vloggers networks resulted from other data. For example, we used the data for featured and related channels as calculated by youtube algorithms for the channels of the 56 relevant vloggers. Although the data did not totally coincide, it showed some overlaps, especially while analysing vlogs with specific content, such as beauty vlogs (one of the smallest communities for our sample- with pink in both networks). Thus, the vloggers ALINA CEUSAN, CARMEN GREBENIȘAN and IOANA GRAMA are connected between them also in the featured channels obtained network. In the case of the vlog channels that have general content, such as entertainment, the distinguished communities identified from our sample are mixed. Although in our network SELLY and BIANCA ADAM have distinct communities, in this network they belong to the same community. Considering that the aim of this paper was not exploring the network of related channels, we will present only some figures with that network for illustrative purposes. A main characteristic that we noticed at the network of featured channels as compared to vlogging preferences' network is the more compact structure of each community. Based on the

data provided by Social Blade (a commercial tool which provides YouTube channels' statistics based on content and viewers) regarding the content of each channel, we approximated the main content that may characterise the biggest communities (See Figures 11 a. to d.).

The main communities correspond to gaming and entertainment, entertainment and music, educational and informational content. While the former communities have more nodes which are more connected between them, but also to other nodes, the latter is smaller and comprises nodes strongly tied between them, but less connected to other communities. Consequently, we may hypothesize that in most of the cases users who watch vloggers from Levi's community do not watch vloggers from other groups. In the case of subscribers of Selly or other vloggers from Selly's community, it is a higher chance that they also watch vloggers outside the community, which means they watch vloggers that produce other types of content.

We did not deepen our research in this topic, as its statistical value is questionable.

Figure 11 a. Network Featured/related channels b. Codrin's community (gaming and entertainment) c. Selly's community (entertainment, music- mostly for teenagers) d. Levi's community (educational content)

Limitations

The research presents a tool of getting an in-depth view of the online contents, grounding on the assumptions of network science. Still, we have to regard this analysis with caution, as it encompasses severe limitations. The distinguished vlogging communities are not representative for teenagers from Bihor county, as the respondents were selected based on opportunity sampling and their number is too small for a quantitative analysis. Moreover, we did not have access to sociodemographic data about our respondents and we did not have control over the conditions in which children answered the data. This seriously questions the results of the study. When the sample is representative and the sociodemographic data about respondents are available to researchers, the identified communities may be included in a variable. The created dependent variable may be included in logistic regressions or other statistical analysis for categorical data. Nevertheless, researchers should acknowledge the dynamic character of the modularity classes (the procedure is based on a high number of iterations, fact that implies that some small changes may appear from one analysis to another- such as the number of communities). Even in these conditions, the communities identified with Gephi (by using modularity function, gender attribute, etc) have a high potential of explaining specific aspects of a certain group. It may also identify vulnerable groups and patterns of reducing the risky potential by eliminating the elements that pose risks.

Conclusion

The paper shows an example of how to use open-source programs to analyse aspects attributed to a specific group. We hold that online data is to be regarded as a useful tool in social sciences. As Internet is ubiquitous in teenagers' lives, exploring the types of online content they are exposed to and the mechanisms through which a certain content is spread within a network provide useful data both for social scientists and educationalists. Online content gives information about users' preferences, values and behavior, thus constituting a valuable clue in indicating students at educational risk. In addition, understanding the diffusion patterns facilitates a targeted intervention in a system, with significantly higher chances of ameliorating the situations.

Used together with content analysis, this method reveals contents that potentially pose a risk in engaging in deviant behavior and shows the way they spread within a network. It identifies the main hubs (the most connected nodes) and the connections that lead to problematic content.

Knowing the hubs, which represent in this case contents found appealing by most of the users allows a targetted intervention in the system. Networks may easily be destroyed by eliminating the hubs. If we apply this theory to the diffusion of content, it follows that the diffusion of problematic content may be significantly diminished by eliminating the main pathways through which it is transmitted, for example: certain videos, specific vloggers, websites, etc. At the same time, the presented method leads to the identification of the contents that pose a threat in leading to a problematic content. Removing these ties leaves the nodes with problematic content isolated. Consequently, their influence is reduced.

As the offered content is limited, this method should be used in addition to other techniques and methods, such as content analysis (for better distinguishing the type of content), interview, focus group, questionnaire. The obtained data from other methods, which illustrate some of the factors that lie behind the connections enrich its findings. Serving in the identification of hidden communities, the proposed method may be used both for qualitative and quantitative study.

Glossary Of Network Science Terms

Area 5000.0: The area shows the size of a graph in the workspace (in Gephi programme). For the Fruchterman-Reingold Algorithm, the default area is 10000.0. Choosing the 5000.0 area instead of the default one reduced the size of the graph and it allowed for a better visualization.

Average degree: Average number of links for each node.
(<https://sites.google.com/a/umn.edu/social-network-analysis/terminology>)

Average weighted degree: Average of sum of weights of the edges of nodes.
(<https://sites.google.com/a/umn.edu/social-network-analysis/terminology>)

Bipartite graph: a graph in which nodes are two different entities. According to Weisstein (2012), a bipartite graph is a set of graph vertices (nodes) decomposed into two sets such that no two graph vertices within the same set are adjacent.

Bipartite network: A bipartite network is a graph with two different types of nodes where nodes from one category can only connect to nodes from the other category (Filho & Neale, 2018)

Degree: the degree of a node represents the number of nodes to which a node is connected
(http://www.stats.ox.ac.uk/~snijders/D1pm_Graph_theory_basics.pdf). In Gephi, the loops, which are the edges that have the same node as their

starting and end point are counted twice, fact that determines researchers to use for comparisons weighted degree, the average degree and the average weighted degree.

Edge: a connection between two entities. (called “nodes” or “vertices”)

Eigenvector centrality: Measures how important is a node in a network based on its connections. The eigenvector centrality of a specific node represents the sum of its connections to other nodes, weighted by their centrality.

(http://www.stats.ox.ac.uk/~snijders/D1pm_Graph_theory_basics.pdf)

Fruchterman-Reingold Layout: the layout algorithms are very important for visualization, as they influence the way the graph looks like. Fruchterman-Reingold layout was developed by Thomas Fruchterman and Edward Reingold in 1991 and simulates the graph as a system of mass particles in which the nodes are the particles and the edges are springs between the particles. The sum of the force vectors determines the way in which a node moves, while the node width establishes how far a node goes in one step. The nodes stop moving when the graph reaches a state of equilibrium. (<https://gephi.org/users/tutorial-layouts/>)

Graph: A graph is a set of vertices (nodes) and a set of edges, in which an edge joins a pair of vertices (nodes). (Fruchterman & Reingold, 1991)

Graph density: The density of a graph measures how close a network is to be completed. A network is considered completed when all the nodes are connected between them, considering that graph density represents the ratio of the number of edges and the number of possible edges (http://www.stats.ox.ac.uk/~snijders/D1pm_Graph_theory_basics.pdf). A density equal to 1 or close to 1 means that all nodes are connected between them, while a density of 0 means that there are no connections between nodes. A small density shows that most of the nodes are weakly connected in the network.

Gravity 5.0: The gravity shows the amount of force between nodes. As it is explained in Gephi documentation, gravity attracts all nodes to the center in order to avoid dispersion of disconnected components. The default value for Fruchterman-Reingold Layout is 10.0. A smaller value implies a higher dispersion, with nodes going farther from the center. We used this measure for gravity in order to keep the same proportion between Area and gravity as the default one (10000.0 Area with 10.0 gravity).

Hub: node that has a higher number of connections than the average. For example, influencers may represent hubs in a network of social media users.

Network diameter: the diameter represents the longest graph distance between any two nodes in the network

Node: called also vertex, the node is the entity observed in a graph. It is represented as a dot.

Tie: connection within two nodes. There can be distinguished three types of ties: strong ties (very connected entities), weak ties (entities that have less connections), latent ties (entities that are not yet connected, but they have the potential to be connected).

Weight: it is associated to edges, representing a label of it. Weights are real numbers and may measure how strong is a connection (example: 1 –weak; 2-strong; 3 very strong) or may be the result of a ranking system (like in these paper 5- the favorite; 4-the second favorite; 3- preference number 3; 4-preference number 4; 5 – preference number 5).

References

- Rubin, A. M. (2009). "Uses and Gratifications An Evolving Perspective of Media Effects." In M. B. O. Nabi, Robin L. (editors), *The SAGE handbook of media processes and effects*, (pp. 147–60). Los Angeles: Sage.
- Balea, B. (2016). "Digital Natives or Not? How do Romanian Adolescents Cross the Boundaries of Internet Common Use." *Studia UBB Sociologia*, 61(1): 59–76.
- Badge, J. L., Saunders, N. F. W., & Cann, A. J. (2012). *Beyond marks: new tools to visualise student engagement via social networks*. Retrieved from <https://lra.le.ac.uk/bitstream/2381/10086/2/published.pdf>
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi : An Open Source Software for Exploring and Manipulating Networks Visualization and Exploration of Large Graphs.
- Biel, J.-I., & Gatica-Perez, D. (2009). "Voices of Vlogging." Proceedings of the 4th International Conference on Weblogs and Social Media, 211–14.
- Biel, J.-I., & Gatica-Perez, D. (2010). "VlogSense: Conversational Behavior and Social Attention in YouTube." *ACM Trans. Multimedia Comput. Commun. Appl.*, 2(20).
- Biel, J.-I., & Gatica-Perez, D. (n.d.). "Wearing a YouTube Hat: Directors, Comedians, Gurus, and User Aggregated Behavior."
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008(10).
- Bou-Franch, P., Lorenzo-Dus, N., GarcésGarc, P., & Blitvich, G.-C. (2012). "Social Interaction in YouTube Text-Based Polylogues: A Study of Coherence."
- Burgess, J., & Green, J. (2018). *YouTube: Online video and participatory culture*. John Wiley & Sons

- Chiang, H.-S., & Hsiao, K.-L. (2015). YouTube stickiness: the needs, personal, and environmental perspective. *Internet Research*, 25(1), 85-106.
- Coletto, M., Luca, M. A., Lucchese, C., & Silvestri, F. (2017). Adult Content Consumption in Online Social Networks. *Social Network Analysis and Mining*, 7, 28
- Coviello, N. E., & Munro, H. J. (1995). Growing the entrepreneurial firm: networking for international market development. *European journal of marketing*, 29(7), 49-61.
- Danon, L., Diaz-Guilera, A., Duch, J., & Arenas, A. (2005). *Comparing community structure identification*. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09), P09008. Retrieved from: <https://arxiv.org/pdf/cond-mat/0505245.pdf>
- van Deursen, A. J. A. M., & van Dijk, J. A. G. M. (2014). The Digital Divide Shifts to Differences in Usage. *New Media and Society*, 16(3): 507–26.
- Van Dijk, J., & Hacker, K. (2003). The Digital Divide as a Complex and Dynamic Phenomenon. The Multifaceted Concept of Access. *The Information Society* 19, 315–26.
- DiMaggio, P., & Garip, F. (2012). Network Effects and Social Inequality. *Annual Review of Sociology*, 38(1), 93–118.
- Ferchaud, A., Grzeslo, J., Orme, S., & LaGroue, J. (2018). Parasocial Attributes and YouTube Personalities: Exploring Content Trends across the Most Subscribed YouTube Channels. *Computers in Human Behavior*, 80, 88–96.
- Figueiredo, F., Benevenuto, F., & Almeida, J. M. (2011). The Tube over Time: Characterizing Popularity Growth of YouTube Videos. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining* 745–54.
- Gerbner, G., & Gross, L. (1976). Living with television: The violence profile. *Journal of communication*, 26(2), 172-199.
- Gerbner, G. (1998). Cultivation analysis: An overview. *Mass communication and society*, 1(3-4), 175-194.
- Grunspan, D. Z., Wiggins, B. L., & Goodreau, S. M. (2014). Understanding classrooms through social network analysis: A primer for social network analysis in education research. *CBE—Life Sciences Education*, 13(2), 167-178.
- Hargittai, E. (2010). Digital Na(t)ives? Variation in Internet Skills and Uses among Members of the 'Net Generation. *Sociological Inquiry*, 80(1), 92–113.
- Hargittai, E., Bakalov, R., Hunter, C.D., Van Rees, K., Steensland, B., Tepper, S., Upright, C., & Wuthnow, B. (2000). Open Portals or Closed Gates? Channeling Content on the World Wide Web S. *Poetics*, 27(27), 233–53.
- Hargittai, E., & Hinnant, A. (2008). Digital Inequality: Differences in Young Adults' Use of the Internet. *Communication Research*, 35(5), 602–21.

- Haridakis, P. M., & Hanson, G. (2009). Social Interaction and Co-Viewing With YouTube: Blending Mass Communication Reception and Social Connection. *Journal of Broadcasting & Electronic Media*, 317-335. doi:10.1080/08838150902908270
- Khan, M. L. (2017). Social media engagement: What motivates user participation and consumption on YouTube? *Computers in Human Behavior*, 66, 236-247.
- Diaconescu, M., Barbovski, M., Baci, C. (2008). *Beneficii si Riscuri Ale Utilizării Internetului in Rândul Copiilor si Adolescentilor. Repere Pentru Elaborarea unui Ghid de Siguranță Pe Internet si de Prevenire a Victimizării Online*. Cluj-Napoca: Presa Universitară Clujeană.
- Mascheroni, G., & Ólafsson, K. (2014). *Risks and Opportunities*. Second Edition.
- Mascheroni, G., Ólafsson, K., Cuman, A., Dinh, T., Haddon, L., Jørgensen, H., Livingstone, S., Neill, B. O., Ponte, C., Stald, G., Velicu, A., & Vincent, J. (2013). *Mobile Internet Access and Use among European Children. Initial Findings of the Net Children Go Mobile Project*. Milano.
- Ragnedda, M., & Muschert, G. W. (eds.) (2013). *The Digital Divide. The Internet and Social Inequality in International Perspective*. New York: Routledge.
- Mocanu, F. (2018). Digital Literacy Among Young Adults in Romania. *Management Dynamics in the Knowledge Economy*, 6(3), 449-469.
- Mogallapu, A. (n.d.) "Scholars' Mine Social Network Analysis of the Video Bloggers' Community in YouTube."
- Molyneaux, H., Gibson, K., O'Donnell, S., & Singer, J. (2007). "New Visual Media and Gender: A Content, Visual and Audience Analysis of YouTube Vlogs."
- Molyneaux, H., O'Donnell, S., & Gibson, A. (2008). Exploring the Gender Divide on YouTube: An Analysis of the Creation and Reception of Vlogs. *American Communication Journal*, 10(2).
- Morgan, M., Shanahan, J., & Signorielli, N. (2017). Cultivation Theory: Idea, Topical Fields, and Methodology. *The International Encyclopedia of Media Effects*, 1-14.
- Morgan, M., Shanahan, J., Signorielli, N., Morgan, M., & Shanahan, J. (2014). Cultivation Theory in the Twenty-First Century. In R. S. Fortner and P. M. Fackler (eds.), *The Handbook of Media and Mass Communication Theory* (pp. 480-97).
- Nelson, D. W. (2016). Millennial Social Networking Behavior from a Uses and Gratifications Perspective. 76(11-B(E)).
- Newman, M. E. J. (2016). *Community detection in networks: Modularity optimization and maximum likelihood are equivalent*. arXiv preprint arXiv:1606.02319."
- Paluck, E. L., Shepherd, H., & Aronow, P. M. (2016). Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, 113(3), 566-571.

- Ragnedda, M., & Kreitem, H. (2018). The three levels of digital divide in East EU countries. *World of Media. Journal of Russian Media and Journalism Studies*, 1(4), 5-26.
- Ruggiero, T. (2000). Uses and Gratifications Theory in the 21st Century. *Mass Communication & Mass Communication & Society*, 3(1), 3-37.
- Salvați Copiii (2019). *Studiu privind Utilizarea Internetului De Către Copiii. Cercetare Socială de tip cantitativ*. București.
- Shao, G. (2009). Understanding the Appeal of User-generated Media: A Uses and Gratification Perspective. *Internet Research*, 19(1), 7-25.
- Smith, W. R. (1956). Product Differentiation and Market Segmentation As Alternative Marketing Strategies. *Journal of Marketing*, 21(1), 3-8.
- So, J. (2012). Uses, Gratifications, and Beyond: Toward a Model of Motivated Media Exposure and Its Effects on Risk Perception. *Communication Theory*, 22(2), 116-37.
- Stafford, T. F., Stafford, M. R., & Schkade, L. L. (2004). Determining uses and gratifications for the Internet. *Decision sciences*, 35(2): 259-288.
- Ștefănescu, S. (2008). Utilizarea Internetului: Pattern-Uri De Consum Ale Adolescenților Din România. *Revista Română de Sociologie*, XIX(3-4), 307-30.
- Stefanone, M. A., Lackaff, D., & Rosen, D. (2010). The Relationship between Traditional Mass Media and "Social Media": Reality Television as a Model for Social Network Site Behavior. *Journal of Broadcasting & Electronic Media*, 54(3), 508-25.
- Sundar, S. S., & Limperos, A. M. (2013). Uses and grats 2.0: New gratifications for new media. *Journal of Broadcasting & Electronic Media*, 57(4), 504-525.
- Tufă, L. (2010). Diviziunea digitală. Accesul și utilizarea Internetului în România, comparativ cu țările Uniunii Europene. *Calitatea Vieții*, 21(1-2), 71-86.
- Velicu, A., Balea, B., & Barbovski, M. (2019). Acces, utilizări, riscuri și oportunități ale internetului pentru copii din România. Rezultatele EU Kids Online 2018. EU Kids Online și DigiLiv-REI.
- Wedel, M., & Kamakura, W. A. (2012). *Market segmentation: Conceptual and methodological foundations* (Vol. 8). Springer Science & Business Media.
- Whiting, A., & Williams, D. (2013). Why People Use Social Media: A Uses and Gratifications Approach. *Qualitative Market Research: An International Journal*, 16(4), 362-69.
- Yoganarasimhan, H. (2011). Impact of Social Network Structure on Content Propagation: A Study Using YouTube Data. *Springer Science. Business Media, LLC 2011*, 111-50.