USER-GENERATED HATE SPEECH: ANALYSIS, LESSONS LEARNT, AND POLICY IMPLICATIONS. THE CASE OF ROMANIA

By

Istvan-Peter Ianto-Petnehazi

Submitted to

Central European University

Department of Political Science

In partial fulfillment of the requirements for the degree of

Master of Political Science

Supervisors: Kate Coyer (CEU)

Stefania Milan (University of Toronto)

Budapest, Hungary

(2012)

Abstract

This thesis is a descriptive case study about use of interactive features of online newspapers in Romania by the members of the audience i.e. 'users' to propagate hate speech: a phenomena labeled user generated hate-speech. To assess the proportions of the phenomena and to test the efficiency of the Romanian legislation and of the site usage policies in identifying and preventing user-generated hate speech a comparative analysis of the participatory features of five major Romanian news sites was performed, which served as basis for the collection of a purposive sample of 84 articles and the respective 6031 comments. The articles were grouped on target minorities and topics that occurred during a period of 13 months from March 2011 to April 2012. A definition of 'hate' was created based on the legislation and the encyclopedic definitions, and expanded into 23 hate-type categories, to provide a codebook for content analysis, which revealed that 37.99 percent of comments in the sample contained hate speech.

Acknowledgements

I express my gratitude to my supervisors Professors Kate Coyer and Stefania Milan, for their guidance, support and patience through the writing of this thesis and my two years at CEU. I dedicate this thesis to my partner Anna without whose love, support and programming skills this thesis and my stay at CEU would have not been possible.

Table of Contents

List of tables and figures

Tables

Figures

Introduction

"First we will go to the streets without weapons. Then we will see" said the headline on the home page of Gandul.info one of the most visited Romanian online newspapers on 16th June 2011, referring to the determination of the leaders of the country's 1.2 million Hungarian minority to stop the proposed territorial reorganization of the country which would dissolve their two counties into a region with Romanian majority.[1] Hours within the publication the article prompted 340 reader comments 70 of which called for the extermination of the Hungarian minority, the murder of their leaders or the rape of Hungarian women. As it turned out on the same day the headline distorted the words of the Hungarian leader, who actually said "peacefully". Almost a year later the number of comments to the article grew to more than 600 and the calls for genocide are still there.

This study is an exploration of interactivity, the most important feature of the transition of newspapers from print to the internet, and its media policy implications. I will exemplify the variety of issues and the difficulties posed to regulators focusing on the audience participation features of online news sites, particularly comments to articles, which opened up access to mass audiences for everyone.[2] The analysis of these online spaces incorporate in one place some of the important questions of the web 2.0 era, as for instance the increased difficulty of differentiating between public and private forums and opinions[3]; the tension between control over content and freedom of speech[4]; the blurring distinction between audiences and

[1] gandul.info. 2011. "Tamas Sandor (DAHR) the Chief of the County Council of Covasna About the Civil Disobedience: 'In the First Phase We Will Get to the Streets Without Weapons. Than We Will See' (Tamas Sandor (UDMR), Şeful Consiliului Judeţean Covasna, Despre „nesupunerea Civică": „În Prima Fază, Ieşim În Stradă Fără Arme. Apoi o Să Vedem".) EXCLUSIV - Gandul." http://www.gandul.info/politica/tamas-sandor-udmr-seful-consiliului-judetean-covasna-despre-nesupunerea-civica-in-prima-faza-iesim-in-strada-fara-arme-apoi-o-sa-vedem-exclusiv-8342275. Later it was revealed that the journalist mistranslated the Hungarian word *bekesen (peacefully)* giving it the sense without weapons

[2] Kaufer, 2004

[3] Braman, 2006

[4] Barendt, 2007; Cammaert, 2009

publishers[5]. In this thesis I will present how the weaknesses of unclear and sometimes obsolete regulations, ineffective authorities and media policy, coupled with facile access to mass audiences provided by the interactive features of websites can be exploited by users to target vulnerable groups with hateful, discriminatory content. I have termed this: user-generated hate speech.[6]

My definition of user generated hate speech includes elements from both the concept of user generated content and hate speech. I define user generated hate speech as *content (text, audio, video, multimedia), usually created by non-professional, and anonymous users, aimed at intimidating/harming particular minority groups (ethnic, sexual, racial) taking advantage of interactive features of websites aimed at the general public or content hosting platforms for being published and to reach its targets.*

Historically compared to broadcast media, the printed press enjoyed significantly larger liberties as for instance the lack of regulations regarding licensing or content[7], which at least in Europe are both under quite heavy state supervision with dedicated state authorities/supervisory bodies[8]. Traditionally the main argument for lesser regulation of the printed press was that being an on-demand medium, i.e. one has to actively seek (buy) a newspaper, while other mediums were more intrusive. With the transition of newspapers to the internet legislators faced the problem of placing the website into an adequate media policy category. Is it the same as the print edition and therefore should be subjected to the same

[5] Benkler, 2006; Schafer, 2011; Valcke and Lenaerts, 2010

6 the term user-generated hate speech is mentioned in Brown-Sica, Margaret, and Jeffrey Beall. "Library 2.0 and the Problem of Hate Speech." Electronic Journal of Academic and Special Librarianship v.9 no.2, no. Summer 2008 (2008). http://southernlibrarianship.icaap.org/content/v09n02/brown-sica_m01.html. But the authors do not provide a definition. A search on Sage online journals and EBSCO host complete did not return results for the term. To my best knowledge this is the first research paper that uses and defines the term.

[7] Braman, Sandra. 2006. *Change of state*. Cambridge (Mass.)[illegible];;London: the MIT press. p.68

[8] An extensive overview of different regulatory toolkits and bodies can be found at K.U.Leuven – ICRI (lead contractor) Jönköping International Business School - MMTC Central European University - CMCS Ernst & Young Consultancy Belgium. Country reports - Study on Indicators for Media Pluralism - Media Task Force | Europa - Information Society and Media. *Independent Study on Indicators for Media Pluralism in the Member States - Towards a Risk-Based Approach.* http://ec.europa.eu/information_society/media_taskforce/pluralism/study/country_rep/index_en.htm.

"relaxed" rules applying to the press, or is it a totally different product and thus new rules are needed? Interactivity further complicated the problem. This is especially evident in the case of user comments, which are difficult to fit into one of the traditional categories of media policy and regulation, audience or editors/journalists. They share the same journalistic space, and potentially the audience of regular articles, but contrary to professional journalists their authors are usually anonymous thus unaccountable and face no consequences for their actions, even if they might fall under legal restrictions as it is the case of discriminatory content.

While it can be argued that by taking the decision to access a certain website, the reader made a conscious decision and thus assumes the risk of facing whatever content is displayed there, usually websites do not warn their readers that they might be also hosting harmful content originating from their users.[9] Such content is usually displayed in the same journalistic space (same page) as the professional text and legitimate user contributions, thus exposing all visitors to harmful content; raising the question whether inadequate participation policies could open up the possibility for the website to be exploited by the users as a delivery platform of readers to hate speech.

On the other hand there are a series of questions for which this thesis cannot offer an answer. It is not my intention to get to the social, economical cultural roots, causes of hate speech nor is to offer a solution that would solve the problem. What this thesis aims to do is analyze and describe the problem of user generated hate speech in Romania and to signal a policy gap, by presenting how unclear regulation and legislation made obsolete by new technical developments can be exploited to deliver discriminatory content specifically to members of the targeted group where it could inflict the most harm.

[9] My analysis of the website terms and conditions revealed that the terms of use for one of the websites in the sample (evz.ro) does contain a warning about potentially harmful content and a disclaimer for any harm caused to readers.

This thesis does not advocate for, nor does it endorse internet censorship, rather it arguments for clearer rules for user participation, and the separation of professional and UGC. If users are to be considered co-authors and the content created by them is an essential part of the media product, as many of the media scholars cited in the next chapters suggest[10], I argue that just as in the traditional press model the newspaper should assume editorial responsibility for them and implement moderating policies that would prevent the access to mass audiences for such content.

First, I will describe the phenomena from legal and media theory perspective presenting different, even conflicting approaches to hate speech regulation, freedom of expression the roles of users in the new online environment, as well as some of the challenges faced by policymakers.

The methodology is a triangulation of content analysis of a sample of comments from the websites of almost all national daily newspapers in Romania, a comparative analysis of the role their audience participation policies play in the existence of user generated hate speech and a review of the Romanian media regulation and anti-discrimination legislation. In spite of a range of anti-discrimination laws that are transposed in the participation policies of the websites, the presence of hate speech is widespread. Using a codebook based on the existing legislation, the sites terms and conditions, the encyclopedic definition of hate speech I l performed content analysis on a purposive or relevance sample[11] of user comments collected from the websites of the four major Romanian newspapers and one news portal site, to assess the extent of hate speech and the analyze the effectiveness of their policies in preventing the abuse of their interactive features to disseminate discriminative content. Resulting in a

[10] Benkler (2006), Deuze (2008), Schafer (2011)
[11] Krippendorf, 2004:113; Ritchie and Lewis, 2008: 78

description of the user generated hate speech phenomena in Romania, the legal and regulatory framework that contributed to its proliferation.

Chapter I. Methodology

This thesis is a descriptive case study about use of interactive features of online newspapers in Romania by the members of the audience i.e. 'users' to propagate hate speech: a phenomena I have labeled user generated hate-speech. The aim of this research is to show how the transformation of newspapers from printed unidirectional products into an online interactive platform[12] resulted in loopholes in media policy that contribute to a wider spread of such content.

According to Stake there are three types of case studies: intrinsic, instrumental, and collective[13]. Intrinsic case studies consist of research undertaken in order to get a better understanding of a particular case, because the case itself has some particular features worth exploring and not aimed at theory building although sometimes it can result in that. On the other hand when performing instrumental case study the case itself serves only facilitate the understanding of a broader phenomena. My research has features from both types of case studies. It is an intrinsic study in the sense that the Romanian media system and its components presented here (the press) have characteristics that warrant a detailed examination such as its evolution from a state control to its present form of deregulated printed and online hybrid[14]; the existence of the Hungarian community, one of the largest minority in Europe and its representation in the media or the widespread intolerance towards ethnic/religious/racial/sexual minorities[15]. On the other hand it is also an instrumental case

[12] Deuze, Mark. 2003. "The Web and its Journalisms: Considering the Consequences of Different Types of Newsmedia Online." *New Media & Society* 5 (2) (June): 203-230. doi:10.1177/1461444803005002004.
[13] Stake, Robert. 2005. Qualitative Case Studies. In *Sage Handbook of Qualitative Research*, 443, 467. 3rd ed. Sage Publications.
[14] For a detailed overview of the development of the Romanian media system from the fall of the communism see. Gross, Peter. 1996. *Mass media in revolution and national development: the Romanian laboratory*. Ames (Iowa): Iowa state university press. - and also Gross, Peter, and Mihai Coman. 2006. *Media and journalism in Romania*. Berlin: Vistas.
[15] INSOMAR. 2009. *Fenomenul discriminarii in Romania - perceptii si atitudini" in anul 2009 - Discrimination in Romania-perceptions and attitudes in 2009*. CNCD - National Anti-Discrimination Council, Romania. http://www.cncd.org.ro/files/file/Fenomenul%20discriminarii%202009.pdf.

study as it aims to contribute to a better understanding of the challenges faced by regulators and journalists alike due to the transformations of journalism caused by the move to the internet, and the more recent developments towards interactivity, such as the question of user anonymity or responsibility for user generated content as it will be presented later. In this sense the case of hate speech serves as a particularly suitable illustration of such challenges and the shortcomings of media policy inadequately adapted to the internet.

The research is based on a year long observation of a sample of five Romanian news sites started in March 2011 with the intention to reveal both the way the selected media organizations handle user participation, and the results of these policies as they appear through the comments sections. The primary research method is content analysis of a sample of articles, comments and site usage policies resulted from the observation in order to answer the following research questions.

I.1. Research questions

The larger question this thesis seeks to better understand is that of responsibility and to reveal the policy approaches that lead to the presence of user-generated hate speech. In order to address this, the thesis focuses on two research questions, one present what exactly constitutes user generated hate speech in practice, and second to see the factors that contribute to its propagation.

RQ1: What is the nature of user generated hate speech?

RQ2: Do the legislative, regulatory environment, the editorial or moderating policies, the type of the media product, the user participation rules of the media organizations contribute to the presence of user generated hate speech?

I.2. Case selection and sampling

The research questions were answered using a dual sample approach that resulted in two connected samples of Romanian online newspapers and of the respective user comments.

In order to study the role of user participation guidelines and approaches to user participation in the existence of user generated hate speech a sample of Romanian news-sites was assembled here the units of analysis were the media organizations/sites. Data collected for this sample includes terms of use/terms of service (TOS) of websites (especially regarding user participation) the comments posting interface, the placement of comments in relation to the content produced by the media organization, the user registration requirements and apparent moderation techniques visible by visiting the site or when posting comments. These allowed a complete overview and comparison of user participation policies of online newspaper segment of the Romanian media system. This sample can almost be considered a census as it includes all national newspapers in Romania with relevant user participation and circulation.

The newspapers and websites were selected based on information from the database of the Romanian Bureau of Circulation Audit (BRAT)[16] shown in Table 1. BRAT for the period of the study March 2011 - April 2012 listed 7 national daily newspapers (*cotidian generalist national*) composing the so called "quality segment of the national press" excluding tabloids. Five of these newspapers were included in the original sample, the other two *Puterea* (The power) and *Curierul* National (The National Courier) were excluded due to very low circulation numbers (around 3000 compared to around 9000 of the lowest in the sample Gandul). Jurnalul (The Journal) (www.jurnalul.ro) was excluded from the sample after a couple months of observation as it became clear that although it had similar terms and conditions as the other sites the level of user participation was low, during the 13 month

[16]BRAT - Romanian Bureau of Circulation Audit. Circulation number for nationwide daily newspapers (cotidian generalist national) for the period march 2011 - march 2012. http://www.brat.ro/index.php?page=compare.

observation had very few articles that would meet the criteria of having at least 20 comments per article to be included into the sample of comments.

According to data from the internet audience study (SATI) of BRAT presented in Table 1 the websites in the sample are amongst the most visited Romanian websites in terms of unique visitors in the category of news-sites, also ranking high amongst audited Romanian websites in general. Other top ranking news sites belong mostly to televisions and were not included in the sample as they the aim of this research was to study the transition of newspapers to online environment, therefore only media organizations/sites that had a printed edition where included. Exceptions were made for gandul.info which at the beginning of the study still had a printed version that was discontinued in march 2011. The newspaper was also kept in the sample as it serves as an illustration of the transition of the press from a print through a hybrid online/print to an online only medium.

	Adevarul		Evz		Gandul		RimaniaLibera		Hotnews
Period	**Print**	**Online**	**Print**	**Online**	**Print**	**Online**	**Print**	**Online**	**Online**
Mar 2012	n/a	2273464	n/a	1448594	n/a	2166162	n/a	717093	1310320
Feb 2012	n/a	2435230	n/a	1530202	n/a	2293050	n/a	731550	1459331
Jan 2012	n/a	2670791	n/a	1606177	n/a	2381772	n/a	775395	1555655
Dec 2011	29102	2417629	16174	1464764	n/a	2082459	36707	626572	1332912
Nov 2011	27222	2504943	15507	1474793	n/a	2090535	37786	692299	1328460
Oct 2011	29764	2487269	15658	1490293	n/a	2118880	38255	727147	1383016
Sep 2011	32937	2278284	15556	1332726	n/a	1851449	39205	665884	1210038
Aug 2011	35899	1843188	16351	1196650	n/a	1671246	39454	592135	1117038
Jul 2011	42849	2167916	15634	1476645	n/a	1815485	39748	702302	1340255
Jun 2011	43415	1956475	16336	1288791	n/a	1752421	40602	624048	1353130
May 2011	43946	1912514	16271	1168603	n/a	1645980	41366	660472	1297142
Apr 2011	45109	1695829	16751	1082108	n/a	1529326	42276	619483	1219350
Mar 2011	45685	1992992	17965	1274258	10333	1625373	41809	775532	1465323

Table 1. Circulation numbers and unique visitors for the websites in the sample. Source: BRAT/SATI

The other exception Hotnews.ro is a natively digital media organization, without a print edition but it was included in the sample due to its strong connection to traditional (i.e. print) newspapers. The site is labeled in Romanian as a news portal and it started out originally as a news aggregator offering a sample of articles/content from other newspapers under the name revistrapresei.ro (the review of the press) but later started adding their own original content and changed the name. However the site still has some of the features of an aggregator as some of the content is collected from the sites of other newspapers, but on the other hand the majority of their content is original production.

An additional step in analyzing the regulatory environment was also to identify legislation relevant to media and hate speech that was also used in the creation of the codebook for the content analysis of user comments.

Comment sample for content analysis

For the second stage of research I performed content analysis on a sample of 6081 comments to 83 articles regarding minorities from the five websites, in order to give an answer to the first research question (RQ1). The study design falls within the category of 'problem driven content analysis' described by Krippendorf as studies where the choices of 'suitable texts' and 'analytical paths' are shaped their potential to answer the research question.[17] Therefore a purposive or relevance sample was assembled as described by Krippendorf by choosing the texts based on their relevance for the research questions in order to give them "a chance of being answered correctly"[18]. Since it is not probabilistic sample, it is not be representative for the population of texts published in the Romanian press. However, I believe that it will be a

[17] Krippendorff, Klaus. 2004. *Content analysis : an introduction to its methodology*. 2nd ed. Thousand Oaks Calif.: Sage,.p. 340
[18] Idem, p. 113

good illustration for the "population of relevant texts".

The main feature of purposive sampling is according to Ritchie et al. that "units have to meet certain criteria to be included" in the sample, i.e. they are 'deliberately' selected because they have particular features or characteristics.[19] In the case of the present research the first selection criteria was to include articles with topics regarding minorities that are of interest to the members of the minorities, and are likely to generate debate. To ensure that the sample includes articles that generated sufficient interest in the form of debate and audience a second criteria was introduced in selecting only articles with at least 20 comments and 500 views. Selective judgment had to be involved due to the first criteria of relevance for minorities as no random sampling method would have resulted in a sample that would provide accurate illustration to the definition presented earlier or an answer to the research question.

Ritchie, Lewis and Elam also caution about the level of researchers deliberation involved in purposive samples and point out the need to provide equal opportunities for the hypotheses to be confirmed or disproved.[20] I believe this requirement is met by my sample as my judgment was only involved on the selection of topics and articles, while the primary units of analysis the comments were preserved as they were on the websites and not altered in any way. Furthermore the list of topics presented in the following section is the result of a thirteen month observation of the five websites, while also providing a large diversity further contributing to meeting this requirement.

[19] Ritchie, Jane, and Jane Lewis. 2003. *Qualitative research practice□: a guide for social science students and researchers*. London; Thousand Oaks, Calif.: Sage Publications. p 78

[20] Ritchie, Jane, and Jane Lewis. 2003. *Qualitative research practice□: a guide for social science students and researchers*. London; Thousand Oaks, Calif.: Sage Publications p. 80

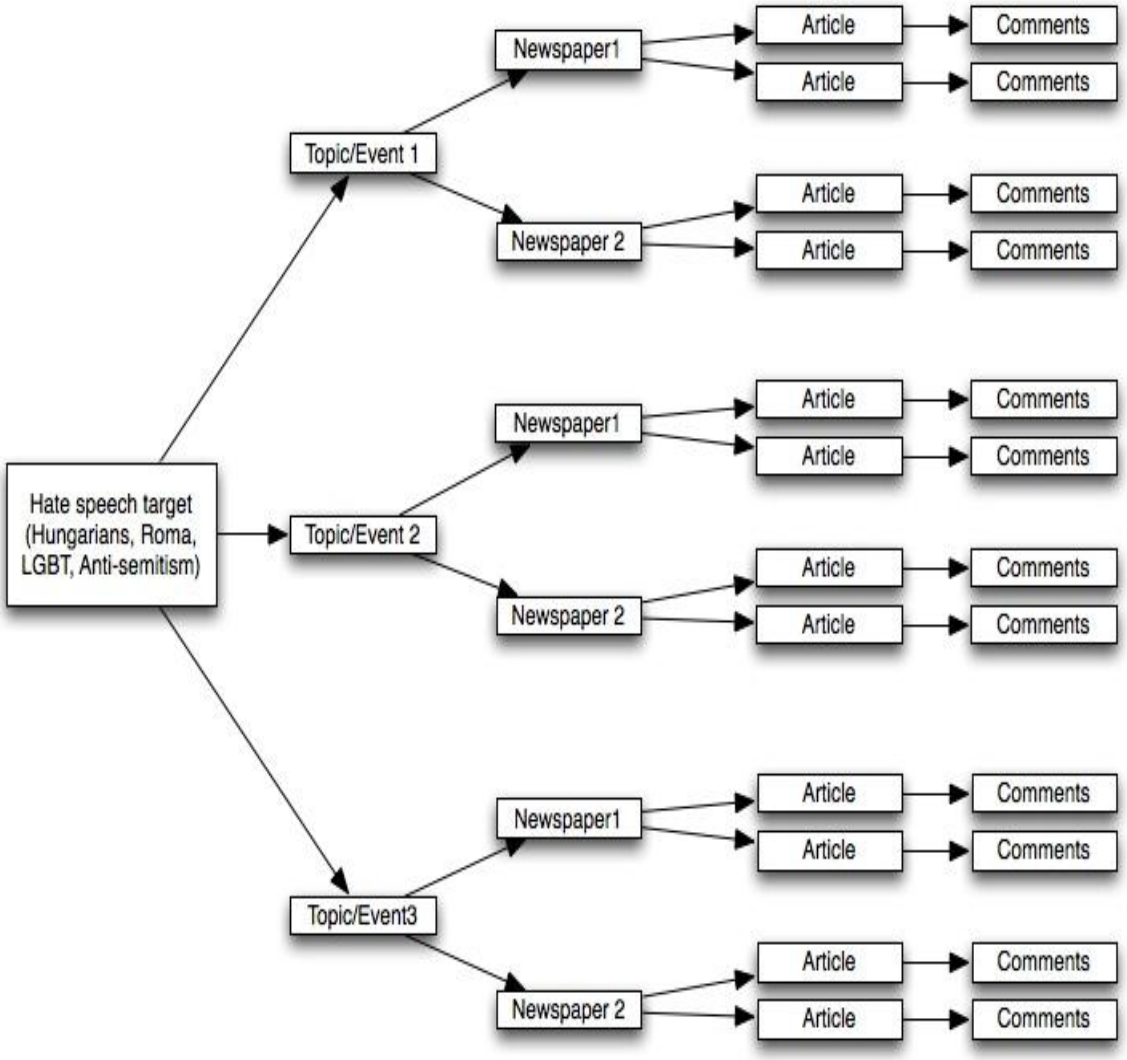


Figure 1: Sample/Database structure

I.3. Research strategy

Figure 1 presents the schematic structure of the sample for the content analysis. As mentioned earlier the units of analysis are individual comments, on the five websites chosen based on circulation data from the Romanian Bureau of Press Circulation audit (BRAT). During the 13 month long observation of the five sites I have collected, 6081 comments to 75 articles in 16 main topics (shown in Appendix 1) regarding minorities that occurred in the Romanian media, and further grouped them into four hate speech target groups: Hungarians, Roma, Homosexuality (LGTB), Anti-semitism/Holocaust (Jewish). Appendix 1 also lists the number of comments in the sample for each issue.

The articles and comments were captured (archived) using the free internet browser extension Zotero that allows the creation of an identical snapshot of the webpages capturing every element.[21] Enlisting the help of a professional programmer, a dedicated software tool was built to extract information from the archived webpages (parse the html files), and arrange it into a database, according to criteria such as, topic, user, article etc.[22] Purpose built software was needed for analyzing my sample, as comments are more than just text in the traditional sense of the word. Beyond the text itself they also contain important database information such as the time when the comments were published, comments published by a given user or preserving the link between comments, the audience votes received by a comment or the characteristics of the debate/dialogue on the website. These data also form the base of my analysis and are important to consider when analyzing an interactive platform such as online newspapers.

[21] www.zotero.org
[22] Online access to the database is available on request by email to janto.petnehazy.istvan@gmail.com

Coding

As mentioned earlier in RQ2 the aim of this research is to signal loopholes in media policy that allowed the existence and propagation of user generated hate speech. In order to objectively identify (code) comments as user generated hate speech a codebook was created based on the existing Romanian legislation[23], the TOS of the five sites, the encyclopedia definition of hate speech[24], and the observation of dominant themes in the comments.[25] According to the literature[26] to ensure reliability and objectivity coding should be done by at least two independent, well trained coders who have not taken part in the development of the codebook. On the other hand authors such as Saldana[27] and Richie and Lewis[28] consider that an individual researcher can also carry out coding. To ensure objectivity two independent coders performed a test of the codebook on a randomly generated sub sample of four articles and the respective 330 comments resulting in Cohen's K for the main categories of hate/non-hate of .72 for tester 1 and the author, and .73 for tester 2 and the author, a satisfying coefficient.[29]

I.5. Social context. Romania and its minorities

Appendix 2 provides an illustration of the various topics regarding minorities that occurred in the Romanian press during the observation period. The largest proportion of the sample consist of topics/articles and comments about the Hungarian minority, this is due partially to historical and to political factors. The 1,2 million Hungarians, (6.5 percent of the population

[23] Government Ordinance (Romania) nr.137/ 31August 2000 (republished)
[24] Kinney, Terry A. 2008. Hate Speech and Ethnophaulisms. In Donsbach, Wolfgang. ed, 2008. *The international encyclopedia of communication*. Malden MA: Blackwell Pub.
[25] The codebook is presented in detail in section IV.3, while the codebook can be found in Appendix 1.
[26] Krippendorf (2004), Neundorf (2002), Berg (2001)
[27] Saldaña, Johnny. 2009. *The coding manual for qualitative researchers*. London; Thousand Oaks, Calif.: Sage.
[28] Ritchie, Jane, and Jane Lewis. 2003. *Qualitative research practice□: a guide for social science students and researchers*. London; Thousand Oaks, Calif.: Sage Publications
[29] The number of comments for this subsample was maximized at 100 per article the rest being deleted after the sample was generated.

according to the census of 2011) are the largest minority in Romania. They live mostly in the region of Transylvania, that became part of Romania after the treaty of Trianon, that ended the World War I. in 1920, and are frequently blamed with separatist tendencies. On the other hand the Hungarian community also has considerable political power, due to the presence in the parliament and the Government (from 2002 until April, 27-th 2012) of the Democratic Alliance of Hungarians in Romania (DAHR) an ethnic party that consistently gathers around 7 percent of the votes in the national elections. Due to DAHR's intense political activity Hungarian politicians and the DAHR are frequently presented (i.e. on the daily basis) in the Romanian media, but not always in ethnic terms.

Romania is suitable case for studying hate speech due to the widespread negative attitudes towards minorities. The Roma minority is regularly blamed for the bad image of the country abroad, and linked to criminality, once anti-Roma discourse reaching as far as the president, who in 2007 was recorded on tape calling a female journalist "filthy gipsy"[30]. Furthermore despite the extremely small number of Jewish people still living in the country there is anti Semitism, and as the results from the polls cited on the next page show homophobia is widespread, homosexuality being decriminalized only in 2001 at the pressures of the European Union.

According to a survey from 2009 on discrimination made at the request of the National Council Combating Discrimination (CNCD)[31] Hungarians are still regarded with suspicion as 33,9 percent believe that they have different agenda than the rest of the citizens; 61,6 percent

[30] for an overview of the case in English and the legal actions taken against the President by the CNCD see: FRA - European Union Agency for Fundamental Rights. "Romania / Traian Basescu V. CNCD, Dosar Nr. 4510/2/2007, Curtea De Apel Bucuresti, Sentinta Civila Nr.2799." http://infoportal.fra.europa.eu/InfoPortal/caselawFrontEndAccess.do?id=165.

[31] INSOMAR. *Fenomenul discriminarii in Romania - perceptii si atitudini" in anul 2009 - Discrimination in Romania-perceptions and attitudes in 2009*. CNCD - National Anti-Discrimination Council, Romania, 2009. http://www.cncd.org.ro/files/file/Fenomenul%20discriminarii%202009.pdf.
This is the so called Hungarian card, frequently used by politicians of all party orientations and widely accepted in the Romanian population, implying that by asking for educational, cultural, linguistic, or collective minority rights the Hungarians seek to disintegrate Romania, and the return of Transylvania to Hungary. For more on the nationalist discourse regarding Hungarians in Romania see Brubaker (2008)

of those questioned believe that Hungarians should address all public services only in Romanian, 35,9 percent disagree with the existence of Hungarian schools and 43,2 percent with higher education in Hungarian. On the other hand according to the survey Hungarians are considered hard working honest people and accepted as co-workers, neighbors and even family members. With respect to anti-Roma views the report found that, a large majority of those questioned would not even accept Roma or homosexual people as neighbors; 74 percent believe that most Roma break the law, and 48 percent said that they are a disgrace for Romania. Regarding views about homosexuality more than 50 percent consider that it should be treated as a mental disease 22 percent confessing of feeling repulsion hearing the word homosexual, furthermore an alarming 10,3 percent believe that women are less intelligent than men. There are several subcategories in the codebook of the content analysis that were built in order to reflect these attitudes[32]; for instance 'denying rights' refers to the view mentioned earlier that Hungarians access to public services and higher education, while the code 'conspiracy/foreign interest threat' refers also to the belief that Hungarians or other minorities "have a different agenda".

According to the latest 2012 survey of the CNCD regarding the discrimination[33] Romanian citizens have a distorted sense about the meaning of discrimination as 12 percent do not consider it to be discrimination if a person is fired from his job for being homosexual, 11 percent for being pregnant and 12 percent if the access of Roma persons is denied into a public place. Moreover 27 percent do not consider a student to be discriminated if his request of exemption from religion classes is denied in a school that has no classes on his religion[34]. The attitudes towards the main minority groups are similar to those found in 2009; 43 percent of respondents having bad or very bad opinion of the Roma, 16 percent of the Hungarian

[32] see Appendix 1.II
[33] TNS CSOP, Romania. 2012. *Perceptions and Attitudes Regarding Discrimination in Romania (Perceptii Si Atitudini Privind Discriminarea in Romania).* Survey. Romania: CNCD, Romania. http://www.cncd.org.ro/files/file/Raport%20de%20cercetare%20CNCD_Discriminare.pdf.
[34] taking religion classes is mandatory in the primary and optional in the secondary schools

minority, while 36 percent had a good or very good opinion of Hungarians and only 14 percent of Roma. For all minority communities there was a high proportion of about 40 percent who reported having "neither good, nor bad opinion". Regarding stereotypes in a multiple-choice question, 46% considered the Roma to be lazy, 45% aggressive, 35% dishonest. Hungarians were viewed as intolerant by 30%, aggressive by 14%. Surprisingly there was a high proportion of 38% who did not respond to the question regarding negative stereotypes about Hungarians.

Regarding the role of the media in discrimination, according to the aforementioned survey 76% of the respondents noticed discrimination based on ethnicity, and 47% based on sexual orientation on television or the press. As for responsibility 48% considered that ordinary people, 46% politicians and 45% journalists were to blame for the discrimination.

In the questions regarding social distance to minorities, 24 percent of the respondents considered "living in Romania" the closest acceptable relationship to homosexuals, 9 percent "to visit Romania" while 16 percent consider that they should not come to Romania at all. When referring to Roma 22 percent accepts them living in the country, 6 percent as visitors and 6 percent considers that they should not be in Romania at all. Hungarians enjoy a greater acceptance, more than half accepting them even as friends (33%) and family (25%) and the proportion of those refusing their presence in Romania is 4 percents while those who would only like to see them as tourists measure to 5 percent. This high level of people refusing even the presence of minorities in the country lead me to introduce into the codebook[35] for the content analysis the subcategory "Exclusion/This is our country which resulted in similar proportions for refusing the presence of the Roma and Hungarian minorities.

[35] see Appendix 1.II

Without entering into further detail regarding the social and inter-ethnic context, the above serves as a brief description of a majority – minority relationship that favors the propagation of hate speech thus making Romania a good candidate for a study of the phenomena. However the central focus of my thesis will be on user generated content, the complexities of regulating it, hate speech being the issue where this is most clearly visible as the shortcomings in regulation coupled with the wide spread intolerance amplify the phenomena making Romania a perfect case for my research.

Chapter II. The networked public sphere and user generated content

According to Van Dijk (2006) in the last century we have witnessed at least four "revolutions" driven by new technologies, that lead to utopian expectations about radical shifts in power and social relations, mostly based on the participatory nature of these technologies: "the notion of tele-democracy in the 1980s, virtual communities and the new economy in the 1990s, and most recently the Web 2.0."[36] An important feature of the latest 'revolution' is the increased interactivity in the media expressed most importantly through the central role of user generated content (UGC). According to the International Encyclopedia of Communication[37] interactivity is an "elusive concept" referring to the "phenomena of mutual adaptation between a media and human user". However as Bucy cited in the same entry points out, it is a "key feature of new media, but we scarcely know what it is". Based on several authors Robinson formulates a somewhat better definition: interactivity is "the ability to manipulate or otherwise modify someone else's content or add new content as audience member"[38]. There is no generally accepted definition for UGC. According to the OECD (2007) it is "content made publicly available over the Internet, which reflects a certain amount of creative effort, and which is created outside of professional routines and practices". Schafer distinguishes between explicit and implicit participation.[39] The former is motivation driven, and includes uploading content, posting, commenting, while the latter is driven by the interface, the automation of user activities and does not implies interaction with others or cultural production: it has a more active form like rating or tagging contents, but includes

[36] Van Dijk (2006, cited in Schafer 2011:25)
[37] Neumann, Russel W. 2008. Interactivity, Concept of. In *The international encyclopedia of communication*, 2318-2321. Malden MA: Blackwell Pub.
[38] Downes and McMillan, 2000; Steur, 1992 cited in Robinson, Sue. 2010. Traditionalists vs. Convergers. *Convergence: The International Journal of Research into New Media Technologies* 16, no. 1 (February 1): 125 -143.
[39] Schäfer, Mirko. 2011. *Bastard culture! : how user participation transforms cultural production.* Amsterdam: Amsterdam University Press. p.52

even being a member of a social or peer to peer (p2p) network or reading an online article.

Benkler talks about a 'networked information economy' which provides a more robust platform for public debate enabling citizens to participate in the public conversation "not as passive recipients but active participants", who can produce their own cultural environment, a culture that will be "more critical, more reflexive" resulting in a "networked public sphere"[40]. Benkler holds that the most important and durable effect of the Internet is that it ended the idea of a public sphere constructed by 'finished utterances by a small set of actors' and that "statements in the public sphere can now be seen as invitations to conversation rather as finished goods"[41]. The central role in this 'networked information economy' is played by the 'user' which is a new kind of relationship to information production in addition to the traditional producer/consumer, as it can be sometimes producer, sometimes consumer or even both at the same time.[42]

Schafer shares some of Benkler's views by agreeing with the fact that in the web 2.0. the role of cultural industries shifted from creator towards platform provider for UGC.[43] However he formulates his "extended cultural industry" model in direct contradiction to Benkler pointing out that this does not happened in order to empower the audience but rather to extend their production mode to the sphere of users, allowing mass media to "employ user activities in a way that clearly questions their status as producers."[44] Comparing his approach to the "participatory culture" model of "community driven appropriation of commercial media text" formulated by Jenkins, Schafer holds that web 2.0 revolution only meant that the mass media extended their production beyond established channels incorporating user activities into

[40] Benkler, Yochai. 2006. The wealth of networks how social production transforms markets and freedom. New Haven [Conn.] :: Yale University Press,
[41] idem p.180
[42] idem p. 139
[43] Schafer, *Bastard Culture 168.*
[44] Idem.p 168)

commercial media production. A similar but more radical view is shared by Fuchs, who talks about the "unpaid labor" of users and content creators exploited by Google,[45] by inserting abusive clauses in the service terms of use (TOS) which allow the company the use its users data as it pleases in order to gain profit. My research also found similar abusive TOS agreements whereby the media companies unilaterally impose contractual relationships where they are in control of all their users access data, but also reap all the benefit of the user-generated content while declining any responsibility for it.

A radically different view is expressed by Couture, who argues that the internet brought on the intrusion of private life into the public forum[46]. Kaufer shows that the internet not only allowed more forms for expression by giving access to mass audiences to individuals, but also eliminated the existing assumptions about the right to speak to the masses, which is now taken for granted by any self-selected speaker, contrary to the previous models where "speakers capacity to public expression was measured by their prior power to assemble a mass audience", raising the question how to turn this quantitative explosion into qualitative improvement of public communication.[47]

II.1. Online news sites

The transformation of traditional journalism into online news is an example where the issues described above about the intermixing and blurring of the categories of audience, publisher, host, public and private can coexist and thus be best examined in one place. The first online

[45] Fuchs, Christian. 2011. "A Contribution to the Critique of a Political Economy of Google." *Fastcapitalism* (8(1)). http://fuchs.uti.at/wp- content/uploads/Google_FastCapitalism.pdf.
[46] Couture, Barbara. 2004. Reconciling Private Lives and Public Rhetoric: What's at Stake? In *Private, the Public, and the Published : Reconciling Private Lives and Public Rhetoric.*, 1-30. Logan :: Utah State University Press, p.18
[47] Kaufer, David S. 2004. The Influence of Expanded Access to Mass Communication on Public Expression: The Rise of Representatives of the Personal. In *Private, the Public, and the Published : Reconciling Private Lives and Public Rhetoric.*, 153-165. Logan: Utah State University Press, p.155

newspapers and dedicated news sites were created at the end of the 1990s starting a migration and transformation process that evolved from an online variant of the paper edition into something specific to the internet[48]. In addition to the extended content possibilities of the new medium in most cases online news sites also provided space for the audience to "talk-back" on the comment sections, placing internet news "somewhere on a continuum between professionally produced content and the provision of public connectivity"[49]. While traditionally journalism's role was to deliver news to audiences, according to Deuze, in the online world interactivity is a more prominent feature, as news sites not only offer content but also a platform for participatory communication. He ranks the level of participation on a scale ranging from 'open' where users can post anything without moderation to closed where comments are under strict editorial control much like the letters to the editor in the traditional media.[50]

[48] . ……(Springer, 2004:3363)
[49] Deuze, Mark. 2008. Internet News. In *The international encyclopedia of communication*, 2447. Malden MA: Blackwell Pub.
[50] Deuze, Mark. 2003. "The Web and its Journalisms: Considering the Consequences of Different Types of Newsmedia Online." *New Media & Society* 5 (2) (June): 203-230. doi:10.1177/1461444803005002004.

II.2.User comments and their effects

In a newsroom ethnography Robinson documented the transition of a traditional newspaper to an online-only news outlet.[51] Her study reveals some basic differences between attitudes of the journalists and the commenting audience, the most significant being that the audience (wrongly) considered commenting a democratic right stemming from the first amendment, a privilege not always supported by the publishers. Although journalists admitted the importance of comments in community building and engagement, they also considered them a journalistic tool for audience feedback, information gathering, and also as way to create revenue by increasing "stickiness" to the site. Most importantly as mentioned earlier, Robinson found that users considered commenting as an exercise of their right to freedom of expression or even a form of journalism, arguing with moderators against the perceived censorship of their comments based on this right. On the other hand, journalists constantly reminded users that "they do not own the place and have no right to it"[52].

Goss analyzed from a critical discourse analysis perspective comments on the website of The Nation a major leftist magazine in the US. The author found that users made frequent use of *topoi* characteristic of the ideological discourse, to reproduce predominant ideologies along the republican/democratic party lines but also the class and gender divisions. Concluding that the "democratizing potential of the internet might be exaggerated" as the discourse of the

[51] Robinson, Sue. 2010. Traditionalists vs. Convergers. *Convergence: The International Journal of Research into New Media Technologies* 16, no. 1 (February 1): 125 -143.
[52] Idem, 138

comments "augment the day to day reproduction of sociological propaganda" even calling it "sociological propaganda in action" [53]

Van Dijk describes *topoi* as standardized and publicizing ready-made arguments that "need not to be defended and serve as basic criteria in argumentation."[54] Constituting "premises that are taken for granted as self-evident and a sufficient reason to accept the conclusion"[55]. *Topoi* are an important element of ideological and especially racist discourse such as for instance "immigrants are "burden to our country" but as van Dijk points out also of anti-racist discourse.

According to van Dijk in the discourse of people belonging to the majority regarding minorities the "preferred topics can be characterized by the concepts: of difference, deviance, transgression, threat."[56] Although I did not analyze comments from the critical discourse analysis perspective it is worth noting that my preliminary analysis also indicates a high use of *topoi* in the discourse of comment sections of the Romanian newspapers, such as the recurring and readily accepted argument that "no country in the world/Europe offers more rights to minorities than Romania." The results of the content analysis also point to the direction described by van Dijk, the three largest categories within the hate comments referring to stereotypes, minorities as representing foreign interest or being a threat and "this is our country" exclusionary arguments.

Ruiz *et al*. analyzed whether online newspapers and their comments sections create a dialogue-fostering environment, corresponding to the online version of 'bourgeois café' in the

[53] Goss, Brian Michael. 2007. "ONLINE 'LOONEY TUNES'." *Journalism Studies* 8 (3) (June): 365-381. doi:10.1080/14616700701276117.
[54] van Dijck, Teun A. *Ideology and discourse A Multidisciplinary Introduction.* Pompeu Fabra University, Barcelona. .53, 63
[55] Idem,.63
[56] Van Dijk, *Ideology and Discourse*, 46.

habermasian public sphere model[57]. They examined ethical guidelines and legal framework for user participation and analyzed the content of comments to assess the presence of the principles of discursive ethics, summarized by Habermas in three set of rules regarding logic and coherence; cooperative search for truth; and agreement based on the best argument. They also built their analysis on Hallin and Mancini's[58] three media systems model, which considers that political systems shape journalism culture and practices, expecting different type of participation in different media systems. Their sample of 16000 comments was drawn from two newspapers from liberal media systems (United States, and the United Kingdom) and three from the polarized pluralist model (France, Italy, Spain). According to their analysis of ethical guidelines, media organizations aim to create an environment favorable to dialogue trying to find an "equilibrium between freedom of speech and mutual respect." In the words of an editor interviewed by Ruiz *et al*.: "The aim of moderation is not censorship, but ensuring that the community participation areas of the site remain, appropriate, intelligent and lawful".[59]

Ruiz *et al*.'s analysis of the legal frameworks points to the direction described by Schafer[60] of media organizations using UGC to include users into their production models, revealing that while newspapers decline legal responsibility for comments, they do reserve the intellectual property rights for them. In fact the researchers found that in most cases when users post a comment, or join a site they implicitly enter in a contract with publishers where they are responsible for the content of their comments including legal liability, but ceding all intellectual property rights to the newspaper. As chapter four will show this is also true in the case of Romania, for all newspapers in the sample.

[57] Ruiz, Carlos, David Domingo, Josep Lluís Micó, Javier Díaz-Noci, Koldo Meso, and Pere Masip. 2011. "Public Sphere 2.0? The Democratic Qualities of Citizen Debates in Online Newspapers." *The International Journal of Press/Politics* 16 (4) (October 1): 463 -487.
[58] Hallin, Daniel C, and Paolo Mancini. 2004. *Comparing media systems : three models of media and politics.* Cambridge, U.K.: Cambridge University Press.
[59] Ruiz et. al, *Public Sphere 2.0*
[60] Schafer, *Bastard Culture.*

The analysis of comments in Ruiz *et a*l. revealed that while it is relatively easy to filter out insults or outright hate using automated methods like profanity filters, it is quite hard to moderate derogatory content, including hate speech as the audience is using a range of tactics to avoid automated methods. They found that comments were generally aligned with the ideological position of the newsroom. Moreover they also confirmed and extended Hallin and Mancini's model to audience participation. The two media systems resulted in different types of user behavior in the comments sections: "communities of debate" in the UK. and US characterized by argumentation, and dialogue in line with the internal pluralism characterizing newspapers in the liberal model. On the other hand in the polarized pluralist model audiences formed "homogenous communities", their participation being described as "dialogue of the deaf" typically users venting their opinion without engaging in dialogue.[61]

Singer *et al*. interviewed 70 journalists from leading newspapers of ten democratic countries about the role of audiences in the online newspapers[62]. They included the type of user generated content discussed in this thesis under the label of participatory journalism defined as "processes of ordinary citizens contributing to gathering, selecting, publishing, distributing, commenting on or publicly discussing, the news that is contained within an institutional media product such as newspaper websites" [63] (p.15). The authors identified five stages of the news production process: access/observation, selection/filtering, processing/editing, distributing, interpreting. Users have the most prominent role in the interpretation stage, with comments being the most widely offered form of user participation. A conclusion of their cross-country research is that journalists view audience members as "active recipients" and not "active participants" expressed by the resistance to open up other stages to user participation, keeping

[61] Ruiz et al, *Public sphere 2.0*.
[62] Singer, Jane B, David Domingo, Ari Heinonen, Alfred Hermida, Steve Paulussen, Thorsten Quandt, Zvi Reich, and Marina Vujnovic. 2011. *Participatory Journalism in Online Newspapers□: Guarding the Internet's Open Gates*. Boston [u.a.]: Wiley-Blackwell.
[63] Singer, Jane B, David Domingo, Ari Heinonen, Alfred Hermida, Steve Paulussen, Thorsten Quandt, Zvi Reich, and Marina Vujnovic. 2011. *Participatory Journalism in Online Newspapers□: Guarding the Internet's Open Gates*. Boston [u.a.]: Wiley-Blackwell.

it “at arms length”. This is evidenced also by the fact that places dedicated to audience such as forums were subject to more relaxed rules, creating “segregated playgrounds” whereas journalists tended to maintain control in spaces shared with professional content.

Reich shows that contrary to the efforts of web-designers to separate user generated and professional content using graphical elements like typography, “in reality the two types of content are inseparable” creating the hybrid nature of online news[64] already pointed out by authors such as Deuze in the definition of online news sites quoted earlier.[65] On the other hand by posting comments users are “authors without responsibilities that go with authorship” which according to Reich is an “intolerable situation” therefore responsibility has to be assigned to users or to the moderator.[66]

According to the study of Singer *et al*. media organizations maintain comments for commercial motivations: they increase traffic to the site, loyalty to the brand as users who comment tend to return to sites, and also stay longer therefore being exposed to more advertisements. From the journalistic point of view, users can also serve as potential sources, improve accuracy by pointing out errors, but the most important aspect is that they represent immediate feedback and information from the audience. However this feedback is heavily distorted and hardly representative as most authors studying comments found that only a minority of visitors actually comments.[67]

The move to online might have eliminated constraints such as space but the ethical constraints remain the same, as Singer points out.[68] On the other hand the control over content has changed radically. Consensus in the countries their the study is that the organizations that

[64] Reich, Zvi. User Comments: The transformation of participatory space. Singer et al, 2011
[65] Deuze, 2008
[66] Reich, 2011
[67] Kim and Hong cited in Lee and Jang
[68] Jane B. Singer: Taking responsibility (p.121) in Singer, Jane B, David Domingo, Ari Heinonen, Alfred Hermida, Steve Paulussen, Thorsten Quandt, Zvi Reich, and Marina Vujnovic. 2011. *Participatory Journalism in Online Newspapers : Guarding the Internet’s Open Gates*. Boston [u.a.]: Wiley-Blackwell.

post-moderate are not responsible for content, but become responsible to post-publication concerns such as the quick removal of offensive content. Contrary to their findings my analysis shows, that on most of the websites in my sample it is possible to find user-generated hate speech years after the publication of the article.

Some journalists interviewed by Singer *et al*. assumed responsibility for comments; in the words of an editor of the Canadian National Post "It's a debate we're hosting and we-re responsible for that debate (…) commentary on the site should uphold journalistic standards". According to Singer the "hands-off" approach prevalent in the United States could be a way to avoid responsibility. The newspaper does not moderate in order to not to appear as an editor who then can become responsible, but the consensus across the book seems to be "nobody knows yet who is responsible for that content"[69]

Cammaerts presents the use of blogs and an online forum to propagate hate speech targeted to immigrants and the Muslim community as reaction to three criminal acts that created interethnic tensions in North Belgium in 2007.[70] As the author shows, while the internet is an ideal platform for genuine deliberation[71] when the debate takes place in a dedicated space such as an extremist forum, it serves more for opinion reinforcement between like-minded individuals.[72] According to Cammaerts the fragmented nature of cyberspace prevents the encounter of hate speech if one does not specifically looks for it. However in my opinion the situation is totally different if such speech is allowed in public places like news sites.[73] Similar to Robinson, Cammaerts also found that forum participants or bloggers posting extremist speech often claim it to be their democratic right. Several of the posts analyzed by the author questioned the rights of immigrants to be in Belgium, making references to their inferiority or

[69] Idem, p. 134
[70] Cammaerts, Bart. 2009. Radical pluralism and free speech in online public spaces. *International Journal of Cultural Studies* 12, no. 6 (November 1): 555 -575.
[71] Coleman and Gotze 2001:17 cited by Cammaerts, 2009
[72] Davies 1999:162 cited by Cammaerts, 2009
[73] Cammaerts, 2009

calling them animals (rats) – enacting what Butler refers to the power of hate speech to "put one at his place"

Lee and Yang consider comments on news sites an "unprecedented interaction of the mass and interpersonal communication".[74] While traditionally newspaper reading was an individual experience in the sense that dissemination and reading took place in different settings and time, with online newspapers the exposure and the reaction to the news takes place simultaneously. Others reactions can serve as an indicator of the general climate i.e. readers could use journalistic content to infer about the public opinion, however in the new interactive context comments that can be perceived as sample of public opinion present directly and in the same place with the article influencing the perception of readers. The authors show that while only about 2,5 percent of readers actually comment,[75] the results of their experiment point to the fact that "exposure to others reaction indicated significant changes in how people make sense of their social environment" . The authors conclude that newspaper comments can "distort the accuracy of social reality perception as people no longer infer about the general climate from the news but from comments".[76]

[74] Lee, E.-J., and Yoon Jae Jang. 2010. What Do Others' Reactions to News on Internet Portal Sites Tell Us? Effects of Presentation Format and Readers' Need for Cognition on Reality Perception. *Communication Research* 37, no. 6 (7): 825-846. doi:10.1177/0093650210376189.
[75] Kim and Hong, 2009 cited in Lee and Jang 2010
[76] Lee, E.-J., and Yoon Jae Jang. 2010. What Do Others' Reactions to News on Internet Portal Sites Tell Us? Effects of Presentation Format and Readers' Need for Cognition on Reality Perception. *Communication Research* 37, no. 6 (7): 825-846. doi:10.1177/0093650210376189. p. 843

Chapter III. Hate speech and freedom of expression

The following chapter presents a definition of hate speech, the main legal approaches and an overview of the arguments and controversies surrounding it and linked to the fact that combating hate speech is in essence a limit on freedom of expression.

The International Encyclopedia of Communication defines hate speech as a "form of verbal aggression, that expresses contempt, ridicule, threat towards a specific group or class of people".[77] Although it refers to hate speech as verbal aggression the author of the entry Terry A. Kinney considers that it also includes all kinds of acts that demean or degrade, and believes that even if they are untrue and expressed by marginal groups they have the "ability to infiltrate our thoughts (…) affecting how we perceive ourselves." However regulating hate speech is difficult because it implies limitations to the freedom of expression. More importantly for the topic of this thesis, Kinney also points out, that the internet created "new communication spaces where this kind of speech can flourish" making legislation a more stringent question. The Encyclopedia of Political Communication offers a similar definition, focusing on the use of "words as weapons" with potential to terrorize, humiliate, degrade and discriminate victims[78].

According to Barendt there are two basic types of legal approaches to hate speech. In the United States "even extreme racist speech is an exercise of freedom of speech and is rightly covered by freedom of expression clauses"[79], and therefore it is protected by the first amendment of the Constitution as evidenced by numerous Supreme Court decisions. One of the more famous First Amendment cases is R.A.V. v. St. Paul, the case of a white teenager

[77] Kinney, Terry A. 2008. Hate Speech and Ethnophaulisms. In *The international encyclopedia of communication*, pp. 2051, 2054. Malden MA: Blackwell Pub.
[78] Rhea, David M. 2008. Hate Speech. In *Encyclopedia of political communication*, 301. Los Angeles. Sage Publications.
[79] Barendt, E. 2007. *Freedom of speech*. 2nd ed. Oxford ;;New York: Oxford University Press. p172

who burnt a cross in the backyard of a black family[80], and whose conviction was overturned by the Supreme Court. With regards to freedom of expression on the internet, it was under First Amendment concerns that the Court eliminated important segments of the Communications Decency Act (CDA) of 1996 aimed to regulate internet pornography.

In Europe as Lichtenberg[81] shows a social responsibility approach is more prevalent that weights the freedom of speech against other rights and protections, therefore making it relative. In Barendt's view limitations of free speech might be necessary not only to preserve social peace but to guard members of targeted groups from psychological injury and damage to their self esteem. In essence it is the expression of the right to equality and non-discrimination. Barendt adds that a strong argument for regulation is that tolerating hate speech would effectively lead to the state "lending respectability to racist groups and attitudes", therefore in this case limitations of free speech express the "right of the society to indicate abhorrence"[82]. Here I argue that a similar effect could occur in the case of user generated hate speech. The presence of discriminatory content on the site of a major newspaper, and the fact that it is tolerated might lead to the newspaper transferring some of its reputation to it, even lending it some credibility, thus making it more harmful.

On the other hand Barendt also warns, that hate speech, no matter how despicable, is a kind of political speech, and arguments used to ban it can be easily used to ban any other form of speech the government or a dominant group in society dislikes, leading to the main anti regulation argument that the "best solution to hate speech is more speech". This stands at the base of the marketplace of ideas argument[83] which claims that such opinions and beliefs

[80] Barendt, 2007, p. 185, In this case the Supreme Court overturned a Minnesota regulation, and thus dismissed the conviction of a teenager who burnt a cross on a black family's lawn
[81] Lichtenberg quoted in Cammaerts, 2010
[82] Barendt, E. 2007. *Freedom of speech*. 2nd ed. Oxford ;;New York: Oxford University Press. p.174
[83] Sorial, Sarah. 2010. Free Speech, Autonomy, and the Marketplace of Ideas. *The Journal of Value Inquiry* 44, no. 2 (1): 167-183. doi:10.1007/s10790-010-9200-x.

should be argued against and rejected on rational grounds, bypassing the need for government regulation, similar to the concept of self-regulation of financial markets.

The most important arguments defending the marketplace metaphor are according to Brison[84], "distrust in government" and the "slippery slope argument" both holding that whatever harm the exercise of absolute free speech causes in the long term it is still better then allowing government intervention. A shortcoming of the marketplace argument, according to Brison is being based on the assumption that the prevalence of "good" ideas in society will show the absurdity of hate speech. However since the "bad" ideas are directed against minorities it can easily happen that just as in financial markets the marketplace of ideas gets overrun by "bad" ideas simply because "good" ideas will be in short supply. This is similar to Delgado and Stefancic's argument who consider that the "marketplace of ideas" is designed to benefit, the majority and those in power even contributing to the marginalization of minorities, the majority's ideas can simply overrun "the market" where minorities already in a weaker position cannot gain access.[85] Parekh presents a reverse 'slippery slope' argument for banning hate speech. He considers that if hate speech is accepted as part of legitimate freedom of expression, those uttering it might feel encouraged, and gradually could even resort to physical violence against the targeted groups. As Parekh puts it, "if anything can be said about a group of persons with impunity, anything can also be done to it"[86]

Barendt considers, that the best argument for hate speech regulation is that it is not a "victimless crime" and mere expression of a political position but it can actually cause psychological harm[87], or as Matsuda said in her famous definition, these are "words that

[84] Brison (1995, quoted in Sorial, 2010)
[85] Delgado, Richard, and Jean Stefancic. 1997. *Must we defend Nazis?□: hate speech, pornography, and the new First Amendment*. New York: New York University Press. p. 89
[86] Parekh, Bhikhu. 2006. "Hate Speech." *Public Policy Research* 12 (4) (February): 213-223. doi:10.1111/j.1070-3535.2005.00405.x. p 217-218
[87] Barendt, 2007, p. 172, -174

wound"[88]. Judith Butler considers that hate speech injures by questioning the addressee's position in the community of speakers, attempting to put the addressee back in 'his place'.[89]

As this brief overview of the various arguments for and against hate speech regulation show it is indeed a controversial issue. While in the practice of the Supreme Court of the USA regulation against hate constitutes unacceptable limit on the freedom of expression, in this thesis I argue for the social responsibility approach. Perhaps the most fitting literature argument for the case of user-generated hate speech is that of Delgado and Stefancic[90]; we could imagine the comment sections as a marketplace of ideas but as the next chapters will show the "market" in question tends to be over flown by hateful ideas.

III.1. Freedom of the press on the internet. Blurring boundaries

This section is a short overview of the challenges faced by regulators approaching user generated content and the internet in general.

Prior to the internet era, delimiting individual freedom of expression and the freedom of the press was not so difficult as relatively few groups, and individuals (pirate radios, community media, newsletters) had access to publishing or broadcasting technology. With the advent of the internet the boundary between mass media and audience became blurred and it is now unclear who should benefit of the special provisions for the press such as for instance the protection of confidential sources. The question is whether freedom of the press is in any way different from the general right to free speech enjoyed by individuals. According to Barendt

[88] Matsuda, 1989, cited in Barendt, 2007, p 173.
[89] Butler, Judith. 1997. *Excitable speech : a politics of the performative.* New York: Routledge. p 4
[90] Delgado and Stefancic, *Must we defend nazis*, 88-89.

there are three perspectives on this issue.[91] In the United States individuals have the same rights as the media. However, it can be argued that in order to fulfill their vital role in democracies the media need special privileges, for instance access rights to official events, documents, places not available to the general public. On the other hand it is unclear how and why should then a journalist be differentiated from a blogger or any individual with access to publishing technologies. The second approach grants some special immunities and privileges to the media but it is again open to objections. Therefore a third approach emerged which grants some privileges to media institutions as long as it is in the public interest for instance such a privilege is the right to protection of sources in some countries.

Regulation of the internet is made complicated by the immediate and global nature of the medium: immediate in this case refers to the ease of distributing content without being filtered by professional gatekeepers[92], while the global nature refers to jurisdictional issues caused by the fact that content can be published and accessed across physical borders. As Barendt shows these features create specific problems in the regulation of hate speech and pornography. Another question is whether the net should be treated as an open forum similar to the streets where citizens have protected rights to express their opinion and to have access to other people's. According to Barendt the net is established by private actors therefore the providers rules apply. Thus in the case of the United States as other countries as well, the first amendment or similar free speech provisions are not engaged, meaning then the assumption of users creating UGC that they have some kind of protected right to free speech in online forums is wrong.

[91] Barendt, E. 2007. *Freedom of speech*. 2nd ed. Oxford ;;New York: Oxford University Press. pp. 417-424.)

[92] the lack of gatekeepers refers to the perceived absence of control, in reality there are many gatekeepers such as ISP's, hosting providers, site administrators making internet in fact into the most controlled media. For more on the issue of controlling and surveillance on the internet see. Morozov, 2011. However, in the case of user generated apparent content the lack of gatekeepers is one of the most important features.

Braman adopts a more nuanced view and distinguishes between public, quasi-public, quasi-private and private forums, differentiated according to ownership, functions, levels of free expression and privacy one might expect.[93] For the purpose of this thesis the most important of these categories are the quasi-private forums. The participatory spaces of news sites, such as comments could fit in this category similarly to restaurants and café's in the offline world. According to Braman although they are privately owned since they serve public functions there should be "some freedom of expression" in these spaces, yet restrictions are also legitimate if they are necessary for the functioning of the service. On the other hand Braman also points out that the case of the internet is complicated by the fact that to gain access one has to accept the providers terms of service (TOS), which became a "de facto communication regulation." It is this "de facto communication regulation role" in which I will examine the TOS of the sites in sample. The findings of authors such as Fuchs, Ruiz *et a*l.[94] and my research as well, shows that the acceptance of the TOS is implied when accessing the site, for both readers and commenters. On the other hand this implied consent only becomes evident at a detailed examination of the TOS, and because usually the users do not have to accept them in order to access the site their legitimacy can be questioned.

In addition to the already existing difficulties in regulating online hate speech, UGC presents further challenges to legislators. As Valcke and Lenaerts show, it is difficult to identify with certainty the traditional categories of author, editor, publisher, hosting provider on which media and early internet regulations are based.[95] Therefore the existing two approaches, the publisher or the Internet Service Provider (ISP) models, cannot readily be applied to UGC platform providers, leaving for the moment UGC in a grey area. Hateful comments on online news articles perfectly illustrate the legislators dilemma. While it can be argued that media

[93]Braman, Sandra. 2006. *Change of state*. Cambridge (Mass.)□;;London: the MIT press., 93, 94
[94] Fuchs, *The political economy of google.* Ruiz et al., *Public sphere 2.o.*
[95] Valcke, Peggy, and Marieke Lenaerts. 2010. Who's author, editor and publisher in user-generated content? Applying traditional media concepts to UGC providers. *International Review of Law, Computers & Technology* 24, no. 1 (3): 119-131. doi:10.1080/13600861003644533.

organization bears full responsibility for what is published on their pages (online or on paper) as under national media law, at the same time they can defend themselves by pointing out that media organizations merely provided a space for comments as in the hosting model, thus not being responsible for third party content.

III.2. Regulating online hate

Reviewing the legislation about online hate speech, Rorive points out that regulators worldwide tend to adopt legislation based on the principle that 'what is illegal offline is illegal online' even if this means authorities crossing traditional jurisdictional boundaries[96]. The 2001 Convention on Cybercrime contains an additional protocol criminalizing hate speech online. However in practice this has proven to be ineffective mostly because of the First Amendment that made the US a safe haven for online hate speech and hate groups worldwide. The European Directive on E-commerce (2002) was more effective although its direct aim was not to regulate content. Such directive instituted a limited liability for ISPs under a 'notice and take down' policy thereby creating economic incentives for ISPs not to tolerate hateful content on their servers. In practice the Directive turned out to be an effective tool to circumvent the First Amendment protection by making hate speech a matter of private law. For example US hosting companies with economic interests in Europe had more incentives to remove hateful content, for which they are generally covered under their terms of service, than risking lengthy judicial processes in Europe. On the other hand, Rorive admits that this

[96] Rorive, Isabelle. 2009. What Can Be Done Against Cyber Hate? Freedom of Speech Versus Hate Speech in the Council of Europe. *Cardozo Journal of International & Comparative Law* 17, no. 3 (October): 417-426.

approach is also problematic and should be treated with reservation because it effectively created economic incentives for private censorship.

Biegel places online hate into the category of "inappropriate conduct" which in his view includes "online activities that constitute intimidation, ridicule or insult" [97] and also "hostile behaviors" as harassment through email, on discussion forums, and dedicated extremist websites.[98] He differentiates it explicitly from "threatening behavior", which refers to "direct, explicit personal threats that may lead to physical injury". My definition throughout this thesis including the codebook for content analysis includes under the term user generated hate speech both types of content ranging from insults to threats with violence as long as it is based on the appartenance of the target to a certain ethnic/religios/sexual orientation group.

In Biegel's terms online hate refers to "words that discriminate on the basis of race, ethnicity, religion, sex, sexual orientation and disability", creating a "discriminatory hostile environment" but he also shows that online hate is hard to address as even the definition of "discriminatory harassment" is disputed. Moreover online hate is viewed not only as less dangerous, but also something that people should be able to tolerate just as "people walking on the streets should be able to tolerate some level of hateful, aggressive, or inappropriate conduct" (p.324). However as Biegel shows the online spaces are different not only that due to the speed on which such content can be disseminated but even more so because of their perceived anonymity, people tend to express views that they would hesitate in other public forum. Therefore the biggest danger of online hate is in the author's view that it could reverse the trend according to which "society no longer tolerates open expressions of prejudice."

According to Biegel another difficulty in fighting online hate is that in case of websites the harassment argument could be invalided as users are not forced to go there and the simple

[97] Biegel, Stuart. 2001. *Beyond our control?* Cambridge (Mass.); London: the MIT press. p. 86,
[98] Idem, p.321

knowledge of existence of such sites is not enough to constitute discrimination. However, I believe that in the case of user generated hate speech, the situation is different as visitors of the site access it looking for other, legitimate and professionally produced content but by means of comments they are inadvertedly exposed to hate.

Chapter IV. User-generated hate speech

A preliminary definition

Based on the above, the definition of user generated hate speech includes features from both the concept of user-generated content and hate speech. Although it could be included in both of them, I argue that it has some specific features that justify the creation of a new label. I define user generated hate speech as content (text, audio, video, multimedia), created by non-professional, anonymous users aimed at intimidating and/or harming particular minority groups (in ethnic and/or sexual, racial term) that takes advantage of the interactive features of websites, as well as of gaps in media regulation taking advantage or exploiting content oriented toward the general public, or content hosting platforms to be published and to reach its target audience. Such content displays some parasitic and viral characteristics: it needs a host such as an interactive website/UGC platform to exist as a parasite, but also the host is the one that transmits it to the victims as in a virus. It also exploits the weaknesses of user generated content, hate speech regulations and of media policy: i.e. anonymity, blurring delimitations of public/private, journalist/audience or the provisions protecting free speech especially the lack of regulation regarding the press. An essential characteristic of such content that differentiates it from dedicated hate blogs/forums/sites is that it is aimed at targets from the general audience and uses mainstream sites to reach it, while the readers of hate-sites are usually people who purposefully look for that content. By parasitizing mainstream sites it can reach a much wider audience. Furthermore by being attached to articles/content whose topic is relevant for the target group (e.g. an article about minority education or a video about an LGBT parade) it relies on the topic of the host to attract members of the target community to both the legitimate content and the hate-speech.

IV.1 Regulatory environment

My preliminary review of existing legislation revealed that currently Romania has several laws that refer or can be applied in cases of hate speech. However none of them refers directly to the press or internet.[99] The introductory part of this section illustrates the challenge posed to regulators by digitization and convergence i.e. the possibility of accessing the same content on different platforms, in this case print and online. Then I present the most important approaches to media regulation: statutory, market control and public responsibility, which will be followed by an overview of the Romanian legislation and regulatory environment regarding media and discrimination.

Theoretical considerations

Media convergence

Traditionally media and telecommunications used to be under different regulatory systems with different rights and responsibilities. According to Braman one of the most visible instances was editorial control[100], which was "unlimited in print, constrained in broadcasting, and prohibited in telecommunications." Digitization and other technological developments allow the same message to travel easily across all three, so today the "inherited legal categories no longer fit empirical realities". Dwyer shares a similar view considering that it is

[99] Romania, Govt. Ordinance *137/31 Aug. 2000., Art. 317 of the Penal Code, Law 107/2006.*
[100] Braman, Sandra. 2006. *Change of state.* Cambridge (Mass.)□;;London: the MIT press. p.68

no longer adequate to treat each medium separately and argues for the opening up the "old media silos"[101].

Media accountability

According to McQuail there are three media accountability frames.[102] The first is legal controls such as statutory regulations limiting media freedom, aiming to coerce some kind of behavior. Secondly comes market control: based on market theory products it disregards that high profit does not equate good content. Finally public responsibility, trust covering the self regulation framework, is considered by McQuail the "most suitable for expressing and implementing public interest, and holding free media to account."[103] However, he admits that this model could be considered weak, as it depends on the will of companies to comply. McQuail makes an explicit argument against regulatory convergence, arguing exactly for its opposite, considering that "diverse, overlapping and even conflicting regulations are more desirable" than unified ones, as more alternatives to accountability create more courts of appeal and less chilling effect. He also considers accountability mechanisms that reward good behavior preferable. In Romania, as the next pages will show, the printed press is left entirely to self-regulation as in the public responsibility model, a feature successfully exploited by user-generated hate speech as the authorities seem to be reluctant to apply the laws regarding discrimination that refer to "all kind of public behavior" including the press.

Regulatory framework in Romania

[101] Dwyer, Tim. 2010. *Media convergence*. Maidenhead□;;New York: McGraw Hill/Open University Press. p.14
[102] McQuail, Dennis. 2005. Accountability of Media to Society: Principles and Means. In *Communication Theory & Research*, 89-102. Sage.
[103] McQuail, Accountability of Media

The main Romanian legislation dealing with media is the law on the Audiovisual that established the National Audiovisual Council (CNA) as the sole authority with attributions in the field of media.[104] The CNA elaborates and periodically revises a media content code[105] that also contains anti-discrimination provisions stating that: "the broadcasting of any programmes containing any kind of anti-Semite, xenophobic manifestations, discrimination of any kind, and the denial, minimization or apologetic presentation of the crimes of the nazi and communist regimes is forbidden." [106] The council also decides on financial or administrative sanctions.

The telecommunications sector, including the internet, is regulated by the National Authority for Management and Regulation in Communications (ANCOM) that deals with issues ranging from mobile phone licensing, competition, consumer protection to keeping up to date, statistical records of the telecom industry.

Currently there is no legislation on the printed or online press, competent state authorities in this domain nor any self-regulatory bodies that could claim national legitimacy. There are some limitations imposed on the press under the general libel, defamation and privacy protection legislation currently included in the civil code. According to the Penal Code "the instigation to hate" is punishable by imprisonment, without specifying any exemptions for the press or requirements for such instigation to be done in public or under certain conditions.[107] On the other hand the Government Ordinance against discrimination specifies that if refers to "any public behavior that does not enter under the effect of the penal law"; therefore it can be considered as also referring to comments that can reasonably be considered as being "public

[104] Romania,. *Legea audiovizualului (Audiovisual Law.)* Law nr. 504 of 11 July, 2002
[105] Romania, *Decision nr 220 of 24 February, 2011 of the National Audiovizual Council (CNA*)
[106] Romania, *Decision nr 220 of 24 February of the National Audiovizual Council (CNA), art. 47.*
[107] Article 317 chapter 4, of the Penal Code (modified in 2006) "Instigation to discrimination. The instigation to hate on grounds of race, nationality, ethnicity, language, religion, gender, sexual orientation, political apartenece, convictions, wealth, social origin, age, disability, chronic uncontagious disease or HIV/AIDS infection is punishable by prison from 6 months to 3 year or fine."personal non-official translation

behavior". The legislation – "combating all kinds of discrimination"[108] refers to content usually described as hate speech and makes no exceptions for the press. The same ordinance established the National Anti-Discrimination Council (CNCD)[109] as an autonomous body named by the Parliament with responsibilities in monitoring and sanctioning discrimination that theoretically would also include the press.

In Romania journalists themselves are skeptical about self-regulation, 54 percent considering that there are no journalists of sufficient credibility to be elected in self-regulation organism. A majority of 70 percent even agrees that a press law would improve quality of journalism while only 34 percent believe that a self regulation and a deontology code would increase ethical behavior, 48 percent also confessing of not being familiar with any deontological requirements. [110]

Legislation regarding hate speech

Article 15 of Government Ordinance nr.137/31August 2000 (republished) about the right to personal dignity, "combating all kinds of discrimination", does not refer to hate speech explicitly but it is formulated in a way to include it by prohibiting: "any public behavior that has the character or nationalist-chauvinist propaganda, or any behavior that has as purpose of creating an intimidating, degrading, hostile, humiliating or offensive atmosphere against, or harms the dignity of a person, group, community in connection with their race, nationality,

[108] Article 15 about the right to personal dignity, of Government Ordinance nr. 137/ 31August 2000 (republished
[109] Consiliul National pentru Combaterea Discriminarii
[110] ActiveWatch Media Monitoring Agency, Centrul Pentru Jurnalism Independent (Center for Independent Journalism - Romania), and IMAS Public opinion resarch agency. 2009. *Autoreglementarea presei in Romania - Self regulation of the press in Romania*. Survey. ActiveWatch-Media Monitoring Agency (Romania), October. http://www.activewatch.ro/uploads/FreeEx%20Publicatii%20/Autoreglementarea%20presei%20din%20Romania.pdf.,35

ethnicity, religion, social category, conviction or sexual orientation."[111] As my research shows there are quite a large number of comments on the websites included in the sample that would fall under the provisions of this legislation.

Holocaust denial was criminalized in 2002 by Government Ordinance 31/2002[112] and penalizes with imprisonment both the "public denial of the Holocaust and its effects" and the public use of "fascist racist and xenophobic symbols" including slogans, or greeting formulas". The ordinance also clarifies that definition of Holocaust refers to acts committed against the Jewish and Roma population done by Nazi Germany and its allies, including Romania. This clarification is important because it extends the effect of the law to a frequently occuring theme in anti-Semite discourse that also appears in the comments that Holocaust refers exclusively to crimes committed by Germany against the Jewish population.

IV.2. Comparative analysis of user participation on the websites

This section presents a comparative analysis of participative features on the five websites. The analysis includes terms and conditions for the use of the site, guidelines for user participation, the existence of registration requirements, the position of comments in the page in relation to professional content, and apparent moderation policies. On the four websites that have dedicated forum sections the participation on these portions was compared with comments to articles on the main page.

[112] Parliament of Romania. *Law Nr. 107 of 27 April 2006* www.cdep.ro/pls/legis/legis_pck.htp_act?ida=64075&frame=0. . Chapter 2, Article 6.

Moderation policies

The cross-country study by Singer et al. identified two main comment management strategies: pre-moderation and post-moderation.[113] Pre-moderation is typical in Germany, where newspapers due to the nazi-past, holocaust denial and hate-speech legislation have stricter moderation policies that in other countries. Post-moderation usually involves some collaborative features: users typically have the option to click on "a report abusive content" link which will then be removed. Another approach entails tracking users and publishing their comments according to "reputation": comments of users who have a track-record of abusive content will be reviewed by moderators, while "super-users" or "trusted users" can post directly or even be granted moderating privileges. An important component in the case of post-moderation is requiring users to register. According to the journalists interviewed by the authors 60-90 percent of the comments are likely to be published. The lowest rates were recorded in Germany and Israel, for stories regarding religious or ethnic tensions, where comments "often cross the line into hate-speech" to a degree that some editors reported turning off commenting functions in case of sensitive stories or switching to pre-moderation if the site used post-moderation before.

Three of the five sites discussed in this study rely on post-moderation, and also allow user contributions in the integrate placement approach presented earlier, meaning that they effectively open their journalistic spaces to audience participation with very low level of control. On the other hand, with the exception of adevarul.ro the two other sites that use post-moderation have a profanity filter in place that filters out certain offensive or obscene words defined in a dictionary.

[113] Singer et al, *Participatory Journalism* 107

The profanity filters of the sites were tested by posting comments that only contained insulting epithets referring to minorities. The tests were performed on 22-23 May, 2012 by posting comments using the name Ion and a fictional email address ik@ionkommment.ro first by posting the sentence "I am commenting" *(comentez)* which was posted on all three sites. In a second step the same user name was used to post comments with insulting epithets to articles about the re-burial of a Hungarian poet that sparked diplomatic tensions between Hungary and Romania.[114] On gandul.info a comment containing a frequently used insult referring to Hungarians and another referring to Jewish people was replaced by an "*" to indicate that it was filtered out. However the profanity filter left in place two derogatory words referring to gay and Roma minorities, and also the word referring to Hungarians was published after it was slightly altered by inserting a dot after the first letter. A similarly easy-to-bypass profanity filter is in place on evz.ro which allowed the word referring to Hungarians after a space was inserted in it, although it was still clearly recognizable as the insulting word. This filter also refused the entire comment if it contained an unmodified insulting word. There was no profanity filter in place on adevarul.ro which allowed all comments with insulting words and even non-sense comments or a text taken randomly from other article but also a comment saying "this was a test of the profanity filter". However there is some kind of moderation or filter in place on adevarul.ro as the content analysis revealed that some comments were replaced with a text suggesting moderation, and users also accused the moderators of the site of censorship for not publishing their comments.

On the other hand on the two sites that use pre-moderation all of the test comments were refused, even those that contained full sentences but were not on the topic of the article. [115] A similar test performed on February 8th 2012 on the sites gandul.info, and evz.ro, adevarul.ro

[114] Anon. "Burial Plans of pro-Nazi Poet Sparks Hungary-Romania Row |." *Europe Online*. http://en.europeonline-magazine.eu/burial-plans-of-pro-nazi-poet-sparks-hungary-romania-row_211590.html.
[115] for screen captures of the profanity filter test see appendix…

and jurnalul.ro[116] revealed that it is enough to know the syntax of an email address in order to post comments even if the comment and the username itself shows that it is not a contribution to the discussion.

Although the above test was performed only once, therefore it cannot form the base of generalization, it is a good illustration of the weaknesses of the computerized moderation. While it can prevent the flooding of the site with obscenity and hate, in order to be efficient its dictionary needs constant updating, fine tuning and human supervision; otherwise it becomes easy to bypass even with a slight alteration of the excluded words, which will still be recognizable to the targets of hate speech, thus harmful.

There is no registration requirement for posting comments on any of the sites in the sample, not even for those using post-moderation, although as Singer *et al*. pointed out registration is an important element when relying on post moderation[117]. According to my tests the sites require an email address to post comments but do not check its validity. Therefore knowing the syntax of an email address is enough to gain access to the participatory spaces of every site in this sample, and to potentially reach audiences of millions of unique visitors drawn by the content provided by the newspapers.[118] Hotnews.ro the only site in the sample that uses community moderation requires registration in order to be able to participate in the moderation. Users of the site can give positive and negative votes to comments and the text of the comments whose total turned negative will be hidden only their title line remaining visible. Although this type of moderation is quite effective in maintaining the overall civility of the discussion its efficiency is reduced in preventing hate speech: in a majority-minority situation the number of users agreeing with comments directed against minority members can

[116] excluded from the final sample
[117] Singer et al., *Participatory Journalism*, 83.
[118] for an overview of the print an online audiences of the sites in the sample see table 2.

be higher thus still allowing discriminatory content. Voting for or against comments is also possible on adevarul.ro but comments remain visible even if their total is negative. On evz.ro users have an option either to vote for a comment or to report it to the administrators of the site. If a user chooses the ”report this comment” link he will have to provide the reason for reporting that contribution. This type of community moderation is also weak in case of hate speech, as usually the software used for controlling user activity will only actually report a comment to a human moderator if a pre-defined number of reports have been filed i.e. there might not be enough users reporting a comment as offensive in order to remove it. My test on evz.ro confirmed that even violent homophobic content remained on the site although I have reported it as discriminatory.

As mentioned earlier hotnews.ro and romanialibera.ro are the two sites that use pre-moderation i.e. comments have to be approved by a moderator before they are published. However as the results of the content analysis will show their pages also contain comments that can be labeled as hate speech.

Placement of comments in the page

All the sites in the sample place user comments on the same page with the articles written by journalists in a chronological or reverse chronological order in an “integrate placement” approach to user participation.[119] The user-generated and the professional content are separated using design/typograhical techniques, for instance by comments being placed in a different text box. However with the exception of adevarul.ro the sites do not use design or layout techniques to create a distance between the professional and the user-generated

[119] Singer et al., *Participatory journalism*, 103

content, the first comments being visible from the end of the article i.e. users who read an article until the end are exposed to hate speech against their will if there is such content amongst the first comments.[120] The designers of adevarul.ro placed the links and recommendations of other articles under the professional content and also the comment posting box thereby separating the two types of content: users have to actively move down the page to reach the comments making them avoidable for those who whish so.

Comparison of terms and conditions or ethical guidelines (TOS)

All newspapers in the sample explicitly prohibit the posting of discriminatory, xenophobic, obscene, insulting or violent content theoretically excluding hate speech from their pages. With two exceptions these rules are set down in terms and conditions of use or terms of service guidelines (TOS), which also contain provisions regarding intellectual property and responsibility for content. As a general feature these guidelines are difficult to indentify as they are placed on the bottom of the front page and with small fonts. I could not locate a TOS for romanialibera.ro; however this is one of the sites in the sample that uses pre-moderation and a warning message is placed on the commenting interface cautioning users that messages containing licentious language, or instigating to hate, racism, xenophobia, homophobia will be deleted. Similar warning messages are displayed on gandul.info and hotnews.ro; the later also warns users that they bear the entire responsibility for the content they publish including for damages resulting from any legal actions against such content. Hotnews.ro is the only site in the sample that states in the posting interface that by clicking on the "send" button the user agrees to the TOS.

[120] For an example see annex …with a derogatory word referring to hungarians placed right under the article and visible in the same screen without moving (scrolling) further down the page.

Responsibility and intellectual property rights for user generated content

Art. 2.2. of the TOS of hotnews.ro states unambiguously that all content published by users at the comment and the forum sections "becomes the property of Hotnews.ro from the moment they are posted on the site."[121] At the same time the terms and conditions has two entire paragraphs whereby Hotnews.ro declines any responsibility for content posted by users stating explicitly that the "user bears sole responsibility for the content of his comments and eventual legal consequences" although if they were published, by that time the intellectual property of the comments thus the potential benefits already belongs to the site. The TOS of evz.ro are almost exactly the same regarding responsibility and intellectual property, but the site also adds that by posting content on the site the user grants an "irrevocable and unlimited license to all his content including for the reproduction, transformation, retransmission on any channel and the creation of derivate works". Evz.ro is the only site in the sample that also contains a disclaimer concerning harms caused to users by any content of the site warning visitors that "by using evz.ro you acknowledge that you expose yourself to content that can be offensive, indecent, repulsive, and you agree to give up any legal rights or reparations that you could claims from evz.ro and you agree to grant evz.ro and its owners/partners with full immunity in the degree allowed by the law for all aspects regarding the use of this site".[122] In other words, although the site retains the full property rights for user generated content including the right for commercial use it declines any responsibility for its property and the harm it might cause to the users/readers of the site. Another problematic aspect of this is that although it assumes implicit consent to relinquishing legal claims, it does not appear before or the moment the user is accessing the site as a warning that the site also contains potentially

[121] Hotnews.ro. "Terms and Conditions of Use of the Hotnews.ro Website (Termeni Si Conditii De Utilizare a Site-ului Hotnews.ro)." http://www.hotnews.ro/stiri-general-5447989-termeni-conditii-utilizare-site-ului-hotnews.htm.
[122] evz.ro. "Terms and Conditions for Evz.ro (Termeni Si Conditii > EVZ.ro)." http://www.evz.ro/termeni-si-conditii.html.

harmful material and does not provide the possibility of declining the agreement by preventing the visit of the site as, for example similar warnings of pornographic sites do. Quite the contrary, the warning is part of the TOS located on the bottom of the front page. Thus by the time the user has the chance to read it, according to the terms quoted earlier he already gave up any rights for seeking reparations against harm caused by the site.

Participation on dedicated forums and comments

Four websites in the sample also have dedicated forum sections separated from the main site. The main difference between the forum and the comments to the articles is that by being separated from the professional content, visitors/readers of the site have to take a deliberate decision to access them by clicking on their links. Another major difference is that users have to register with a real email address in order to comment in these places, and the address is also checked, although in some cases it is also possible to comment as a "guest" without registration. With the exception of romanialibera.ro registered users can open discussion topics, thus enjoying a greater freedom in shaping the discussion, although some topics are usually created by the administrators of the site. A general characteristic of these forums across the sample is that the participation numbers are almost incomparably smaller than the comments on the main sites. For instance even on the most popular forum in the sample, Hotnews.ro, on Apr 27th, 2012 when the government of Romania was dismissed due to a vote of no confidence, there were only fifteen posts in the dedicated forum section that only had 300 views while the article on the same topic on the front page of had 221 comments and more than 50000 views (53778)[123]. The same is true for MyAdevarul, the forum of

[123] Hotnews.ro. "The Ungureanu Government Has Fallen. The Motion of No Confidence Was Approved with 235 Votes for and 9 Against (Guvernul Ungureanu a Picat. Motiunea De Cenzura a Fost Aprobata Cu 235 De Voturi Pentru Si 9 Impotriva)." http://www.hotnews.ro/stiri-politic-12103841-live-text-ora-9-00-parlamentul-dezbate-supune-vot-motiunea-cenzura-opriti-guvernul-satajabil-asa-nu-niciodata.htm.

adevarul.ro, that only had 53 comments on May 23, 2012 far less then on the comment sections in the main site. Although I could not identify a TOS for commenting on the main site, MyAdevarul has detailed guidelines for behavior (netiquette) on the forum, forbidding obscene, rasist or homophobic comments but also contributions written in upper case (meaning shouting). The operators of the site opened all the discussion topics on Romanialibera.ro. On April 27, 2012 the only active topic was the "question of the day" where 13 readers responded to the question "who do you think the president will nominate as prime minister?", while a similar article on the main site had 61 comments. The forum of evz.ro had around 300 comments in total, on May 23, 2012 the latest comment was posted five days earlier and the most debated topic had only 118 comments in total. At the same time an article on the main site about the communist past of a member of the government had more than 200 comments only hours after it was published.

Consequences of the TOS: who is responsible for user comments?

In my opinion the low popularity of the dedicated forums compared to the comments to articles might suggest that users who comment on the main site are drawn there due to the increased exposure of the later. Although forums offer greater freedom for users, they also have considerably smaller audiences. Users have the possibility to open their own topic but they also have to attract their readers and participants by having an interesting title, description or discussion starter, which might prove to be difficult without the added extra exposure that's available on the main site. In these regards the discussions in the forums are similar to one's individual website or even the "old" media model presented earlier from Kaufer's argument when those who desired to address mass audiences also had to assemble

their own public[124]. Comments to the articles on the other hand eliminate this requirement, without the need "to assemble an audience" or even have something interesting to say all users who know the syntax of an email address can have mass audiences readily assembled by the media organizations. Although as the analysis of the TOS of the five sites shows, users bear the full responsibility for their comments, the anonymity provided by the lack of registration requirements allows them to publish all kinds of content with very low risk of ever being held accountable for it. On the other hand media organizations do not consider themselves accountable for user generated content although it is their property as they appropriate the copyrights for it, including the right to the potential financial benefits while declining any potential disadvantages.

The result of these approaches is that on one side users are in fact exploited by the media organization, providing it with free content and bearing all the responsibility for it while relinquishing all the rights and benefits, confirming Schafer's and Fuchs's view presented earlier[125]. At the same time the attitude of the media organizations contradicts the basic ethical principle that one is responsible for one's property, placing comments into a gray area where nobody is accountable for them. Resulting in media content potentially reaching millions of readers for which nobody bears effective responsibility. In theory according to the TOS presented in this section users bear full responsibility for the content they publish through the sites. In practice however this responsibility is hardly enforceable due to the anonymous nature of the comments, and the complicated legal process resulting from the technical characteristics of the internet. A person seeking to hold users accountable for the comments would need several court orders just to identify the person behind the nick/user name. First a court order would be needed to obtain the access data from the site, but this would only result in an IP address which will in turn require another court order to get the connection and

[124] Kaufer, *The Influence of Expanded Access*, 155.
[125] Schafer, *Bastard Culture*, *168*. Fuchs, *Political Economy of Google*.

subscriber data from the Internet Service Provider. The process can be further complicated if the user posting the hateful comment accessed the site from outside of the country. These steps require that both the news site and the ISP track and keep detailed records (logs) of the connection and access data for their subscribers and users, which is highly problematic from the perspective of citizens rights to privacy. Moreover it would be almost impossible for a regular individual seeking reparations for hate speech to identify a user who posted a comment from a public connection such as those available in cafés. To sum up despite the fact that Romania has adequate legislation regarding discrimination, the participation models presented in this section lead to a situation where users can post all manner of hate speech in the extent allowed by the sites who can also gain financial benefits from user participation while persons harmed by such content have almost no possibility in getting remedies. In short, the answer to the question posed in the title of this segment is: in practice nobody.

IV.3. Content Analysis of Comments

As the two previous sections have shown, despite the fact that there is not a distinct law regarding hate speech, discriminatory content included under the term is prohibited by a range of regulations in the existing Romanian legislation. Additionally, all the sites in the sample forbid the posting of such messages either in distinct guidelines, TOS or warning messages. However even a superficial preliminary analysis of the comment sections reveals the presence of comments containing insults, threats but also extreme violence or even calls for the murder or rape of persons belonging to target groups. There seems to be a consensus on disallowing comments that are hateful, discriminatory, xenophobic, instigate violence or hate, but there is no definition on either of the sites about what exactly is meant by these terms. In order to assess the effectiveness of both the legislation and the ethical guidelines of the sites I have assembled a sample of articles and comments according to a methodology described in the first chapter.

Codebook: Assessing effectiveness of sites participation policies and anti-discrimination legislation

Since the sites do not provide definitions or description for what they mean by discriminatory content I have used instead the academic definitions from two major encyclopedias in the field of communication, the International Encyclopedia of Communication and the Encyclopedia of Political Communication. Although the exact meaning and definition of the term hate speech is subject to controversy in my opinion the definitions in the two encyclopedias reflect at least a general agreement in academia as well as the communication profession on what is understood under hate speech. The two encyclopedic definitions were

extended with the definition of discrimination from the Romanian legislation against discrimination[126] resulting in the following definition of hate speech that formed the base of the codebook.

[126] Romania, *Government Ordinance 137/31.08.2000*

> Comments containing speech aimed to terrorize, humiliate, degrade, abuse, threaten, ridicule, demean, and discriminate based on race, ethnicity, religion, sexual orientation, national origin, or gender;[127]expressing prejudice, and contempt, promoting or supporting discrimination, prejudice and violence; seeking to distort the history of targeted groups, to eliminate their agency, to create and maintain derogatory cultural, racial, and ethnic illusions about targeted groups. Also including pejoratives and group based insults, that sometimes comprise brief group epithets consisting of short, usually negative labels or lengthy narratives about an out-group's alleged negative behavior.[128] Discrimination is considered to be any differentiation, exclusion, restriction or preference based on group appartenance and any other criteria, that is aimed or has the effect of restricting, limiting recognition, use or exercise in conditions of equality, of human rights, and of fundamental freedoms, or of rights recognized by law, in the political, economic, social and cultural and any other domains of the public life[129].

The above definition incorporates discriminatory content addressed by the three laws presented earlier including the provisions of the Penal Code regarding instigation to hate[130] and the law criminalizing holocaust denial ("seeking to distort the history of targeted groups").[131]

The definition of hate presented earlier was then expanded into 23 subcategories referring to types of hate speech occurring in the comments sections which are presented in detail with their definitions in the codebook and coding protocol in Appendix 2. These subcategories were developed from elements of the definition above and are grounded on the existing Romanian legislation, and the academic definition as well. For instance a frequently occurring theme in the comments that was labeled "exclusion/this is our country" refers to comments that invalidate minority groups claims for rights or even ask for their expulsion on the grounds the majority group is the rightful "owner" of the country and therefore people belonging to minorities have no legitimacy to ask for rights, keep their language, customs and traditions or even exist in the territory of the country. Beyond the three anti-discrimination laws mentioned earlier this type of discriminatory argument is against the Constitution of

[127] Encyclopedia of Political communication, 2007:301
[128] International Encyclopedia of Communication, 2007:2051
[129] Art. 2 of OUG 137/31 Aug. 2000
[130] Article 317 of the Penal Code of Romania
[131] Romania, *Law Nr. 107 of 27 April 2006.*

Romania itself which states in its fourth article that "Romania is the common and indivisible homeland of all its citizens, without any discrimination on account of race, nationality, ethnic origin, language, religion, sex, opinion, political adherence, property or social origin."[132] Moreover comments denying the Holocaust and making the apology or praise of leaders or organizations involved labeled in my codebook with "Holocaust denial/minimization", "Holocaust blame shifting", and "Holocaust, Fascism apology justifications" are explicitly liable of prison sentences according to the aforementioned law criminalizing Holocaust denial[133]. As mentioned in section I.5 the social attitudes regarding minorities revealed by the surveys of the CNCD have also served as indicators and guides in the creation of the subtypes.

The purpose of creating a codebook based on these sources and using it as a tool for content analysis is threefold. First my intention was to show that the existing Romanian legislation without the need of additional media regulation could efficiently be used to identify and to restrain content falling into the definition of hate speech. On the other hand I also intended to show that despite legislation and ethical guidelines that in theory should prevent the presence of such content into the media and the entire public sphere, this type of content not only occurs sporadically, but it is quite an extended phenomena pointing to a loophole in the media policy that allows the presence of such content. Thirdly my intention was to create a codebook that could in practice be used as a tool for moderation, as not only it expands the academic definitions into subtypes making it easier to identify such content, but it is also built on the basis of the existing legislation and it is also taking into consideration the social realities and the most frequently occurring themes and attitudes.

[132] Article 4.2 of the Constitution of Romania
[133] Romania, *Law 107/2006.*

Coding frame

The comments were coded on two levels. The first level codes were 'hate', and 'non-hate', 'hate' referring to all comments that could fit into the definition provided earlier; comments could only be coded in one of these codes (binary coding). The second level codes referred to types of 'hate' and multiple codes could be assigned to one comment with the exception of 'legit' i.e. legitimate content according to site guidelines and the legislation that could not be assigned to comments that have any other sub-codes. Non-hate comments that should not have been allowed according to the terms and conditions or terms of use of the sites were coded with 'insult', 'violence', 'junk/spam' – all other comments that have not been assigned a code were automatically assigned by the software the code 'legit' i.e. legitimate discussion. 'Hate' referred to comments targeted to members or groups/communities, while 'insult', 'violence' 'profanity" in the non-hate group referred to comments targeted at individuals without making reference to their group appertenance. 'junk/spam' – refers to comments that have no comprehensible content, contain advertisements, or other similar content. The nicknames/usernames of the users and the subject lines were also considered as being part of the comment as they frequently contain insults in an attempt to bypass the profanity filters of the sites. The definitions in the codebook were formulated with the intention to allow their use for identifying any kind of group based discriminatory content directed against all types of targets not just a minority/majority situation. Therefore the terms 'group A' and 'group B' were used instead of majority or minority, where 'group A' refers to in-groups while 'group B, C, D' to out-groups. This approach also allows the identification of hateful content directed by members of minority groups against the majority. Even though some of these comments coming from members of minority groups were posted as responses to previous provocations or verbal attacks in my opinion that does not justify the use of

discriminatory language; therefore they were also coded into the hate category.

Content Analysis: findings

Despite the existing legislation prohibiting discrimination, the guidelines (TOS) and the warnings that the sites do not accept or will delete xenophobic, instigating, hateful comments, the proportion of hate comments in the entire sample is 37.99% while legitimate comments i.e. contributions to the discussion that are not hateful, insulting or threatening only account for 61.08% of the comments in my sample. Although generalizability of these finding is limited to the purposive sampling the presence of hate speech in such large proportions points to major deficiencies in managing user participation from the part of site administrators and authorities.

Figure 2 shows the proportion of hate speech on comment sections on the individual sites. The highest percentage was found on gandul.info where 48.29% of user contributions in the sample have been included into the hate category. On evz.ro the proportion of legitimate comments is also decreased by the presence of the large amount of comments containing insults and profanity in the non-hate category which make up 4.64%, the largest proportion in this category between the five sites.[134] Surprisingly the lack of profanity filter on adevarul.ro is not abused excessively the site having the lowest proportion of insults/profanity within the post-moderation group both in the hate 16.62% compared 27.17% on Gandul.info and 21.52% on evz and non-hate categories, although the other two sites have profanity filters. More importantly even though the site shows visible signs of moderation 32.39% of comments on adevarul.ro were in the hate category, including an alarming 3.44% of

[134] Data not shown here: see Appendix 3.

comments that calling for the extermination murder or rape of persons belonging to the target groups.[135]

Contrary to the other three sites Hotnews.ro and Romanialibera.ro use pre-moderation i.e. comments are not posted instantly but need the approval of a moderator; hotnews.ro also using community moderation. Nevertheless moderation seems to be inconsistent on Romanialibera.ro as the proportion of hate comments is even slightly higher than on adevarul.ro that published all comments instantly. Admittedly the site had the second lowest proportion of 'extermination/murder/rape' comments (1.01%) and of insults (8.60%) which suggests that moderators watch more closely comments that are easy to identify as infringing.

The proportion of legit comments of on hotnews.ro is almost 10% higher than the other sites. This is maintained using a dual moderation system of pre-filtering abusing content by moderators and community moderation in form of voting, also showing that it is possible to efficiently moderate comments. Furthermore hotnews.ro is the only site in the sample where community moderation not only has visible consequences by hiding comments whose total turned negative but the site also requires users to register in order to vote for comments. This could lead to an increased proportion of registered users participating in the discussion who might not post hate-speech to avoid the banning of their account. Besides moderation, this feature might also explain the lower amount of hate-speech.

[135] For an illustrative sample in approximate English translation of comments in the 'extermination/murder/rape' subcategory see Appendix…

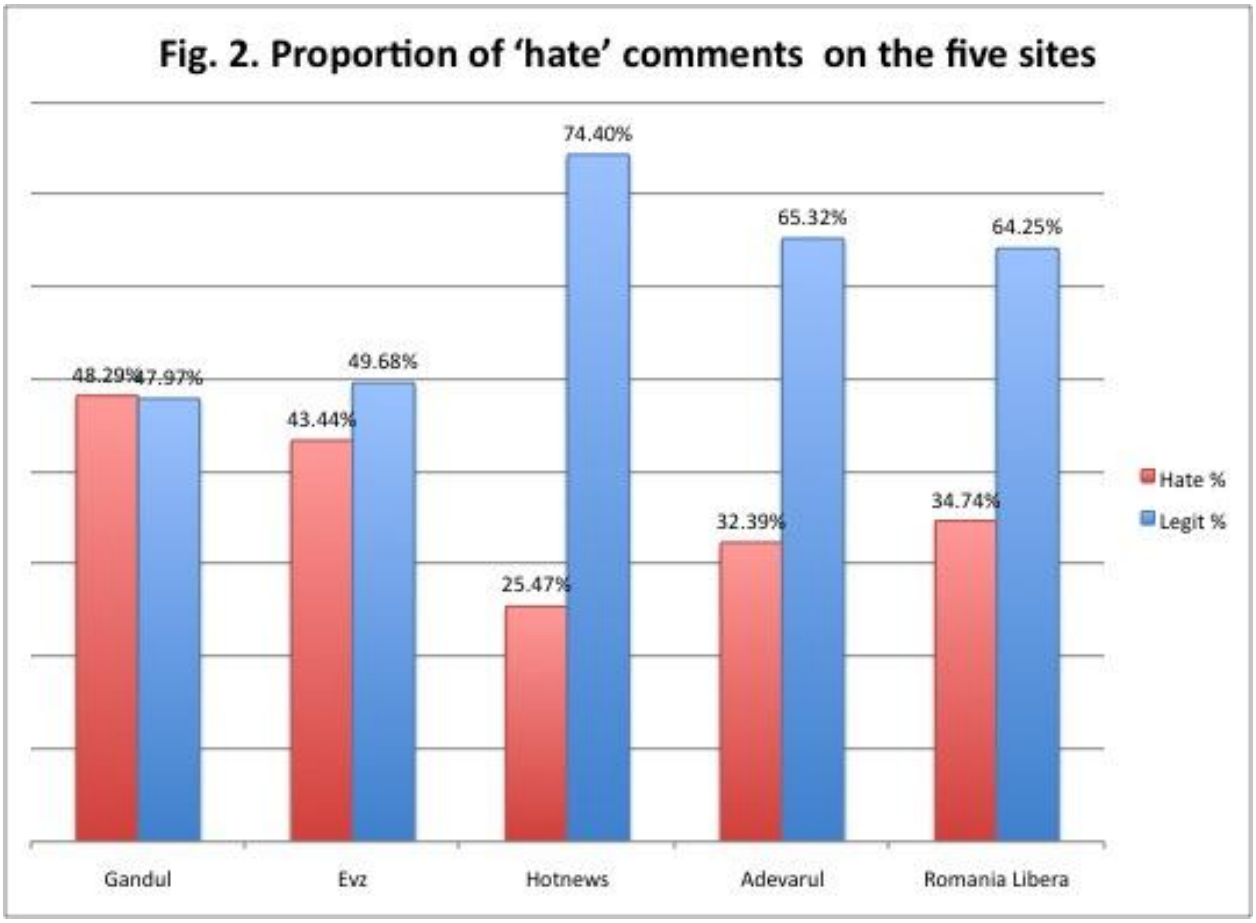


Table 2. Hate speech types

HATE	**Comments**	**Percent**	**Percent of Hate**
Insults	1106	18.08%	47.59%
Stereotypes/Generalization/Prejudice	522	8.53%	22.46%
Conspiracy/Foreign interests/Enemies/Threat	397	6.49%	17.08%
Exclusion/This is our country	341	5.57%	14.67%
Extermination/Murder/Rape	245	4.01%	10.54%
Superiority/Inferiority/Normality	194	3.17%	8.35%
Denying rights (political/civil)	186	3.04%	8.00%
Expulsion	165	2.70%	7.10%
History	158	2.58%	6.80%
Threats	148	2.42%	6.37%
Violence	141	2.31%	6.07%
Animals/Sub-human	120	1.96%	5.16%
Religious extremism	88	1.44%	3.79%
Holocaust-denial/minimization	86	1.41%	3.70%
Holocaust - blame shifting	81	1.32%	3.49%
Discrimination	73	1.19%	3.14%
Holocaust, Fascism - appology/justifications	71	1.16%	3.06%
Moderated	59	0.96%	2.54%
General hate/Discrimination	58	0.95%	2.50%
Homosexuality-Pedophilia	50	0.82%	2.15%
Disgrace for the country	25	0.41%	1.08%
Hate-Spam	18	0.29%	0.77%
Sterilization	5	0.08%	0.22%
NON-HATE	**Comments**	**Percent**	**Percent of Non-hate**
Legit	3597	58.80%	94.76%
Insult/Profanity	143	2.34%	3.77%
Thrash/Spam	44	0.72%	1.16%
Threat/Violence	8	0.13%	0.21%
HS target responding	6	0.10%	0.16%

Proportion of hate speech types

Table 2 shows comments in the hate category divided into the 22 hate types.[136] Some of the categories as for instance 'insults' are straightforward while the classification of others as hate speech might be subject to interpretation. Comments coded "exclusion/this is our country" or "conspiracy/foreign interest/threat" might not be considered by all moderators as discriminatory. On the other hand as I have shown these do fit into the legal definition of discrimination, and could even be included under the effect of the penal law. Presenting an entire minority as enemies, threat to the state or undermining society could be interpreted as instigation to hate. The inclusion of the two types into the category of hate speech is also justified in the light of the survey results showing a proportion of respondents refusing even the presence of minorities in Romania. Although I have used a different sampling, data collection and analysis methodology the proportion of comments in the "Exclusion/This is country" category (5.57%) is similar to those found by the CNCD's 2012 survey[137] where 6% of the respondents considered that Roma should not come to Romania, while 4% hold the same view about Hungarians. The survey thus confirms and confers external validity to my findings.

Insults make up the largest proportion of 'hate' comments (47.91%) although this is perhaps the most easy to manage hate speech type. A large proportion of these comments contain frequently used derogatory words such as "*bozgor*" referring to hungarians[138], "crow" (*ciora*) referring to Roma, "*jidan"* (pejorative version of Jew), "*homolau*" (distorted pejorative reference to Homosexual) with slight variations compared to the dictionary form. Their publication could be prevented by a regularly updated and well configured profanity filter.

[136] for the definitions of the categories see Appendix I.2
[137] CNCD, *Perceptions and attitudes.*
[138] Word of unknown meaning and origin allegedly meaning person without a country – it is used in reference to Hungarians in Romania implying that Romania is not their country.

Even in the case of comments not containing the specific demeaning terms, their character is quite obvious and relatively easy to moderate. The second largest category; comments expressing generalization stereotypes and prejudice totaled 7.68% for the entire sample and 21.40% of the hate comments. The legal categorization of these comments is indeed more difficult as it might be hard to differentiate them from legitimate opinions, on the other hand comments such as those presented in Appendix 4.1 clearly place a negative label on all the members of a community and have the potential to instigate to hate against them. Moreover the first two come from hotnews.ro a moderated site, illustrating that moderators have a different interpretation for instigating and xenophobic content. Regarding sexual minorities the preliminary analysis indicated that homosexuality is frequently associating with pedophilia, therefore a separate code was created for it. For the subsample of articles regarding homosexuality comments labeled with this code measured 4.14 percent. The comment in appendix 4.2. also labeled with the stereotype/prejudice label provides an illustration for this type.

While the classification of the comments in the categories presented above might be subject to some interpretation, comments denying the Holocaust, if prosecuted could even lead to prison sentences.[139] However, they still make up more than a quarter of the comments on the articles regarding the Holocaust: 9.55% holocaust denial; 8.73% claims that the victims somehow deserved the holocaust typically for being guilty for the crimes of the communism and 6.72% making an apology of leaders and organizations guilty of the genocide or seeking

[139] Law 107/2006 Romania

justifications for their actions.[140] The example bellow illustrates comments liable for prison sentences, the last two were also coded with 'extermination/murder/rape" (E/M/R). [141]

> "What Holocaust???? There was no such thing. Only the *ji dans*[142] sustain this high and strong. But who brought the communism to the world? The *ji dans*"[143]

> "the Deportation of the Jews in the 2-nd World War was legitimate. They were pro-communists (…) All the countries had camps for the hostile population"[144]

> "I don't deny anything, but let me express a regret: TOO BAD THAT NEITHER HITLER OR ANTONESCU FINISHED THE JOB. Did I deny something? No I did not. Regarding the jews I wish them to remain as many as I have baptized"[145]

> "All the time *jidans* and holocaust their suffering and all the fables repeated obsessively. Why? (…) We had enough of the filthy jidans and their fairytales!!! DEATH TO THE JIDANS!"[146]

The first comment shows the ease of bypassing profanity filters with the word *jidan* divided in two syllables. The second comment was posted on a moderated forum raising the question of liability of the moderators for allowing it. Similarly the last comment was also posted to an article that had signs of moderation as 4 (8.16%) of the 49 comments were deleted. Nevertheless the moderation software or human moderator left in places 6 comments containing death calls (coded E/M/R) and another 16.32% of comments in the holocaust minimization and blame shifting category. The total amount of hate for this article presenting the story of a Holocaust survivor[147] was 44.90% hate, 53.06% legit. This article also shows that the administrators of the adevarul.ro site do not use the voting system as community

[140] , typically praising Ion Antonescu Romania's leader during the second World War, convicted to death and executed for being guilty in genocide or his organization the Iron Guard.
[141] See appendix 4.2 for the Romanian original of the comments, the number after # is the unique identification number assigned to every comment during the content analysis. Access to the database is available on request.
[142] pejorative term referring to Jews modified in order to bypass the profanity filter
[143] #1422 posted by Anton Escu on Mar 6th, 16:58
[144] #6427 posted by observer on 18:45, 23 June 2011, on romanialibera.ro
[145] #1404 posted by rsss on Mar 8th, 08:57
[146] #6299 posted by anti-evrei on 2012-03-20 10:50:55 on adevarul.ro
[147] Adevarul.ro. "The Oldest Survivor of the Holocaust: 'I Will Be Laughing Until the End of My Life. It Does Not Help with Anything If You Cry' (VIDEO Cea Mai Bătrână Supraviețuitoare Din Lume a Holocaustului: „Voi Râde Până La Sfârșitul Vieții. Nu Ajută Cu Nimic Dacă Plângi")."
http://www.adevarul.ro/life/VIDEO_Cea_mai_batrana_supravietuitoare_din_lume_a_Holocaustului-_-Voi_rade_pana_la_sfarsitul_vietii-_Nu_ajuta_cu_nimic_daca_plangi_0_667133281.html.

moderation, as the comment above received 23 negative and 9 positive votes but it was still in place months later. The user in the third comment shows that he is aware of the legislation criminalizing Holocaust denial, but also that he can act with impunity pointing to the failure of the current approach to user participation.

The most disturbing type of hate speech is the category labeled "Extermination/Murder/Rape" referring to extremely violent comments that contain open and explicit calls or threats for murder, genocide or the rape of persons in another group. There might be some space for interpretation for the other categories, on whether the content they refer to is under the extent of the law or media ethics, or if they are within the limits of free expression, but in my opinions the attitudes expressed in the 245 comments (4.01 percent of the entire sample) in this category go well beyond what can be considered to be protected by the right to free expression.

The fact that they still appear and in such large numbers suggests that the media organizations either are indifferent for the content on their pages or it could mean that the 'E/M/R' comments are left there intentionally. As shown in Table 3 Comments advocating for murder or extermination occur across the entire sample, in the highest proportion on articles regarding the Roma (8.80%) while for the rest of the target groups the proportions are 5.42% for Jewish people, 3.74% Hungarians, 2.24% LGTB. This category makes up 10.54% of the hate comments; its proportion is highest on gandul.info where it makes up for 7.48% of the sample, followed by evz.ro with 4.16% and adevarul.ro 3.44%. It is not excluded completely even on the sites that have pre-moderation but the proportions are significantly smaller: 1.01% on romanialibera.ro and 0.54% on hotnews.ro. The following are examples of the comments that are clearly recognizable as extremely violent hate speech and would be quite easy to filter out or moderate.

> "SPIT THE HUNGARIANS WHEREVER YOU FIND THEM/ RAPE THE BOZGOR WOMEN AND THEN KILL THEM/ BURN DOWN THE HUNGARIAN BUSINESSES/ SPARE BULLETS! SHOOT TWO HUNGARIANS AT ONCE. DEATH TO THE BOZGORS[148]
>
> "that's why I say that a good gipsy is a dead gipsy, these are not humans, they are as damaging as the rats"[149]
>
> "I don't like hungarians but I would gang-rape the blonde one in the picture with the boys from my gang! I bet that after a session she will change her name into a Romanian one"[150]
>
> "I've been saying for a long time that the hungarians have to be killed or deported to their Panonia."[151]
>
> **…**being homalau (derogatory reference to homosexual) is a choice!!! … for those incurably homalau the FINAL SOLUTION should be applied…, …The homalau-s have to be treated as the pedophiles"[152]
>
> "I couldn't stand Russians my entire life, but the Hungarians I swear I would align all in a row and shoot them from the first 'till the last" [153]

As mentioned in the methodology all the comments in the sample were posted to articles about minorities, which in my view increase their harm potential. Allowing open calls for genocide against a minority on the webpage of an important national newspaper can serve as a catalyst for further hateful content. If a moderately intolerant person who already has some negative opinions about a minority reads such comments, he can be under the impression as Biegel points out[154], that society at large shares his views, and that such views are legitimate; if he then finds comments that even call for the murder and are not deleted, he might even feel encouraged to voice his views and the comment section can soon turn into a 'hate contest' were users start to compete on posting more violent content.[155] On the other hand it is reasonable to suppose that there is a higher chance that readers/visitors from the minority

[148] #526 posted by Alin on 2011-12-29 17:39:00 on gandul.info
[149] #2088 posted by Laurentiu on 2012-04-19 17:43:40 on adevarul.ro
[149] #1149 posted by Daul Ab Uci on 2012-03-19 13:31:58 on adevarul.ro

[151] #6504 posted by mihai on 20:00, 16 June, 2011 on romanialibera.ro
[152] #6790 posted by Misu on 09:20 | 22 April, 2011 on romanialibera.ro
[153] -#960 posted by zau zau on Feb 29th, 2012, 23:57 on evz.ro
[154] Biegel, *Beyond Our Control*
[155] for an example see the

groups will read articles about/regarding their community, the sites effectively helping the hateful comment to reach its target.

Distribution of hate based on target groups and topics

Table 3 shows the proportion of hate content divided along the five target communities. The data shows surprisingly small variations. Since the articles were collected on different times, and on different topics; the number of articles collected also differed this relative stable proportion of hate speech along the target groups in the sample suggest that the population of users posting hate speech is also relatively stable. In the light of the survey results regarding social distance the lower amount of hate for the articles about homosexuality is surprising, but this might be due to distortions caused by the nature of the sample.

Table 3. Proportion of hate speech against target groups

Topic	Articles	Comments	Hate	Hate Percent	Legit	Legit Percent
Hungarian	41	3640	1377	37.83%	2193	60.25%
Jewish	18	848	334	39.39%	467	55.07%
LGTB	17	1184	431	36.40%	693	58.53%
Roma	7	409	173	42.30%	219	53.55%

In the distribution of hate amongst article topics[156] shown in Figure 3 the articles relating criminal acts committed by Roma prompted the worst reaction, with 54.05% of the comments being hate-speech. This is followed by content directed against Hungarians, hate making up 52.59% of the comments on articles about the 'Territorial reorganization' topic. This refers to the failed initiative to reorganize the territory of the country into regions, abandoned due to

[156] See table .. in appendix..

the refusal of the Hungarian party (DAHR) to vote for it in the Parliament.[157] The articles presenting the position of the Hungarian politicians generated the highest proportion of hate along the topics; beside the 28.02% of comments containing insults, 13.55% were in the 'exclusion/this is our country" category generally expressing the view that Hungarians do not have say in the way Romania is organized since it is not their country.

The negative attitudes were also fueled sometimes by the media organizations. For instance the article with the largest amount of hate in the entire sample is on this topic[158]. In the material published on gandul.info the journalist distorted a statement of a Hungarian county leader to imply that Hungarians would even resort to violence to stop the reorganization prompting more than 300 comments 78.67% of which were hate. [159] Moreover 23.33% of the comments to this article were in the E/M/R category; of the 74 such comments within the topic 70 were posted to this single article. It should be noted that almost a year after it was published in May 2012 the article including the extremely violent calls to genocide were still on the gandul.info site now totaling 674 comments and 28514 views[160]

Despite the above example my findings indicate that the amount of hateful comments is independent of the title or the occasional instigating: articles with tendentious titles can receive less hate comments whereas well intended articles can prompt higher amount of hate speech, suggesting that the target group or the topic is attracting hate comments not the wording of the article. The unexpectedly high proportion of the legit comments for the

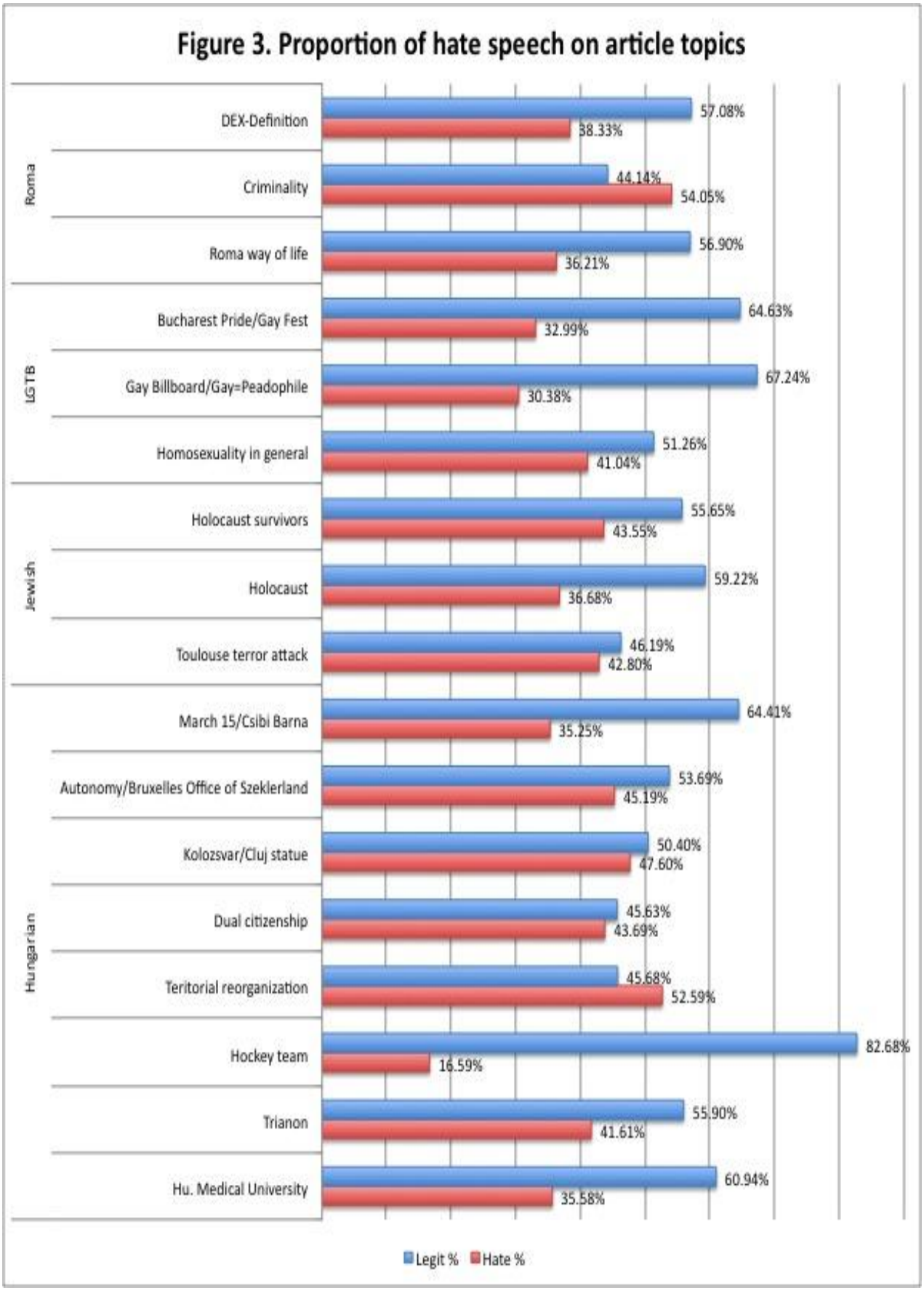
[157] The main motivation of the Hungarian party for the refusal of the reorganization was that the two counties where Hungarians are in majority would have been placed intentionally into regions where the proportion of the Hungarians would be significantly lower, thus they would be in minority all over the country. Since the president intended to go ahead with the reorganization plans despite the refusal of the DAHR, the party threatened with street demonstrations and civil disobedience in protest. At the time the Hungarian party's vote were needed to obtain majority in the parliament

[158] also cited in the Introduction

[159] gandul.info. 2011. "Tamas Sandor (DAHR) the Chief of the County Council of Covasna About the Civil Disobedience: 'In the First Phase We Will Get to the Streets Without Weapons. Than We Will See' (Tamas Sandor (UDMR), Şeful Consiliului Județean Covasna, Despre „nesupunerea Civică": „În Prima Fază, Ieşim În Stradă Fără Arme. Apoi o Să Vedem".) EXCLUSIV - Gandul." http://www.gandul.info/politica/tamas-sandor-udmr-seful-consiliului-judetean-covasna-despre-nesupunerea-civica-in-prima-faza-iesim-in-strada-fara-arme-apoi-o-sa-vedem-exclusiv-8342275. Later it was revealed that the journalist distorted the word peacefully giving it the sense without weapons

[160] the article was collected into the sample on the day it was published on 16.06.2011 when it had 348 comments and 10016 views.

‘hockey team’ topic, might be due to the fact that for an article with high comment count from adevarul.ro (usually un-moderated) the author participated in the discussion and moderated the comments himself resulting in only 8.28% of hate comments. However, this is singular case and remains an outlier.



Within the LGTB target group the articles in the topic "homosexuality in general" received 41.04% hate and 57.08% legit comments. These articles deal with the gay community in general presenting in neutral or even slightly positive tones the life of homosexual couples, 'coming out' stories, the issue of gay-marriage and adoption. Nevertheless 22.11% of the comments contained insults some of them extremely derogatory. Although the apparent purpose of the majority of the articles was to diffuse stereotypes by presenting the gay community as everyday "normal" people, the second most frequent type of comments (9.38%) repeated negative stereotypes. There was a high proportion of religious extremism (7.54%) pointing to the strong position of the Romanian Orthodox Church and its negative attitude towards sexual minorities. In fact most of the comments in this category were based on church literature or posted actual sermons of orthodox monks. The comments on the 'denying rights' category (5.36%) generally denied the right of gay couples or homosexuality to be present in public, while the "conspiracy/foreign interest/threats" category includes comments implying that homosexuality is undermining the morality of the society or that the gay community represents the interests of global conspiracy that forced Romania to grant rights to homosexuals.[161]

The article with the most comments in this topic[162] presented a lesbian couple, the coming out story of a man and some elements of the social life of the gay community. Since it was written in a positive tone it could have contributed to the diffusion of stereotypes about homosexuality.[163] Since the media visibility of sexual minorities is also law it is reasonable to suppose that homosexuals would read the few articles presenting their group in positive terms. However since 47% of the more than 300 comments contained hate, the positive message of the article was distorted. Members of sexual minority groups accessing the article

[161] Homosexuality was decriminalized in 2001 at the pressures of the European Union.
[162] evz.ro. "Sexul Pe Furiş Al Homosexualilor Din Bucureşti (The Secret Sex-life of the Homosexuals in Bucharest)." http://www.evz.ro/detalii/stiri/sexul-pe-furis-al-gay-ilor-din-bucuresti-952300/pagina-comentarii//toate-comentariile.html#comentarii.
[163] See the survey result presented in the section regarding the social context

had to face 142 hate comments, 3% of which called for their extermination 6.33% containing threats, 4% linking or identifying homosexuality with pedophilia, 10.33% religious extremism and 26.33% insults. Therefore by allowing un-moderated hate comments evz.ro effectively delivered the victims to the hate speech.

Conclusions

To assess the proportions of the phenomena and to test the efficiency of the legislation and of the site usage policies in identifying and preventing user-generated hate speech performed a comparative analysis of the participatory features, and then assembled a purposive sample of 83 articles from the sites of most important Romanian newspapers on topics regarding minorities and the respective 6031 comments. The articles were grouped on target minorities and topics that occurred during a period of 13 months from March 2011 to April 2012. A definition of 'hate' was created based on the legislation and the encyclopedic definitions, and then it was expanded into 23 hate-type categories, to provide a codebook that would allow the identification and classification of hate comments. The content analysis revealed that 37.99 percent of comments in my sample contains hate speech. The amount of hate shows relative stability along target groups being around 40 percent, suggesting that there is a relatively stable group of users who post hate comments regardless of target groups or topics. Although it was based on a purposive thus not representative sample the analysis lead to results similar to those found by surveys regarding discrimination[164] where the proportion of respondents who refused the presence of the Roma and Hungarian minority was similar to the proportion of the hate comments in the "exclusion/this is our country" category suggesting that comments reproduce the level of negative attitudes found in the society.

2.The nature and enabling factors of user-generated hate speech

In this study, I set off with the aim of identifying the loophole in media policy provisions on online press that allows for the exploitation of the newspapers' participatory spaces and comment sections as platforms to disseminate hate speech. I argue that this loophole is originated by the fact that newspaper online editions are treated under the same policy as the

[164] CNCD, *Perceptions and attitudes, 2009, 2012.*

print version, although they are different products. A good policy regarding print newspapers is not necessarily a good policy for news-sites. The hands-off, regulation-free approach to newspapers works for the print edition where all the published content has a responsible (and identifiable) author or editor who are accountable for it. However on the website, where editorial material, supposedly produced following the ethics of the profession, and UGC posted by anonymous users appear side by side, there remains a large segment of content for which it is hard if not impossible to identify an accountable author. Therefore in online publishing the regulation-free approach originally aimed at protecting press freedom becomes a weakness that can be successfully exploited by malicious users to publish user-generated hate speech.

Based on the empirical analysis presented in chapter four, I can offer an answer to the research question posed in the beginning about the nature of user-generated hate speech. User-generated hate speech in Romania is composed mainly of group-based insults, but it also contains threats, violent language and even calls for murder or genocide, reproducing in form of comments the negative attitudes towards minorities already widespread in the Romanian society. It shows a parasitizing and virus-like behavior by exploiting the "weaknesses" of the system such as the lack of a consistent regulation regarding UGC, or features designed initially to encourage citizen debate such as the possibility to participate in an online debate using only a nick- or user-name, and it exploits professional content (the news article) to attract and reach its targets and to multiply such as a virus by means of its host.

With regards to my second research question, I have identified two main factors enabling UGHS, namely the deferral of all responsibility for UGC to users and the comment management approaches enacted by the newspapers' editorial teams. In my view, this is because currently there are no statutory or self-regulated provisions regarding specially UGC on sites where it is present alongside with professional content, neither regarding the

management of user participation, leaving the sites free to choose whether to moderate comments or not, and their preferred moderation method.[165] My study shows that the amount of UGHS is much lower in proportion on sites that pre-moderate (i.e., filter) messages before posting them - much like it used to be in the traditional journalism model where editors filter the content created by their team of journalists.

The presence of UGHS in such large amounts shows the failure of the current legislative model and of regulatory authorities in applying the existing anti-discrimination regulation. These laws were transposed in the site guidelines, participation rules, and terms of service, resulting in all the sites analyzed in this research prohibiting xenophobic, discriminatory, hateful, instigating, and racist comments. However, given the presence of such large amount of content in obvious violation of their own guidelines, it seems that administrators and editors consider that they fulfill their duty in the prevention of UGHS just by announcing it in the TOS, and transferring then all the responsibility to users. At the same time, on all the sites analyzed in this study, and similarly to the findings of Ruiz *et a*l.[166] the online newspapers retain the intellectual property rights for the comments including the right to republish or otherwise gain commercial benefits, effectively exploiting the users' "free labor" as suggested by Fuchs and Schafer[167]. This results in a situation where users face all the legal consequences while the sites gain all the financial benefits. Transferring responsibility to users could also exonerate sites from any duty regarding the moderation of comments, which might explain why they tolerate hate speech to this extent.

The authorities also fail in enforcing the provisions of the existing anti-discriminatory legislation that can be applied to UGHS. As shown in the examples provided in chapter four, a large segment of UGHS (especially that in the 'Extermination/Murder/Rape", "Holocaust"

[165] pre or post moderation
[166] Ruiz et al., *Public Sphere 2.0*
[167] Fuchs, *Political Economy of Google*; Schafer, *Bastard Culture.*

and "Insult" categories) can relatively easily be recognized as hate speech under the existing legislation, -yet judging from their continued presence on newspaper websites regulatory authorities do not seem to take any measures to have them removed. The reasons behind this inefficiency are not clear, nor I have empirical data to investigate them at this stage. However the authorities' indifference ends up harming the minority groups affected by hate speech content posted online, as, given the lack of judicial support by the state, holding individual users accountable for the content they post is extremely difficult. Furthermore, the undisturbed presence of hate speech in the comments ends up being unfair towards journalists. To put it with Hlavach and Freivogel[168], there is no good explanation for the preferential treatment of a category of content creators. My view is that since both publish in the same space and share the same audience, they should be subject to the same rules.

Tolerating UGHS is also to be considered a failure on the side of the newspapers. Instead of fulfilling their role of information providers they allow their pages and content to serve as a host and a delivery platform of hate speech. The higher level of exposure provided by the host makes UGHS more dangerous. The sites in my sample are amongst the most visited Romanian sites[169], each with around 1.5 million unique visitors per month. As I have shown in the analysis of their participatory features, the access to these mass-audiences trough the comment sections is open, without effective control, to everyone with sufficient knowledge of the syntax of an email address, providing a readily assembled audience to hate speech which would be hardly accessible in any other way, thus contributing to the large scale dissemination of discriminatory views. In absence of the exposure guaranteed by newspaper websites, hate content would not disappear but it would remain marginalized on extremist websites, hate-blogs, hate-forums or their offline counterparts. Moreover probably it would not be accessed by as many people in the target groups as it happens with online newspapers:

[168] Hlavach and Freivogel, *Ethical Implications of Anonymous Comments.*
[169] see Table 1.

on the sites in this sample, by appearing alongside with articles about minorities, UGHS is able to reach people belonging to minorities who presumably read in larger number articles about their community.

The analysis of the commenting interface revealed that with the exception of three of the five sites allow comments to be published instantly relying only on poorly configured profanity filters to prevent the posting of offensive or obscene words. The results of the content analysis indicate that the largest type of hate speech are insults (18.08 percent of all comments) although this would be the easiest type to prevent, only by regularly updating said profanity filter.

2. Preventing user-generated hate speech

Contrary to countries like the United States, where the First Amendment prevents the regulation of hate speech, in Romania the law prohibits quite clearly the discriminatory behaviors generally associated with hates speech. This is visible, for example, in the fact that all the sites analyzed in this study transposed the legislation into their guidelines prohibiting that type of messages, which however does not prevent online hate speech from happening. Therefore the question is not whether hateful comments on websites should be filtered, rather why are they still there despite the legislation and what could be done to prevent the phenomena.

In my view there could be three solutions to user-generated hate speech: a) the separation of the readers' comments from the main page; b) amendments to the legislation to make newspapers responsible for their electronic pages, and c) self-regulation by newspapers. The last would be the optimal solution as it would prevent state interference into the media system.

The legislative option would require the government to elaborate a special law on online media (where it would clarify the question of responsibility for user-generated content on mainstream sites differentiated from dedicated UGC sites[170]), or the participation management requirements for the service providers. The modification of the existing legislation could add content monitoring function with regards to discrimination to the CNCD also extended to websites, which would allow for the institution to issue take down notices for UGHS - but this second solution would not prevent the initial posting of such messages. Neither of these solutions is desirable, as they would increase the potential for state interference and abuse in the media, especially in presence of authoritarian government. The simplest statutory solution would be a rigorous enforcement of the legislation already in place, and the prosecution of hate speech cases in comment sections, which combined could act as a deterrent. However, this last solution could have a chilling effect on legitimate comments, and it is again open for state abuses.

In my opinion the optimal solution would be self-regulation, a public trust approach as described by McQuail.[171] Media organizations could agree to an ethics code for audience participation which would also contain good practice recommendations regarding moderation and a commitment by the participating newspapers/sites towards moderating comments, based on a commonly agreed set of guidelines similar to the codebook in appendix 2. As it happens with the codebook, such guidelines could also incorporate elements of the legislation, which in the current approach is totally ineffective. Having a document similar to the codebook would ensure that the terms “discriminatory”, “instigating”, “racist”, and “hateful” already present despite being prohibited in the guidelines of the sites finally have a shared meaning. The codebook itself could also be published in the TOS to make users aware of the actual meaning and extent of the prohibited categories.

[170] such as the Youtube video sharing platform
[171] McQuail, *Accountability of Media*

A major obstacle in implementing such a solution is the lack of a joint self-regulatory body in Romania and the financial costs of moderation, which imposes a burden on media companies. On the one hand the moderation itself is costly, but there is also the indirect cost of potentially reduced advertisement revenues. As the content analysis shows, 37.99% of comments contain hate speech: eliminating them would cut in almost half the levels of user participation. Additionally if the hate comments would not be published the comments reacting to them would also disappear, resulting not only in reduced participation but also in reduced advertising revenues as visitors on the site would spend less time reading comments or reacting to them. On the other hand, taking the example of the article with the highest proportion of UGHS (78.64 percent), the 674 comments currently posted on the site (even disregarding multiple contributions) represent still an insignificant fraction of the 28514 views of the article, who presumably went there to read the editorial and not the user-generated content.

On the other hand the unregulated and unrestricted nature of the comments creates a responsibility-free space on the online newspaper which is not present in its offline counterpart, and that can also be used for manipulating the public's perception about the issues presented in the articles or even for the intimidation of a given group – something media are not able to do in the offline or even in the professional areas of the online world due to the threat of possible legal actions. Media organizations and journalists both offline and online are constrained by professional guidelines, ethical rules, laws and other similar formal or informal regulations regarding content; breaching these has legal and moral consequences on their professional reputation. On the other hand, as the data presented earlier shows, there are no such limits in the comments. Views that cannot be published due to societal norms, laws or ethics in an article can be published in the comment sections of an article on the same topic, while still benefiting from the same audience. In many cases such as

the two topics with the highest amount of hate-speech: the territorial reorganization and Roma criminality, the amount of hate comments and their nature is evident at a first glance. It is unlikely that the administrators of the sites were not aware of having such content displayed on their pages. Therefore, one has to wonder about their motivations for allowing it even in clear infringement of their own guidelines. Returning to the article presented at the beginning in the thesis the fact that the site used an inflammatory title and allowed such high amount of extremely hateful and violent comments might also be intended as an intimidation of the minority.

4. Directions for further research

As shown in the previous section, a major question regarding user-generated hate speech in Romania is the motivations of newspapers and authorities for tolerating it. Based on the findings in this thesis a next step would be to try to expore the media organizations' motivations for choosing a particular moderation method, and for keeping hate speech content accessible in evident violations of their own guidelines. The role attributed by Romanian journalists to comments is also worth exploring. An equally important question to be answered is the effect user-generated hate speech has on the target groups. As it was mentioned the findings of the content analysis show similarities to survey results about discrimination, therefore an important issue to clarify would be if the large proportions of hate comments have a role in reproducing or maintaining negative attitudes towards minorities.

References:

"Ana Birchall wins lawsuit vs. Iosif Buble", June 6, 2011. http://www.nineoclock.ro/ana-birchall-wins-lawsuit-vs-iosif-buble/ (accessed June 6, 2011).

"SATI - Studiul de Audienta si Trafic Internet", n.d. http://www.sati.ro/index.php?page=rezultate_site&o=name&sort=ASC&siteid=nespecificat&letter=toate&filter_type_period=1&filter_period=2011-03-01&filter_category=3#nespecificat (accessed April 29, 2012).

ActiveWatch Media Monitoring Agency, Centrul Pentru Jurnalism Independent (Center for Independent Journalism - Romania), and IMAS Public opinion resarch agency. *Autoreglementarea presei in Romania - Self regulation of the press in Romania*. Survey. ActiveWatch-Media Monitoring Agency (Romania), October 2009. www.activewatch.ro/uploads/FreeEx%20Publicatii%20/Autoreglementarea%20presei%20din%20Romania.pdf (accessed October 1, 2012).

———. *Media Freedom in Romania 2009*. ActiveWatch-Media Monitoring Agency (Romania), 2009. http://www.activewatch.ro/uploads/FreeEx%20Publicatii%20/FreeEx%20Report%20-%20May%203%202010.pdf.

Adler, Amy. "What's Left?: Hate Speech, Pornography, and the Problem for Artistic Expression." *California Law Review* 84, no. 6 (December 1, 1996): 1499-1572.

Barendt, E. *Freedom of speech*. 2nd ed. Oxford;;New York: Oxford University Press, 2007.

Benkler, Yochai. *The wealth of networks how social production transforms markets and freedom*. New Haven [Conn.]:: Yale University Press,, 2006.

Berg, Bruce. *Qualitative research methods for the social sciences*. 4th ed. Boston: Allyn and Bacon, 2001.

Biegel, Stuart. *Beyond our control?* Cambridge (Mass.); London: the MIT press, 2001.

Braman, Sandra. *Change of state*. Cambridge (Mass.);;London: the MIT press, 2006.

BRAT - Romanian Bureau of Circulation Audit. "Circulation number for nationwide daily newspapers (cotidian generalist national) for the period march 2011 - march 2012", n.d. http://www.brat.ro/index.php?page=compare (accessed April 29, 2012).

———. "Circulation numbers for newspapers in the sample", n.d. http://www.brat.ro/index.php?page=compare (accessed April 29, 2012).

Brown-Sica, Margaret, and Jeffrey Beall. "Library 2.0 and the Problem of Hate Speech." *Electronic Journal of Academic and Special Librarianship* v.9 no.2, no. Summer 2008 (2008). http://southernlibrarianship.icaap.org/content/v09n02/brown-sica_m01.html.

Butler, Judith. *Excitable speech: a politics of the performative*. New York: Routledge, 1997.

Cammaerts, Bart. "Radical pluralism and free speech in online public spaces." *International Journal of Cultural Studies* 12, no. 6 (November 1, 2009): 555 -575.

Couture, Barbara, and Ebooks Corporation. *Private, the Public, and the Published: Reconciling Private Lives and Public Rhetoric*. Logan:: Utah State University Press,, 2004.

Couture, Barbara. "Reconciling Private Lives and Public Rhetoric: What's at Stake?" In *Private, the Public, and the Published: Reconciling Private Lives and Public Rhetoric.*, 1-30. Logan:: Utah State University Press,, 2004.

Delgado, Richard, and Jean Stefancic. *Must we defend Nazis?: hate speech, pornography, and the new First Amendment*. New York: New York University Press, 1997.

Deuze, Mark. “Internet News”. Malden MA: Blackwell Pub., 2008. in Donsbach, Wolfgang., ed. *The international encyclopedia of communication.* Malden MA: Blackwell Pub., 2008.

———. “The Web and its Journalisms: Considering the Consequences of Different Types of Newsmedia Online.” *New Media & Society* 5, no. 2 (June 2003): 203-230.

Duffy, Margaret E. “Web of Hate: a Fantasy Theme Analysis of the Rhetorical Vision of Hate Groups Online.” *Journal of Communication Inquiry* 27, no. 3 (July 1, 2003): 291 -312.

Dwyer, Tim. *Media convergence*. Maidenhead;;New York: McGraw Hill/Open University Press, 2010.

FreeEx, ActiveWatch Media Monitoring Agency, and Reporters Without Borders (RSF). *Media Freedom in Romania 2010*. ActiveWatch-Media Monitoring Agency (Romania), 2011. http://www.activewatch.ro/uploads/FreeEx%20Publicatii%20/FreeEx%20Report%20-%20May%203%202010.pdf.

FreeEx, and ActiveWatch Media Monitoring Agency. *Media Freedom in Romania 2008*. ActiveWatch-Media Monitoring Agency (Romania), 2009. http://www.activewatch.ro/uploads/FreeEx%20Publicatii%20/Freeexeng_2008_dtp.pdf.

Gaiser, Ted J, and Anthony E Schreiner. *A guide to conducting online research*. Los Angeles; London: SAGE, 2009.

Goss, Brian Michael. “ONLINE ‘LOONEY TUNES’.” *Journalism Studies* 8, no. 3 (June 2007): 365-381.

Gross, Peter, and Mihai Coman. *Media and journalism in Romania*. Berlin: Vistas, 2006.

Gross, Peter. *Mass media in revolution and national development: the Romanian laboratory*. Ames (Iowa): Iowa state university press, 1996.

Hallin, Daniel C, and Paolo Mancini. *Comparing media systems: three models of media and politics*. Cambridge, U.K.: Cambridge University Press, 2004.

Hine, Christine. *Virtual methods: issues in social research on the Internet*. Oxford; New York: Berg, 2005.

Hlavach, Laura, and William Freivogel. “Ethical Implications of Anonymous Comments Posted to Online News Stories.” *Journal of Mass Media Ethics* 26, no. 1 (January 2011): 21-37.

INSOMAR. *Fenomenul discriminarii in Romania - perceptii si atitudini" in anul 2009 - Discrimination in Romania-perceptions and attitudes in 2009*. CNCD - National Anti-Discrimination Council, Romania, 2009. http://www.cncd.org.ro/files/file/Fenomenul%20discriminarii%202009.pdf.

K.U.Leuven – ICRI (lead contractor) Jönköping International Business School - MMTC Central European University - CMCS Ernst & Young Consultancy Belgium. “Country reports - Study on Indicators for Media Pluralism - Media Task Force | Europa - Information Society and Media.” *Independent Study on Indicators for Media Pluralism in the Member States - Towards a Risk-Based Approach.*, n.d. http://ec.europa.eu/information_society/media_taskforce/pluralism/study/country_rep/index_en.htm (accessed November 15, 2011).

Kaufer, David S. “The Influence of Expanded Access to Mass Communication on Public Expression: The Rise of Representatives of the Personal.” In *Private, the Public, and the Published: Reconciling Private Lives and Public Rhetoric.*, 153-165. Logan: Utah State University Press,, 2004.

Kinney, Terry A. “Hate Speech and Ethnophaulisms”. Malden MA: Blackwell Pub., 2008.

Krippendorff, Klaus. *Content analysis: an introduction to its methodology*. 2nd ed. Thousand Oaks Calif.: Sage, 2004.

Lee, E.-J., and Yoon Jae Jang. “What Do Others’ Reactions to News on Internet Portal Sites Tell Us? Effects of Presentation Format and Readers’ Need for Cognition on Reality Perception.” *Communication Research* 37, no. 6 (July 2010): 825-846.

McQuail, Dennis. “Accountability of Media to Society: Principles and Means.” In *Communication Theory & Research*, 89-102. Sage, 2005.

Neuendorf, Kimberly. *The content analysis guidebook*. Thousand Oaks Calif.: Sage Publications, 2002.

Neumann, Russel W. "Interactivity, Concept of". Malden MA: Blackwell Pub., 2008.

Organisation for Economic Co-operation and Developement. "PARTICIPATIVE WEB: USER-CREATED CONTENT", n.d.

Rebillard, F., and A. Touboul. "Promises unfulfilled? 'Journalism 2.0', user participation and editorial policy on newspaper websites." *Media, Culture & Society* 32, no. 2 (March 2010): 323-334.

Reich, Zvi. "User Comments: The transformation of participatory space." edited by Jane B Singer, n.d.

Ritchie, Jane, and Jane Lewis. *Qualitative research practice : a guide for social science students and researchers*. London; Thousand Oaks, Calif.: Sage Publications, 2003.

Robinson, Sue. "Traditionalists vs. Convergers." *Convergence: The International Journal of Research into New Media Technologies* 16, no. 1 (February 1, 2010): 125 -143.

Rorive, Isabelle. "What Can Be Done Against Cyber Hate? Freedom of Speech Versus Hate Speech in the Council of Europe." *Cardozo Journal of International & Comparative Law* 17, no. 3 (October 2009): 417-426.

Ruiz, Carlos, David Domingo, Josep Lluís Micó, Javier Díaz-Noci, Koldo Meso, and Pere Masip. "Public Sphere 2.0? The Democratic Qualities of Citizen Debates in Online Newspapers." *The International Journal of Press/Politics* 16, no. 4 (October 1, 2011): 463 -487.

Sanders, Teela. "Researching the Online Sex Work Community." In *Virtual methods : issues in social research on the Internet*, edited by Christine Hine, pp. 67-80. Oxford; New York: Berg, 2005.

Schäfer, Mirko. *Bastard culture! : how user participation transforms cultural production*. Amsterdam: Amsterdam University Press, 2011.

Singer, Jane B, David Domingo, Ari Heinonen, Alfred Hermida, Steve Paulussen, Thorsten Quandt, Zvi Reich, and Marina Vujnovic. *Participatory Journalism in Online Newspapers : Guarding the Internet's Open Gates*. Boston [u.a.]: Wiley-Blackwell, 2011.

Sorial, Sarah. "Free Speech, Autonomy, and the Marketplace of Ideas." *The Journal of Value Inquiry* 44, no. 2 (January 2010): 167-183.

Stake, Robert. "Qualitative Case Studies." In *Sage Handbook of Qualitative Research*, 443, 467. 3rd ed. Sage Publications, 2005.

Sunstein, Cass. *Going to extremes : how like minds unite and divide*. Oxford ;;New York: Oxford University Press, 2009.

Valcke, Peggy, and Marieke Lenaerts. "Who's author, editor and publisher in user-generated content? Applying traditional media concepts to UGC providers." *International Review of Law, Computers & Technology* 24, no. 1 (March 2010): 119-131.

van Dijck, José. "Users like you? Theorizing agency in user-generated content." *Media, Culture & Society* 31, no. 1 (January 1, 2009): 41 -58.

van Dijck, Teun A. *Ideology and discourse A Multidisciplinary Introduction*. Pompeu Fabra University, Barcelona, n.d.

Annexes

Appendix 1. Minority related issues in the Romanian press

Target	Minority Related Topics	Total Articles	Total Comments
Hungarians	March 15. 2011 – Hungarian national holiday./ The hanging of a puppet representing a Romanian national hero by a Hungarian extremist	4	295
	Autonomy/ the opening of a representation office in Bruxelles, for Szeklerland the region of Romania with Hungarian majority population and for which the Hungarian population seeks automomy	6	447
	Territorial reorganization – a proposition of the government to reorganize Romania in larger administrative units, resulting in changes in Hungarians loosing the majority status in the two counties where they are the majority	8	753
	Hungarian Medical University – The creation at the Medical University of Targu Mures/Marosvasarhely of a faculty of medicine in Hungarian language (February-March 2012)	7	489
	Hockey team – In December 2011 on an official hockey match between Romania and Hungary, the Romanian national hockey team composed entirely of hungarians, sang along the anthem of Hungary.	8	820
	Trianon – Events remembering the treaty that awarded Transsylvania to Romania in 1920.	4	483
	Dual citizenship – Hungarians in Romania asking for the Hungarian citizenship	1	103
	Kolozsvar/Cluj statue – Protest of Hungarian leaders for the unauthorized placement of a plaque with controversial conternt on the statue of a Hungarian king in the city of Kolozsvar/Cluj	3	250
LGTB	The Bucharest Pride/Gay Fest march in 2011	5	294
	Gay Billboard - An LGTB rights activist organization placed billboards in several cities with the image of a newborn wearing a wristband with the word 'homosexual' to illustrate that homosexuality is not a choice.	4	293
	Homosexuality in general – Interviews about the life/coming out of homosexuals. Stories about violence against homosexuals.	8	597
Jewish	Holocaust survivors – Interviews about the holocaust	4	124
	Holocaust – Articles about Romania's role in the	10	488

	holocaust/ Holocaust denial by politicians		
	Toulouse terror attack – Articles about the attack against a jewish school in Toulouse in march 2012	4	236
Roma	DEX-Definition – Changes of the official academic definitions in the Dictionary of the Romanian Language (DEX), for the words, 'Roma' (Rrom), Gipsy (Tigan), Homosexual	3	240
	Criminality – Association of Roma persons with criminality	2	111
	Roma way of life	2	58

Appendix: 2. Coding protocol and codebook for user generated hate speech

I. Coding protocol and codebook for user generated hate speech

Before starting coding please read the following pieces of Romanian legislation and definitions that form the basis of the codes in this codebook.

Legislation:

Constitution of Romania:

Art. 4.2

"Romania is the common and indivisible homeland of all its citizens, without any discrimination on account of race, nationality, ethnic origin, language, religion, sex, opinion, political adherence, property or social origin."

Art.6.1

The State recognizes and guarantees the right of persons belonging to national minorities to the preservation, development and expression of their ethnic, cultural, linguistic and religious identity.

Government Ordinance nr. 137/31 August, 2000 – Prohibiting discrimination of any kind

The principle of equality of citizens and the exclusion of privileges and discrimination are guaranteed especially in exercising the following rights:

b) the right to safety and protection by the state against any violence by any individual, group, or institution.

c) political rights, namely electoral rights, the right to participate at the public life and to have access to public offices

(2) Principiul egalităţii între cetăteni, al excluderii privilegiilor şi discriminarii sunt garantate în special în exercitarea următoarelor drepturi:

b) dreptul la securitatea persoanei şi la obtinerea protectiei statului impotriva violentelor sau maltratarilor din partea oricărui individ, grup sau institutie;

c) drepturile politice, şi anume drepturile electorale, dreptul de a participa la viata publica şi de a avea acces la functii şi demnitati publice;

ART. 2

“According to the present ordinance discrimination is considered to be any differentiation, exclusion, restriction or preference based on race, nationality, ethnicity, language, religion, social category, sex, sexual orientation, age, handicap, not contagious chronic disease, HIV infection, appartenence to a defavorized category and any other criteria, that is aimed or has the effect of restricting, limiting recognition, use or exercise in conditions of equality, of human rights, and of fundamental freedoms, or of rights recognized by law, in the political, economic, social and cultural and any other domains of the public life”

“(1) Potrivit prezentei ordonante, prin discriminare se întelege orice deosebire, excludere, restrictie sau preferinta, pe baza de rasa, nationalitate, etnie, limba, religie, categorie socială, convingeri, sex, orientare sexuală, varsta, handicap, boala cronica necontagioasa, infectare HIV, apartenenta la o categorie defavorizată, precum şi orice alt criteriu care are ca scop sau efect restrangerea, înlăturarea recunoaşterii, folosintei sau exercitării, în conditii de egalitate, a drepturilor omului şi a libertătilor fundamentale sau a drepturilor recunoscute de lege, în domeniul politic, economic, social şi cultural sau în orice alte domenii ale vietii publice.”

Art 2.5

Constitutes harassment and is penalized any act based on criteria of race, nationality, ethnicity, language, religion, social category, convictions, gender, sexual orientation, appartenance to a defavorized category, age, handicap, refugee or asylum seeking status, or any other criteria that creates an intimidating, hostile, degrading or offensive environment.

(5) Constituie hartuire şi se sanctionează contraventional orice comportament pe criteriu de rasa, nationalitate, etnie, limba, religie, categorie socială, convingeri, gen, orientare sexuală, apartenenta la o categorie defavorizată, varsta, handicap, statut de refugiat ori azilant sau orice alt criteriu care duce la crearea unui cadru intimidant, ostil, degradant ori ofensiv.

Art 15.

“It is considered a contravention any public behavior that has the character of nationalist-chauvinist propaganda, or any behavior that has as purpose of creating an intimidating, degrading, hostile, humiliating or offensive atmosphere against, or harms the dignity of a person, group, community in connection with their race, nationality, ethnicity, religion, social category, conviction or sexual orientation.”

“Constituie contraventie, conform prezentei ordonante, dacă fapta nu intra sub incidenta legii penale, orice comportament manifestat în public, având caracter de propaganda nationalist-şovină, de instigare la ura rasială sau nationala, ori acel comportament care are ca scop sau vizează atingerea demnitătii ori crearea unei atmosfere de intimidare, ostile, degradante, umilitoare sau ofensatoare, îndreptat impotriva unei persoane, unui grup de persoane sau unei comunităti şi legat de apartenenta acestora la o anumită rasa, nationalitate, etnie, religie, categorie socială sau la o categorie defavorizată ori de convingerile, sexul sau orientarea sexuală a acestuia.”

Definitions:

"*hate speech* ---the use of words as weapons that terrorize, humiliate, degrade, abuse, threaten, and discriminate based on race, ethnicity, religion, sexual orientation, national origin, or gender" *(Encyclopedia of Political communication, 2007:301)*

"Obscene, defamatory, slanderous, or hateful, speech that holds a reasonable potential to be harmful" (Lederer & Delgado 1995).

"a form of verbal aggression that expresses hatred, contempt, ridicule, or threats toward a specific group or class of people" (Asante 1998).

"Verbalizations, written messages, symbols, or symbolic acts that demean and degrade, and, as such, can promote discrimination, prejudice, and violence toward targeted groups."

"Hate speech functions to distort the history of targeted groups, to eliminate the agency of targeted groups, to create and maintain derogatory cultural, racial, and ethnic illusions about targeted groups, and as a vehicle for expressing pejoratives" (Asante 1998).

(Hate speech and ethnophaulism - in International encyclopedia of Communication, 2007:2051)

Based on the above legislation and the encyclopedic definitions for the purpose of this codebook hate speech is defined as:

Comments containing speech aimed to terrorize, humiliate, degrade, abuse, threaten, ridicule, demean, and discriminate based on race, ethnicity, religion, sexual orientation, national origin, or gender (Encyclopedia of Political communication, 2007:301) Expressing prejudice, and contempt, promoting or supporting discrimination, prejudice and violence. Seeking to distort the history of targeted groups, to eliminate their agency, to create and maintain derogatory cultural, racial, and ethnic illusions about targeted groups . Also including pejoratives and group based insults, that sometimes comprise brief group epithets consisting of short, usually negative labels or lengthy narratives about an outgroup's alleged negative behavior. (International encyclopedia of Communication:2051). Discrimination is considered to be **any differentiation, exclusion, restriction or preference** based on group appartenance and any other criteria, that is aimed or has the effect of restricting, limiting recognition, use or exercise in conditions of equality, of human rights, and of fundamental freedoms, or of rights recognized by law, in the political, economic, social and cultural and any other domains of the public life (Art. 2 of OUG 137/31 Aug. 2000)

Coding frame:

Comments will be coded on two levels. The first level codes are 'hate', and 'non-hate' comments can be coded in one of these codes. The second level codes refer to types of 'hate' and multiple codes can be assigned to one comment with the exception of 'legit' that cannot be assigned to comments that have any other sub-codes. Non-hate comments that should not have been allowed according to the terms and conditions or terms of use of the sites will be coded with 'insult', 'violence', 'junk/spam' – all other comments that have not been assigned a code from the this group will be automatically assigned by the software the code 'legit' i.e. to legitimate discussion. 'Hate' refers to comments targeted to members or groups/communities, while 'insult', 'violence' 'profanity" in the non-hate group refer to comments targeted at individuals without making reference to their group appertenance. 'junk/spam' – refers to comments that have no content or contain advertisements, or other similar content. The nicknames/usernames of the users and the subject lines are also considered as being part of the comment. In the sub-codes 'group A' refers to in-groups while 'group B, C, D' to out-groups.

II. Codebook for user generated hate speech

Hate

Comments containing speech aimed to terrorize, humiliate, degrade, abuse, threaten, ridicule, demean, and discriminate based on race, ethnicity, religion, sexual orientation, national origin, or gender (Encyclopedia of Political communication, 2007:301) Expressing prejudice, and contempt, promoting or supporting discrimination, prejudice and violence. Seeking to distort the history of targeted groups, to eliminate their agency, to create and maintain derogatory cultural, racial, and ethnic illusions about targeted groups . Also including pejoratives and group based insults, that sometimes comprise brief group epithets consisting of short, usually negative labels or lengthy narratives about an outgroup's alleged negative behavior. (International encyclopedia of Communication 2007:2051)

This is a top level code. Please assign it to comments that contain any of the elements of the above definition. After you coded the comment with the top-code 'hate' you may choose additional sub-codes referring to the type of hate speech in the comment. You may also choose a sub-code first in this case the comment will also be automatically coded with the top level code.

'Hate' type sub codes

Insults	Comments that contain insults/ derogatory epithets/labels based on or referring to group appartenance. Examples: bozgor, boaghen, sogor, huni, (Hungarians); ciora, cioroi (Roma), homolau, curist, gaozar (Homosexuals), Jidani, Jidraci (Jews); Valahi, Rromania, mitici, soldoveni (Romanians), Papisti (Catholics). Judgement should be used in the case of 'valah' when it is used in referring to history i.e. The Country of Valahia, or Supplex libellul Valahorum, and in case of \'mitici\' when it is used auto-ironically
Violence	Comments that make open threats or calls to violence against members of communities also including comments that advocate for violent actions against members of communities. Please also add the code 'extermination/murder/rape for comments with extreme violence for example that call/advocate/threaten with murder of a minority group or persons belonging to that group. Also add the extreme label for comments calling or suggesting the rape, torture of people belonging to group B.
Extermination/Murder/Rape	Comments with extreme violence that call/advocate for the extermination, murder of a minority group or persons belonging to that group. Example: 'The best solution would be to get rid/hang all of group B.'
Threats	Comments that contain implied threats, without explicit violence if members of group B do not modify, their behavior, or abandon their claims for rights. Example: "You should stop what your group is doing or else…". "We tolerated your behavior/claim/existence but our patience is coming to an end", 'You should not provoke us because....'
Superiority/Inferiority/Normality	Comments that claim that group A or (people belonging to group A) is superior according to some criteria (ethnicity/language/race/religion/sexual orientation/gender) to group B, or that the group A is what is considered to be normal, thus superior. Also including comments that argue that group B or persons belonging to group B have no rights, or some of their rights should be limited due to their inferiority. Comments that claim that the inferior group should submit to the will/adopt some of the characteristics (language/religion/sexual orientation) of the superior group due to its superiority. Comments that argue for the preferential treatment of the

	superior group
Stereotypes/Generalization/Prejudice	Comments arguing that just by being member of group B or all the members of group B have certain negative characteristics/behaviors, that are despicable, or anti-social and would justify their discrimination, certain actions against them or invalidate their claims for certain rights or for equal treatment. Example: “We should be suspicious of group B because it consists of separatists who want to dismember the country.” “All of group B are criminals.” “All members of group B hate/despise us.” “Members of group B are incapable of living in our country/society.” “People of group B are immoral who will corrupt our (A) youth”. “Group B has some despicable customs/traditions that threaten our society”. “Group B is not to be trusted because of characteristic X.”, “Group B are thieves so they should be sterilized” \'All B-s are terrorists\'
Exclusion/This is our country	Comments that claim that the majority group is the rightful “owner” of the country and therefore: invalidate claims for rights of group B based on the argument that the country belongs to a group A therefore group B has no legitimacy to ask for rights/exist/keep its customs or traditions on the territory of the country. Also comments implying that members of group A have a tolerated status/are guests/ have less grounds for claims because the country belongs to group A or because the majority of the country is in group A. Comments that call for the expulsion of group B based on the argument that the country belongs to group A. Examples: “this is our country if you don’t like it you are free to go to …. “ “This is our country so you should be do whatever we want you to do”, “This is our country so you have no right to ask for X here”, “This is our country so you should not keep your language/customs/traditions/sexual orientation”
Animals/Sub-human	Comments that compare or call the members of a group to animals/pests, similar to animals/pests or sub-human
Holocaust - blame shifting	Comments that shift the blame for the holocaust on the victims. Examples: “The jews have themselves to blame for the holocaust”. “The jews deserved what happened to them”. The Jews brought communism to Romania so they deserved what happened to them.
Homosexuality-Pedophilia	Comments that argument explicitly or implicitly that homosexuality is related to, leads to pedophilia or that homosexual people have pedophile tendencies or are pedophiles.
History	Comments that disqualify the claims for rights or justify the discrimination or mistreatment of people belonging to group B, based on acts or injustices allegedly done by members of that group to group A along the history. Comments that call for actions against a minority based on historical arguments.
Religious extremism	Comments that threaten or call for action against or for limiting civil (secular) rights of group B, insult demean, or express contempt for group B based on religious arguments.
Conspiracy/Foreign interests/Enemies/Threat	Comments that imply that members of a group B are part of conspiracy against the country/society, serve or some foreign or malicious interests. Comments that imply that by being member of a group or seeking rights for that group, its members or leaders are enemies of the state/people/society, or that they are a threat.
Denying rights (political/civil)	Comments that dispute or deny civil or political rights of members of minority groups including rights for political representation/political

	activity, right to demonstrate, right to appear or speak in public on the ground that they are a minority or belong to group B. Call for group based actions to prevent the access to rights. Example: “Let\'s all true A get out to vote so not to allow the B-s to get into the parliament” “B-s should not be allowed to appear/speak in public”, “B-s have no right to have political represenation/education”
Expulsion	Comments that explicitly call for the expulsion of a group from the territory, with or without specific reasons or arguments for that action Ex. \'Out with B-s from the country\' “We should get rid of B’s” “All B-s should be deported”
Holocaust-denial/minimization	Comments that seek minimize the role of Romania in the holocaust, or claim that there was no holocaust in Romania. Examples: “We had no part in the Holocaust” “There was no holocaust in Romania”, “The leaders of the time are true heroes” (Criminal offense according to the Penal Code)
Holocaust-appologetic/justifications	Comments that seek to present persons involved in the holocaust as heroes or find justifications for their actions.
Disgrace for the country	Comments that argue that group B is a disgrace for the country or it is to blame for the bad image of the country.
Discrimination	Comments that call/advocate for discrimination - Discrimination is considered to be any differentiation, exclusion, restriction or preference based on group appartenance and any other criteria, that is aimed or has the effect of restricting, limiting recognition, use or exercise in conditions of equality, of human rights, and of fundamental freedoms, or of rights recognized by law, in the political, economic, social and cultural and any other domains of the public life (Art. 2 of OUG 137/31 Aug. 2000)
General hate/Discrimination	Comments with discriminatory content which does not fit into any of the above categories - Comments containing speech aimed to terrorize, humiliate, degrade, abuse, threaten, ridicule, demean, and discriminate based on race, ethnicity, religion, sexual orientation, national origin, or gender (Encyclopedia of Political communication, 2007:301) Expressing prejudice, and contempt, promoting or supporting discrimination, prejudice and violence. Seeking to distort the history of targeted groups, to eliminate their agency, to create and maintain derogatory cultural, racial, and ethnic illusions about targeted groups . Also including pejoratives and group based insults, that sometimes comprise brief group epithets consisting of short, usually negative labels or lengthy narratives about an outgroup’s alleged negative behavior. (International encyclopedia of Communication:2051).

Non-hate

Insult/Profanity	Comments containing direct personal insults/derogatory epithets addressed to individuals or the author of the article not based on group appartenance. Including non violent profanity, vulgarity
Threat/Violence	Comments containing direct personal threats addressed to individuals not based on group appartenance.
Thrash/Spam	Comments which have no textual content, have no argument, or text relating to the topic of the article or to the newspaper. Usually contain

	advertisments.
Legit	All non-coded comments will be automatically coded by the software as 'legit' – i.e. legitimate comments that respect the ethical guidelines of the site and the legislation

Appendix 3. Results of the content analysis

Table 3.1. Proportion of Hate / Legit in the entire sample

	Comments	Percent
Hate	2324	37.99%
Legit	3597	58.80%

Table 3.2 Codes report

HATE	Comments	Percent	Percent of Hate
Insults	1106	18.08%	47.59%
Stereotypes/Generalization/Prejudice	522	8.53%	22.46%
Conspiracy/Foreign interests/Enemies/Threat	397	6.49%	17.08%
Exclusion/This is our country	341	5.57%	14.67%
Extermination/Murder/Rape	245	4.01%	10.54%
Superiority/Inferiority/Normality	194	3.17%	8.35%
Denying rights (political/civil)	186	3.04%	8.00%
Expulsion	165	2.70%	7.10%
History	158	2.58%	6.80%
Threats	148	2.42%	6.37%
Violence	141	2.31%	6.07%
Animals/Sub-human	120	1.96%	5.16%
Religious extremism	88	1.44%	3.79%
Holocaust-denial/minimization	86	1.41%	3.70%
Holocaust - blame shifting	81	1.32%	3.49%
Discrimination	73	1.19%	3.14%
Holocaust, Fascism - appology/justifications	71	1.16%	3.06%

HATE	Comments	Percent	Percent of Hate
Moderated	59	0.96%	2.54%
General hate/Discrimination	58	0.95%	2.50%
Homosexuality-Pedophilia	50	0.82%	2.15%
Disgrace for the country	25	0.41%	1.08%
Hate-Spam	18	0.29%	0.77%
Sterilization	5	0.08%	0.22%
NON-HATE	**Comments**	**Percent**	**Percent of Non-hate**
Legit	3597	58.80%	94.76%
Insult/Profanity	143	2.34%	3.77%
Thrash/Spam	44	0.72%	1.16%
Threat/Violence	8	0.13%	0.21%
HS target responding	6	0.10%	0.16%

Table 3.3 Proportion of hate comments along the sites

3.3a. Newspapers hate/legit

Newspapers	Articles	Comments	Hate Count	Hate Percent	Legit Count	Legit Percent
Gandul	16	1524	736	48.29%	731	47.97%
Evz	21	1250	543	43.44%	621	49.68%
Hotnews	10	746	190	25.47%	555	74.40%
Adevarul	24	2004	649	32.39%	1309	65.32%
Romania Libera	13	593	206	34.74%	381	64.25%

3.3b. Hate speech types

HATE		**ALL**	**Gandul**	**EVZ**	**Hotnews**	**Adevarul**	**Romania Libera**
Insults	**Count**	1106	414	269	39	333	51
	% of All	18.08%	27.17%	21.52%	5.23%	16.62%	8.60%
	% of Hate	47.59%	56.25%	49.54%	20.53%	51.31%	24.76%
Hate-Spam	**Count**	18	4	10	1	3	0
	% of All	0.29%	0.26%	0.80%	0.13%	0.15%	-
	% of Hate	0.77%	0.54%	1.84%	0.53%	0.46%	-
Moderated	**Count**	59	4	0	0	52	3
	% of All	0.96%	0.26%	-	-	2.59%	0.51%
	% of Hate	2.54%	0.54%	-	-	8.01%	1.46%
Sterilization	**Count**	5	1	3	0	1	0
	% of All	0.08%	0.07%	0.24%	-	0.05%	-
	% of Hate	0.22%	0.14%	0.55%	-	0.15%	-

HATE		ALL	Gandul	EVZ	Hotnews	Adevarul	Romania Libera
Violence	**Count**	141	76	29	7	27	2
	% of All	2.31%	4.99%	2.32%	0.94%	1.35%	0.34%
	% of Hate	6.07%	10.33%	5.34%	3.68%	4.16%	0.97%
Extermination/Murder/Rape	**Count**	245	114	52	4	69	6
	% of All	4.01%	7.48%	4.16%	0.54%	3.44%	1.01%
	% of Hate	10.54%	15.49%	9.58%	2.11%	10.63%	2.91%
Threats	**Count**	148	70	32	13	28	5
	% of All	2.42%	4.59%	2.56%	1.74%	1.40%	0.84%
	% of Hate	6.37%	9.51%	5.89%	6.84%	4.31%	2.43%
Superiority/ Inferiority/ Normality	**Count**	194	45	47	16	52	34
	% of All	3.17%	2.95%	3.76%	2.14%	2.59%	5.73%
	% of Hate	8.35%	6.11%	8.66%	8.42%	8.01%	16.50%
Stereotypes/ Generalization/ Prejudice	**Count**	522	170	98	53	148	53
	% of All	8.53%	11.15%	7.84%	7.10%	7.39%	8.94%
	% of Hate	22.46%	23.10%	18.05%	27.89%	22.80%	25.73%

HATE		ALL	Gandul	EVZ	Hotnews	Adevarul	Romania Libera
Exclusion/This is our country	**Count**	341	135	39	48	88	31
	% of All	5.57%	8.86%	3.12%	6.43%	4.39%	5.23%
	% of Hate	14.67%	18.34%	7.18%	25.26%	13.56%	15.05%
Animals/Sub-human	**Count**	120	43	30	12	26	9
	% of All	1.96%	2.82%	2.40%	1.61%	1.30%	1.52%
	% of Hate	5.16%	5.84%	5.52%	6.32%	4.01%	4.37%
Holocaust - blame shifting	**Count**	81	8	24	0	19	30
	% of All	1.32%	0.52%	1.92%	-	0.95%	5.06%
	% of Hate	3.49%	1.09%	4.42%	-	2.93%	14.56%
Homosexuality-Pedophilia	**Count**	50	6	24	1	11	8
	% of All	0.82%	0.39%	1.92%	0.13%	0.55%	1.35%
	% of Hate	2.15%	0.82%	4.42%	0.53%	1.69%	3.88%
History	**Count**	158	39	14	19	60	26
	% of All	2.58%	2.56%	1.12%	2.55%	2.99%	4.38%
	% of Hate	6.80%	5.30%	2.58%	10.00%	9.24%	12.62%

HATE		ALL	Gandul	EVZ	Hotnews	Adevarul	Romania Libera
Religious extremism	**Count**	88	2	43	4	28	11
	% of All	1.44%	0.13%	3.44%	0.54%	1.40%	1.85%
	% of Hate	3.79%	0.27%	7.92%	2.11%	4.31%	5.34%
Conspiracy/Foreign interests/Enemies/Threat	**Count**	397	128	99	35	95	40
	% of All	6.49%	8.40%	7.92%	4.69%	4.74%	6.75%
	% of Hate	17.08%	17.39%	18.23%	18.42%	14.64%	19.42%
Denying rights (political/civil)	**Count**	186	58	41	37	32	18
	% of All	3.04%	3.81%	3.28%	4.96%	1.60%	3.04%
	% of Hate	8.00%	7.88%	7.55%	19.47%	4.93%	8.74%
Expulsion	**Count**	165	67	19	11	55	13
	% of All	2.70%	4.40%	1.52%	1.47%	2.74%	2.19%
	% of Hate	7.10%	9.10%	3.50%	5.79%	8.47%	6.31%
Holocaust-denial/minimization	**Count**	86	7	27	0	18	34
	% of All	1.41%	0.46%	2.16%	-	0.90%	5.73%
	% of Hate	3.70%	0.95%	4.97%	-	2.77%	16.50%

HATE		ALL	Gandul	EVZ	Hotnews	Adevarul	Romania Libera
Holocaust, Fascism - appology/justifications	**Count**	71	12	24	0	8	27
	% of All	1.16%	0.79%	1.92%	-	0.40%	4.55%
	% of Hate	3.06%	1.63%	4.42%	-	1.23%	13.11%
Disgrace for the country	**Count**	25	7	3	6	8	1
	% of All	0.41%	0.46%	0.24%	0.80%	0.40%	0.17%
	% of Hate	1.08%	0.95%	0.55%	3.16%	1.23%	0.49%
Discrimination	**Count**	73	16	25	7	20	5
	% of All	1.19%	1.05%	2.00%	0.94%	1.00%	0.84%
	% of Hate	3.14%	2.17%	4.60%	3.68%	3.08%	2.43%
General hate/Discrimination	**Count**	58	29	9	3	9	8
	% of All	0.95%	1.90%	0.72%	0.40%	0.45%	1.35%
	% of Hate	2.50%	3.94%	1.66%	1.58%	1.39%	3.88%

NON-HATE		ALL	Gandul	EVZ	Hotnews	Adevarul	Romania Libera
HS target responding	**Count**	6	4	0	0	1	1
	% of All	0.10%	0.26%	-	-	0.05%	0.17%
	% of Non-Hate	0.16%	0.51%	-	-	0.07%	0.26%
Insult/Profanity	**Count**	143	45	58	1	35	4
	% of All	2.34%	2.95%	4.64%	0.13%	1.75%	0.67%
	% of Non-Hate	3.77%	5.70%	8.19%	0.18%	2.58%	1.03%
Threat/Violence	**Count**	8	1	3	0	4	0
	% of All	0.13%	0.07%	0.24%	-	0.20%	-
	% of Non-Hate	0.21%	0.13%	0.42%	-	0.30%	-
Thrash/Spam	**Count**	44	8	26	0	8	2
	% of All	0.72%	0.52%	2.08%	-	0.40%	0.34%
	% of Non-Hate	1.16%	1.01%	3.67%	-	0.59%	0.52%
Legit	**Count**	3597	731	621	555	1309	381
	% of All	58.80%	47.97%	49.68%	74.40%	65.32%	64.25%
	% of Non-Hate	94.76%	92.65%	87.71%	99.82%	96.61%	98.20%

3.4. Proportion of hate speech against target groups

Table 3.4.a. Target- hate / legit

Topic	Articles	Comments	Hate	Hate Percent	Legit	Legit Percent
Hungarian	41	3640	1377	37.83%	2193	60.25%
Jewish	18	848	334	39.39%	467	55.07%
LGTB	17	1184	431	36.40%	693	58.53%
Roma	7	409	173	42.30%	219	53.55%

Table 3.4.b. Hate speech types based on target groups

HATE	Hungarian		LGTB		Jewish		Roma	
Insults	684	18.79%	225	19.00%	117	13.80%	75	18.34%
Sterilization	-	-	-	-	-	-	5	1.22%
Moderated	32	0.88%	9	0.76%	8	0.94%	10	2.44%
Hate-Spam	8	0.22%	10	0.84%	-	-	-	-
Violence	100	2.75%	28	2.36%	5	0.59%	7	1.71%
Extermination/Murder/Rape	136	3.74%	27	2.28%	46	5.42%	36	8.80%
Threats	121	3.32%	15	1.27%	5	0.59%	7	1.71%

HATE	Hungarian		LGTB		Jewish		Roma	
Superiority/ Inferiority/ Normality	101	2.77%	68	5.74%	14	1.65%	10	2.44%
Stereotypes/ Generalization/ Prejudice	259	7.12%	104	8.78%	85	10.02%	74	18.09%
Exclusion/This is our country	304	8.35%	14	1.18%	18	2.12%	5	1.22%
Animals/Sub-human	45	1.24%	29	2.45%	20	2.36%	26	6.36%
Holocaust - blame shifting	-	-	-	-	74	8.73%	4	0.98%
Homosexuality-Pedophilia	-	-	49	4.14%	1	0.12%	-	-
History	153	4.20%	-	-	5	0.59%	-	-
Religious extremism	2	0.05%	81	6.84%	5	0.59%	-	-
Conspiracy/ Foreign interests/ Enemies/Threat	229	6.29%	61	5.15%	84	9.91%	23	5.62%
Denying rights (political/civil)	121	3.32%	57	4.81%	2	0.24%	6	1.47%
Expulsion	147	4.04%	8	0.68%	6	0.71%	4	0.98%
Holocaust-denial/ minimization	1	0.03%	2	0.17%	81	9.55%	2	0.49%

HATE	**Hungarian**		**LGTB**		**Jewish**		**Roma**	
Holocaust, Fascism - apology/ justifications	2	0.05%	1	0.08%	57	6.72%	11	2.69%
Disgrace for the country	8	0.22%	1	0.08%	2	0.24%	14	3.42%
Discrimination	41	1.13%	22	1.86%	7	0.83%	2	0.49%
General hate/Discrimination	41	1.13%	6	0.51%	7	0.83%	-	-
NON-HATE	**Hungarian**		**LGTB**		**Jewish**		**Roma**	
HS target responding	1	0.03%	5	0.42%	-	-	-	-
Insult/Profanity	57	1.57%	33	2.79%	37	4.36%	14	3.42%
Threat/Violence	1	0.03%	3	0.25%	3	0.35%	1	0.24%
Thrash/Spam	12	0.33%	21	1.77%	9	1.06%	2	0.49%
Legit	2193	60.25%	693	58.53%	467	55.07%	219	53.55%

Table 3.5. Hate speech on article topics

Topic	Subtopic	Articles	Comments	Hate	Hate Percent	Legit	Legit Percent
Hungarian	Hu. Medical University	7	489	174	35.58%	298	60.94%
	Trianon	4	483	201	41.61%	270	55.90%
	Hockey team	8	820	136	16.59%	678	82.68%
	Teritorial reorganization	8	753	396	52.59%	344	45.68%
	Dual citizenship	1	103	45	43.69%	47	45.63%
	Kolozsvar/Cluj statue	3	250	119	47.60%	126	50.40%
	Autonomy/Bruxelles Office of Szeklerland	6	447	202	45.19%	240	53.69%
	March 15/Csibi Barna	4	295	104	35.25%	190	64.41%
Jewish	Toulouse terror attack	4	236	101	42.80%	109	46.19%
	Holocaust	10	488	179	36.68%	289	59.22%
	Holocaust survivors	4	124	54	43.55%	69	55.65%
LGTB	Homosexuality in general	8	597	245	41.04%	306	51.26%
	Gay Billboard/Gay=Peadophile	4	293	89	30.38%	197	67.24%
	Bucharest Pride/Gay Fest	5	294	97	32.99%	190	64.63%

Topic	Subtopic	Articles	Comments	Hate	Hate Percent	Legit	Legit Percent
Roma	Roma way of life	2	58	21	36.21%	33	56.90%
	Criminality	2	111	60	54.05%	49	44.14%
	DEX-Definition	3	240	92	38.33%	137	57.08%

Appendix 4. Examples of hate comments

4.1. Stereotypes/Generalization/Prejudice

"They (*the hungarians*) despise and hate us, and you ask US to be tolerant?. …Why is it our fault that the Hungarians choose chauvinism (in fact they don't choose it, its in their being)" posted by **george on hotnews.ro on 2011-12-21 17:47:00**

"You (*the hungarians*) have occupied the land in question (*the Szekler region of Romania*) by force, and you have terrorized and drove out all the romanians who tried to keep their identity. …You are nice on the outside but on the inside you would kill any Romanian you'd meet" **posted by Detinutul secuiesc** on **2011-12-20 14:47:09 on Hotnews.ro**

"Shameful. They (*the Roma*) spill out children one after another and get state social assistance from the hard work of the working man. Get to work, no more begging from the state" **posted by Costin** on **Jan 24th, 17:13 on evz.ro**

"That's no wonder! That's what the gypsies are: thieves, vulgar and dirty. All the filth of humanity has gathered at this ethnic group. The Romanians have to work to pay their state child support and social assistance." **posted by martha** on **2012-04-19 18:51:18 on adevarul.ro**

4.2. Homosexuality=Pedophilia

"The homosexuals are sick people on the border/intersection with pedophilia. Anal sex with women is the first step towards homosexuality" #2587, **posted by Ed____** on **Nov 3rd, 08:49 on evz.ro**

4.3. Holocaust denial

"What Holocaust???? There was no such thing. Only the *ji dans*[172] sustain this up and strong. But who brought the communism to the world? The *ji dans* "#1422 **posted by Anton Escu** on **Mar 6th, 16:58**

"the Deportation of the Jews in the 2-nd World War was legitimate. They were pro-communists…. All the countries had camps for the hostile population" #6427 **posted by observer** on **18:45, 23 June 2011, on romanialibera.ro**

"I don't deny anything, but let me express a regret: TOO BAD THAT NEITHER HITLER OR ANTONESCU FINISHED THE JOB. Did I deny something? No I

[172] pejorative term referring to Jews modified in order to bypass the profanity filter

did not. Regarding the jews I wish them to remain as many as I have baptized" #1404 **posted by rsss** on **Mar 8th, 08:57**

"All the time *jidans* and holocaust their suffering and all the fables repeated obsessively. Why? Why don't you write about the children murdered in Cambodia, Rwanda, Sierra Leone or the blacks killed in America or the indians??? We had enough of the filthy jidans and their fairytales!!! DEATH TO THE JIDANS!" #6299 **posted by anti-evrei** on **2012-03-20 10:50:55 on adevarul.ro**