V
Research studies of healthcare issues

In the context of this thesis research, the proposed process was used twice to study healthcare issues. The first study deals with the hepatitis C disease. This study was motivated as an extension of the seminal work from Madeira [Mad11]. Despite a different approach, the previous work provided a scientific background to conduct such studies. The second study aimed at a relevant social problem, drug abuse. The research was defined in order to help health agents to understand the problem and as training material for medical doctors. An overview of the results of these studies is presented in this chapter. The purpose here is to provide a glimpse of the results achieved by the proposed process and how it benefits from the computational support proposed in this thesis.

V.1 Hepatitis C

The psychological complications and physical symptoms that arise from hepatitis C are well-known and described as a consequence and conditioning factor for recurrence [Caa12]. Issues such as treatment side effects, perspective of recurrence, and the need for radical lifestyle changes, brings challenges to HCV carriers. It would be impossible to tolerate such obstacles for a long time without the social support from spouses, relatives, friends and other HCV carriers. It is believed that the support coming from specialized virtual communities (VC) represents an important resource for HCV patients who encounter obstacles in adapting to everyday difficulties. The internet offers several tools for organization of virtual networks of chronic patients, which are presented here as an object of study.

The purpose of this research study was to analyze the main social media sources on the Internet in Brazil, especially the online communities. The methodology can be summarized as an incremental internet search in decreasing dimensions of coverage, as described in detail in Chapter III. The results are presented below, organized into the three levels of analysis: The Internet, Social Network Sites, and Online Community.

The Internet

The most common terms associated to HCV — referred to here as primary Keywords (**Kp**) – were identified based on information provided by the Google Insights[1]). The main *Kps* associated with "*Hepatite C*" found on sites indexed by Google were (in Portuguese): *CONSULTA HEPATITE C; CURA HEPATITE C; EXAME HEPATITE C; HEPATITE C CRONICA; HCV; MEDICAMENTO HEPATITE C; MEDICO HEPATITE C; PREVEN-CAO HEPATITE C; REMEDIO HEPATITE C; SINTOMAS HEPATITE C; TRANSMISSAO HEPATITE C; TRATAMENTO HEPATITE C; VACINA HEPATITE C; VIRUS HEPATITE C.* These expressions make it seem that casual and indistinct interest were combined with "specialized" searches.

Social Network Sites

Orkut was chosen because of its long existence (established in January 2004) and its popularity in Brazil. In Orkut, thematic discussions are organized into "topics" in which messages are posted. The site also has special features for searching in which Kps were applied to disclose "specialized" (carriers) VCs. VCs that mentioned any of Kps at least once were selected. We assessed the "relevance weight" among Orkut HCV communities by choosing the ones in which Kps were more frequent. Associations of Kps applied among the most popular and active VCs can neutralize bias caused by arbitrary choices in the pre-selection process. This criteria is based on other algorithms like "page rank" [Brin98], which estimates the relevance of a site using the number of highest expression links directed to it. From this set of VCs, the most representative were chosen by its time of existence, number of members, and mainly by the Kps frequency in discussions. This Dominant Community was considered for study.

The search using these Kps in Orkut resulted in 1,476 communities. The highest occurrences of Kps were found in: *Hepatite C* = 588 topics, *HIV - BR* = 208; *hepatite C* = 183; *Hepatologia: doenças do fígado* (Hepatology: liver disease) = 107; and *hepatite c informações* = 129. These communities had 9066 members (number overestimated, as users might belong to multiple communities) and over 20,000 messages.

Online Community

"Hepatite C" (ID 216788 on Orkut - `http://www.orkut.com.br/Main# Community?Cmm=216788`) was presented as the Dominant Community because it gathered the largest volume of public content. A hepatitis C carrier created this community a few months after the Orkut's launch. It is not connected to

---

[1]`http://www.google.com/insights`

any business, government or other institution and in 2,453 days added 1,292 members - 68% women and 32% men.

The relationship of the Dominant Community with its peers was studied through the Community Association Map (CAM) - (Figure V.1). This CAM defines the interrelationships between communities to portray and confirm their dominance around a core of common interests (Section IV.1).
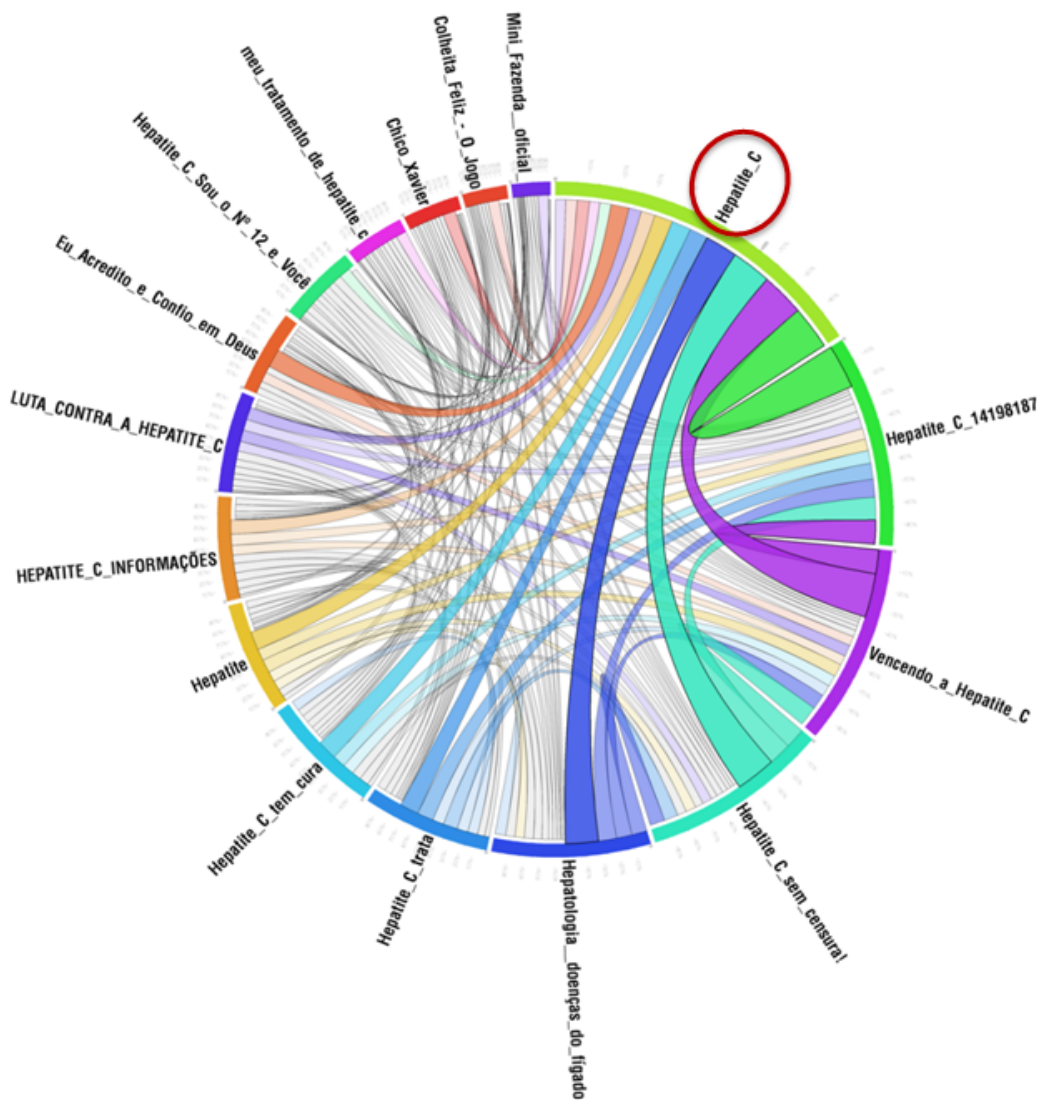

Figure V.1: The Community Association Map (CAM) shows interrelationships between communities of users with the Dominant Community and confirm their dominance around a core of common interests.

The CAM generation is split over many interactions, executed by different programs. Scripts implemented in Lua[2] process the data, while the iMacros script executor on Internet Explorer Browser automates Orkut's system access. iMacros was designed to automate repetitive tasks on the web and is available

[2]Progamming language available at `http://www.lua.org`

at `http://imacros.net`. The final step of visualization plotting is done by a program specialized for this purpose (*i.e.* Circos). The full process execution took two weeks, mainly because of the data gathering step. Details of the CAM generation are presented below.

1) Data gathering

The application considered Orkut communities, which were chosen mainly because of the experts' interests. The data acquisition is scraping-based due to an Orkut system limitation (no API for this purpose is available). The data acquisition was an exhaustive crawling within a defined boundary, having its starting point in a community of interest.

Instead of taking all members, the process considered as input only users that posted messages on the community forum (467 active users). From these users, the crawler was able to retrieve 48,218 associations with 35,954 communities.

2) Model and measurement

Most of the 35,954 communities have few users associated with them. The use of these communities in the model only adds an extra effort in the model processing. A filter to remove the communities with less than 20 users is applied to build a more concise model with 30 communities. The communities' relationships model is built based on a combination of these communities.

Table V.1 shows the top 16 of the 30 communities considered in the modeling step, those with the most users. The first column has the name of the community, followed by the column with the total of members informed by the Orkut's system, and the last column is the number of identified users. The data collection took place in March 2010. Membership fluctuation is frequent in those systems. Of the 467 users that posted messages on the forum, only 382 were members of the studied community at the time of the study. The table provides information about user association, but does not have any information regarding the relationship among the communities. The full relationship model is a combination of all 30 communities, which is represented by 435 meaningful relationships.

3) Plot the visualization

In the final step of the process, graph plotting, the last filter is important to provide a smooth visualization. A plot with the 15 communities related to the analyzed community has 120 weighted relationships. It is the combination of 16 communities (15 + analyzed community).

Community Name Translation	Community Members	Users Associated
Hepatite C Hepatitis C	1,285	382
Hepatite C 14198187 TN: it has the same name, but it is a different community and its id is the 14198187	423	111
Vencendo a Hepatite C Beating Hepatitis C	602	107
Hepatite C sem censura! Hepatitis C without censorship!	505	100
Hepatologia: doenças do fígado Hepatology: diseases of the liver	1,719	83
Eu Acredito e Confio em Deus I believe and trust in God	6,014,818	54
Hepatite C tem cura Hepatitis C has cure	154	50
Hepatite Hepatitis	282	48
Hepatite "C" trata Hepatitis C treatable	182	48
Hepatite C informações Hepatitis C information	77	37
Colheita Feliz - O Jogo TN: a famous Brazilian farmville-like game	1,706,040	34
Chico Xavier TN: popular medium in Brazil's spiritism movement	424,776	33
LUTA CONTRA HEPATITE C Fight against Hepatitis C	141	32
Hepatite C: sou o Nº 12 e você? Hepatitis C: I'm number 12 and you?	78	28
Mini Fazenda [oficial] TN: another famous Brazilian farmville-like game	956,159	28
meu tratamento de hepatite c my hepatitis c treatment	56	27

Table V.1: Top 16 communities with more users associated

Figure V.1 shows the plot for the model considering the top 15 communities related to the Hepatite C community. The image is built using a special component of Circos, the Table Viewer[3]. Table Viewer has many configuration

---

[3]http://mkweb.bcgsc.ca/circos/presentations/articles/vis_tables2

variables, and it must be customized for better visualization according to each expert's preferences. The last step is completed in two phases, first generating the input file for Table Viewer and then executing it.

The communities are displayed in the graph following clockwise orientation in descending order of the relationship weight with the analyzed community (Hepatite C). The connection lines between communities show the relationship weight by thickness and transparency. Broad and opaque relationships are stronger than narrow and translucent ones. One can compare Table V.1 with Figure V.1 to see the difference in the community ordering, because the first has its order based on associated users and the second relies on the model to calculate which communities are more related to the Hepatite C community. The relationship model plot reveals a more recent interest of the users at the moment of the research, because 85 users are not associated with the Hepatite C community anymore, and the plot shows this property in the new ordering when displaying the related communities.

The plot in Figure V.1 reveals the interests of active users in other communities related to hepatitis, meaning that they are users looking for information and support in other communities related to the disease as well. The names of related communities are in Portuguese because this is the language used in the Hepatite C community. In order to help the reader that does not know Portuguese to understand the relationships, a literal translation of the community names is presented in the Table V.1.

It is important to remember that the Community Association Map (CAM) is more interesting when compared to other research artifacts, as part of the whole content analysis [Koz09]. However, it is interesting to note that most of the associated communities are related to Hepatitis C, showing the strong interest of the active users in discussing the disease.

The Orkut environment has all kinds of communities and, even without weighting the relationships among the communities considering their size, the graph shows that the users are engaged with other communities regarding the topic of interest. The trend showed in the graph is the user interest in other communities related to the disease. In this case, no special interest stands out of the associated communities related with brands. Two other hot topics are *religion*, also identified in the discourse analysis as an important matter of the treatment, and *games*, considered by far the most popular application in social networking sites. Therefore, it reinforces the validity of the discourse analysis of this community content in the sense of getting a broader perception of what its user population says and seeks.

Content of messages posted on "Hepatite C"

We developed scripts to collect and tabulate all messages published on all topics over eight years of the community's existence. Then we studied the most frequent terms in messages to identify the themes of greatest interest. A word cloud (Figure V.2), generated in Wordle (`http://www.wordle.net`), provides a graphic representation of word frequency, with each word's size determined by the relative number of times it was cited (*i.e.* word frequency). The most used word in the forum is "Nao", which means "No" in Portuguese. Although one might think that it would imply that it is a negative community, it seems this happened because of ordinary language use. This conclusion is backed by the discussion analysis carried later. The other distinguished words, such as "tratamento" (treatment), "deus" (God), "medico" (doctor); are high frequency words in discussions about technical matter as much as in discussions about emotional support.

Figure V.2: Word cloud of the Dominant Community forum.

The message volume in the *Dominant Community* remained low until the first quarter of 2009, thereafter the number of messages increased significantly, as shown in Figure V.3, which matches with the popularization of the site in Brazil [Ara12]. Of the total of 47,005 terms found on all topics in 8 years of available content, the most recurring words in frequency analysis (discarding articles, prepositions and numerals) are presented in Table V.2, which also presents a subset of other prominent words (that refers to HCV therapy and potentially toxic drugs to carriers). From the most cited drugs, interferon and ribavirin shared six of the eight stronger relationships among drugs cited in the same message. These findings suggest a demand by the strong association

between ribavirin and interferon - its continuing use is a frequent problem to patients of Brazilian healthcare system [Sou08]. The three types of information that are most frequently found with this association are *preço* (price), *bula* (medication user instructions) and *efeitos colaterais* (side effects).
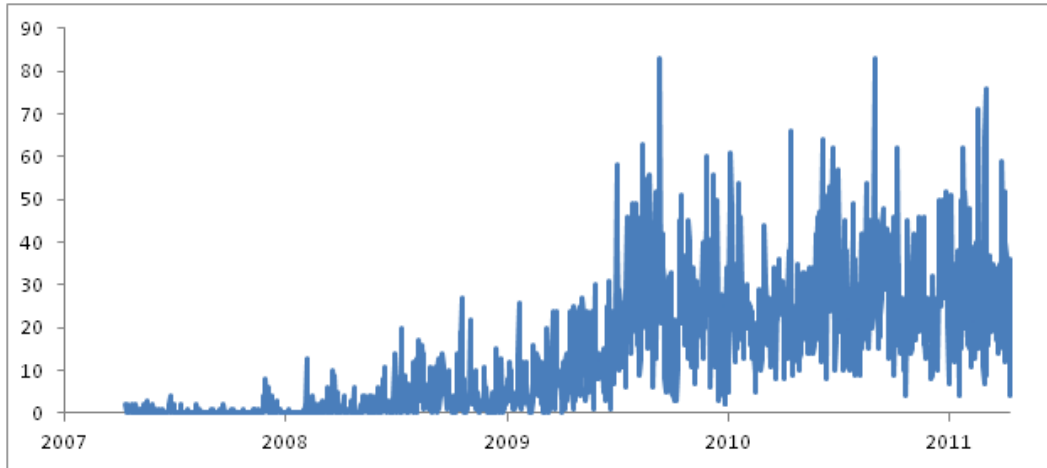

Figure V.3: Historical evolution of messages per day posted at the Dominant Community forum.

Discussion and synthesis of results

Recurring demands of hepatitis C patients are presented here as a low cost method easily applicable for guiding qualitative researchers on data collection. Hypotheses, linking concepts and "bounding ideas" are essential to the portrayal of social support and can be easily weakened by bias and personal assumptions - which can be preventable by the content analysis. HCV carriers are vulnerable to several psychological conditions and depressive symptoms are usually identified and reported among them [Lit12, Shi12]. The identification of recurring demands in communities with chronic diseases may expand our comprehension about their needs for social networks, presenting demands perhaps underestimated by public health policy makers. It's interesting to notice that the results presented here support other conclusions recently reached by other methods. Sousa [Sou08] describes carriers expressing their suffering strongly attached to expectations of obtaining medicine and healing in the context of the passage of the weeks, months and years of survival. The elements described herein provide a basis for further, more detailed, research, in which the PRD are consolidated into central ideas for the construction of analysis categories. The primary purpose of the paper was to furnish unbiased material to a qualitative approach, which could reach findings applicable beyond the immediate boundaries of the study. According to Qualitative Research literature [Jan91, Jup06] it's especially effective in cultural research

Most recurring words in frequency analysis	Citations
Treatment	9581
God	4077
Hepatitis	3329
Physician	2800
Virus	2674
Interferon	2281
Husband	971
Well	3411
Do	2822
Patients	1279
Recurrent words related to the passage of time	
Day	3156
Always	2852
Week	2285
Then	2170
Years	1771
Months	1454
During	1199
Therapies for the control of HCV	
Interferon	1554
Ribavirin	1048
Erythropoietin	317
Pegasys©	201
Folic acid	157
Pegintron©	130
Filgrastim	109
Silymarin (alternative treatment)	63
Potentially toxic drugs to CHCV	
Tylenol©	121
Omeprazole	80
Paracetamol	69

Table V.2: Subset of prominent words from the *Dominant Community* forum (translated to English).

which deals with values, opinions and perspectives which can be generalized in a broader view.

Another notable aspect of this research method refers to the study of virtual communities through algorithms, a field underused so far despite many interesting alternatives. In addition to reduced costs compared to conventional field research, there is opportunity to capture discourses posted

in moments of desperate need for support. Here, Internet virtual communities seem to transcend their merely informative context [Fer11], and acquire a unifying force aimed at overcoming great obstacles [Gre11]. Besides posting messages on topics for mutual enlightenment and social support, maybe HCV carriers feel more comfortable talking about personal difficulties and living conditions when compared to the conventional medical consultation environment. In general, stigmatized diseases or health conditions encourage individuals to take advantage of the Internet as an important source of information and environment for sharing experiences [Ber05, Gre11, Mad11]. We believe, however, that such preferences are not limited to the possibility of hiding identities in the face of uncomfortable topics, but also include the role of social support. The analysis identified frequent use of words that suggest a need for spiritual support (God: 4077 citations), social support (husband, 971). But above all words indicating the need for material support/care were mentioned: *treatment* (9581), *doctor* (2800), and *interferon* (2281), and the association between "Interferon" and "Ribavirin", which has also been described by Sousa [Sou08]. These results reinforce evidence that patients with chronic diseases have a distinct profile of engagement in virtual communities, shown by an agile dissemination of certain content and thematic consistency associated with interest in news about innovative therapies (new formulations of interferon; alternative therapies).

V.2 The drug crack cocaine abuse

Here some results are presented from a study about motivations for drug abuse to start and cease, specifically with regard to the drug crack cocaine in Brazil. The initial research motivations were to study: 1) why do people start to abuse drugs; 2) why do they continue abusing them; and 3) why do they cease to abuse them. All three stages were conducted, providing a descriptive panorama about drugs on the Internet, especially with regard to the Brazilian Internet audience, and revealing the reality of a support community of users of crack cocaine. As a result of the community content analysis, the report compiled answers to the follow questions: 1) what are the leading factors to crack; 2) what are the optimal turning points at which to start a treatment; 3) what are abstinence maintenance factors; 4) what favors the drug abuse to restart; 5) what criticism exists for official health treatment; and 6) which kind of help are the codependents looking for.

The Internet

Following the first stage recommendation of the considered approach, Google Insights for Search was used to get an overview of the search trends

about the term "crack" in Brazil. Because of the ambiguity of the term "crack", which is also used by users looking for illegal software and licenses on the Internet, the configuration service was set to retrieve only searches related to the category of health.

Social Network Sites

Since the major source of social media in Brazil is Google's Orkut, this social networking site was used as the investigation platform for the stage two. A search for the term "crack" in the Orkut system, considering filters for location (Brazil) and language (Portuguese), gave 995 results in September of 2011. The next step was to select communities where content would have contextualized discourses about people's experience related to drug use. Categorizing the 995 communities, it was possible to identify 278 (28%) communities related to selection objective, 360 (36%) communities that seemed not to be directly related to the selection objective, and 357 (36%) communities that were not using the term "crack" in reference to a drug (e.g. instead referred to programs and password cracking). Narrowing the research, the 278 identified online communities were filtered to 13 communities considering other criterias such as: 1) possessing more than 300 members; 2) having been in existence for more than 6 months; 3) having exhibited recent activity; and 4) having content publicly available. The last step of this stage was to choose one community for evaluation. The community "Crack, Nem Pensar - AJUDA"[4] was elected for in-depth analysis because, out of the 13 remaining communities, it is the oldest and has the most members (11,102). In a quick evaluation of its content, the community presented an intense conversation among its members, which in later analysis showed an average of 3.3 messages per day since its creation in July 16 of 2004.

Online Community

The community analysis focused on participating members who had engaged in conversation in the community forum. It is important to make this distinction, as all members have the potential to follow the discussions, but most choose not to participate (*i.e.* lurkers). This analysis is based on the content of the participants who have posted messages in the forum. At the time of the study, September 2011, the community forum had 434 participants, 384 topics and 8655 messages, representing a total of 76.646 words, or 4.515.087[5] characters. The content analysis was conducted by applying the Discourse of Collect Analysis technique. Considering the high volume of data and efforts required for content analysis, a data cut was performed to focus the

---

[4]http://www.orkut.com/Main#Community?cmm=175318
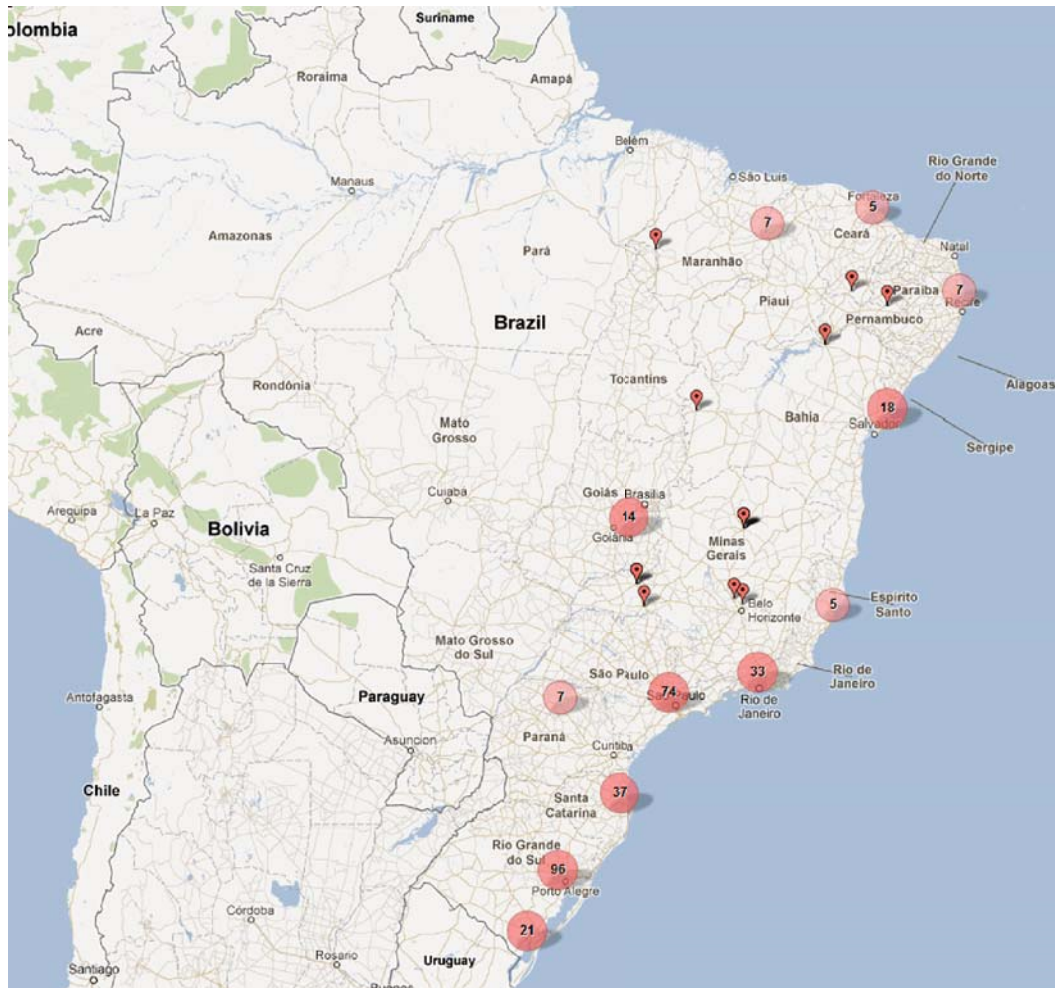[5]The B-42 Gutenberg's Bible has around 3 million characters.

Figure V.4: Geographical distribution of the participants from the analyzed online community.

investigation in a suitable content analysis to the study objectives. From the original data set, 39 (10%) topics were selected, with 129 (30%) participants and 925 (11%) messages, totalizing 107.488 (14%) words, or 602.332 (13%) characters.

From the participant data available, there were 57% men and 43% women identified. The retrieved location of the participants was consolidated in a map as show in Figure V.4[6].

Discussion and synthesis of results

The community analysis identified that the speeches of dependents and codependents (the family and friends of dependents) are intermingled and complement each other, therefore both require care and attention. The reality of these people (e.g. life experiences recorded in the discussion) is transcript through discourse syntheses that answers to the study research questions. A compilation of the results in Portuguese has been submitted to a journal but

---

[6]available online at `http://batchgeo.com/map/536db2e5aac00f746005efc6334542c4`

it is still being processed. The study results have been a subject of discussion in a seminar organized by the Sírio-Libanês Hospital in January of 2012 in São Paulo (Brazil), with attendees from the Brazilian government, health organizations and general public. This example gives an indication of the hard task researchers might face when performing content analysis of online forums and the valuable outcome that can be achieved from its analysis.