Working together
www.rcis.ro

Revista de cercetare și intervenție socială

ISSN: 1583-3410 (print), ISSN: 1584-5397 (electronic)
Selected by coverage in Social Sciences Citation Index, ISI databases

EMPIRICAL SOCIAL RESEARCH TO IDENTIFY CLUSTERS OF CHARACTERISTICS THAT UNDERLIE THE ONLINE EVALUATION OF ACCOMMODATION SERVICES

Olimpia I. BAN, Versavia ANCUSA, Victoria BOGDAN, Ioan Gh. TARA

Revista de cercetare și intervenție socială, 2015, vol. 50, pp. 293-308

The online version of this article can be found at:
www.rcis.ro, *www.doaj.org* and *www.scopus.com*

Published by:
Expert Projects Publishing House

On behalf of:
„Alexandru Ioan Cuza" University,
Department of Sociology and Social Work
and
Holt Romania Foundation
REVISTA DE CERCETARE SI INTERVENTIE SOCIALA
is indexed by ISI Thomson Reuters - Social Sciences Citation Index
(Sociology and Social Work Domains)

Empirical Social Research to Identify Clusters of Characteristics that Underlie the Online Evaluation of Accommodation Services

Olimpia I. BAN[1], Versavia ANCUSA[2], Victoria BOGDAN[3], Ioan Gh. TARA[4]

Abstract

In tourism, online reviews are tools for researching the perceptions of the tourists. The evaluation that all review websites do (Booking.com, TripAdvisor.com, Expedia.com, etc.) involves making an average between the grades of 5, 6 or 7 features, each having an equally weighted average. It is likely that the average obtained does not reflect actual consumer satisfaction, because for every individual, the characteristics have different importance weights and may even vary in content from one individual to another or from one situation to another. The analysis of the reviews associated to the assessments would be revealing but it is difficult to do, give their descriptive nature and the high number. The purpose of this paper is to find a method which allows the identification of the characteristics of hotels' offer, from the reviews posted online, which are of interest to consumers. This paper proposes a method capable of operating with a large amount of descriptive data (15,200 reviews), complex neural networks, and to identify clusters of characteristics of hotels' offer, useful for configuring and improving the offer. The method is tested through an empirical sociological research on all the reviews posted at a time by the AmFostAcolo Romanian website.

Keywords: satisfaction, accomodation, online evaluation, complex network, Romania.

[1] University of Oradea, Department of Economy, Faculty of Economic Sciencies, Oradea, ROMANIA. E-mail: oban@uoradea.ro

[2] Politehnica University of Timisoara, Timisoara, ROMANIA. E-mail: versavia.ancusa@upt.ro

[3] University of Oradea, Department of Finance & Accounting, Oradea, ROMANIA. E-mail: vbogdan@uoradea.ro

[4] University of Oradea, Department of Finance & Accounting, Oradea, ROMANIA. E-mail: ioantara@yahoo.com

Introduction

Online reviews are defined "as peer-generated product evaluations posted on company or third-party websites (Mudambi & Schuff, 2010). They work as word-of-mouth on Internet (eWOM). Online reviews have become a major source of information on the quality of the product, both for consumers and for marketers (Hu, Liu & Zhang, 2008; Zhang, Ye, Law & Li, 2010; Levy, Duan, & Boo 2013). There have been studies on: the influence of eWOM on the other consumers (Filieri & McLeay, 2013; Min, Lim & Magnini, 2015), the influence of price on the quality and the perceived value of hotel services in the online reviews (Ye Li, Wang & Law, 2014), the link between quality and quantity of reviews and sales (Blal & Sturman, 2014; Ye, Law & Gu, 2008; Vermeulen & Seegers 2009), the response of hotel managers to the online reviews (Park & Allen, 2013), what determines the consumers to write reviews online (Melián-González, Bulchand - Giduma1 & López-Valcárcel, 2013) and less on the content of the reviews and characteristics of the products described in these reviews (Pekar & Ou, 2008; Yi, Bunescu & Niblack, 2003; Popescu & Etzioni, 2005).

Chevalier and Mayzlin (2006) concluded from the research made that the purchasing decision may be more influenced by the content of the reviews than the ratings of the accommodation. As a source of information, they are comparable with questionnaires and interviews. The studies find that online reviews provide more reliable information than those collected directly about the perception of product and service quality, satisfaction, dissatisfaction (Ye *et al.*, 2014; Zhang *et al.*, 2010). Starting from the idea that "most researchers have concluded that the motivation for eWOM occurs when a guest has an extremely good or extremely bad experience" (Melián-González *et al.*, 2013: 275), we can assume that in the reviews posted will be found those characteristics of the offer which, in one way or another, have attracted the attention and prompted tourists to publicly express their opinion. It is possible that those characteristics have a larger percentage in the overall satisfaction of the consumer, meaning to be more important than others for the tourist.

Research methodology

The purpose of the study is testing a method of analysis which should allow work with a large amount of descriptive data in the form of online reviews, for extracting useful information for tourism managers. Objectives of research: (1) To identify the characteristics of online reviews, other than those suggested by websites, which could be the basis of the offer configuration, in a wider compatibility with what the consumers seek; (2) To find evaluation patterns, clusters showing a good number of images (pictures, characterizations) of accommodation, made by Romanian consumers.

Hypotheses of research

H1: There are features that repeat at a frequency obvious in the reviews posted on the review websites (which we assume to have greater importance in the evaluation system of the reviewer).

H2: There can be found between 1 and 10 evaluation clusters that show the most important areas in the evaluations made by online reviews.

H3: The evaluation clusters have different weights as frequency and do not coincide with the evaluation characteristics proposed by the website investigated.

The research base was the travel reviews website I Was There (amfostacolo.ro), which is a Romanian website where you can gather and where you can post holiday impressions, you can watch pictures; you can see and compare deals and can make bookings. The website includes websites and accommodation facilities in over 60 destinations worldwide, with related evaluations and network moderators of the destination. The amfostacolo.ro website uses several indicators to evaluate the satisfaction, catching also the evaluation part beyond the 5 quality features used (accomodation, services, meal, natural environment, entertainement/ relaxation). Data were gathered from the website on 04.30.2015 (posted at the time) and entered into an excel document, from where we selected the desired information as specified in the Table. 1.

On the AmFostAcolo.ro website, as it happens on all current reviews websites (Booking.com, TripAdvisor.com, Expedia.com, etc.), there is no distinction between the overall score of the evaluation and the degree of satisfaction. It is considered that the degree of satisfaction is given by the average of evaluating the five characteristics, which is untrue because it does not take into account the percentage of various features in the evaluationt. As a result, there is no weighting of the evaluation characteristics according to the different importance that each consumer individually attaches to them. The analysis of the evaluation of characteristics suggested by the websites shows an ignorance of some in the evaluation, which makes us believe that they are not important for the evaluator, or at least in this particular evaluation situation (Figure 1). Consequently, we can not correctly and fully capitalize the information that these websites offer us.

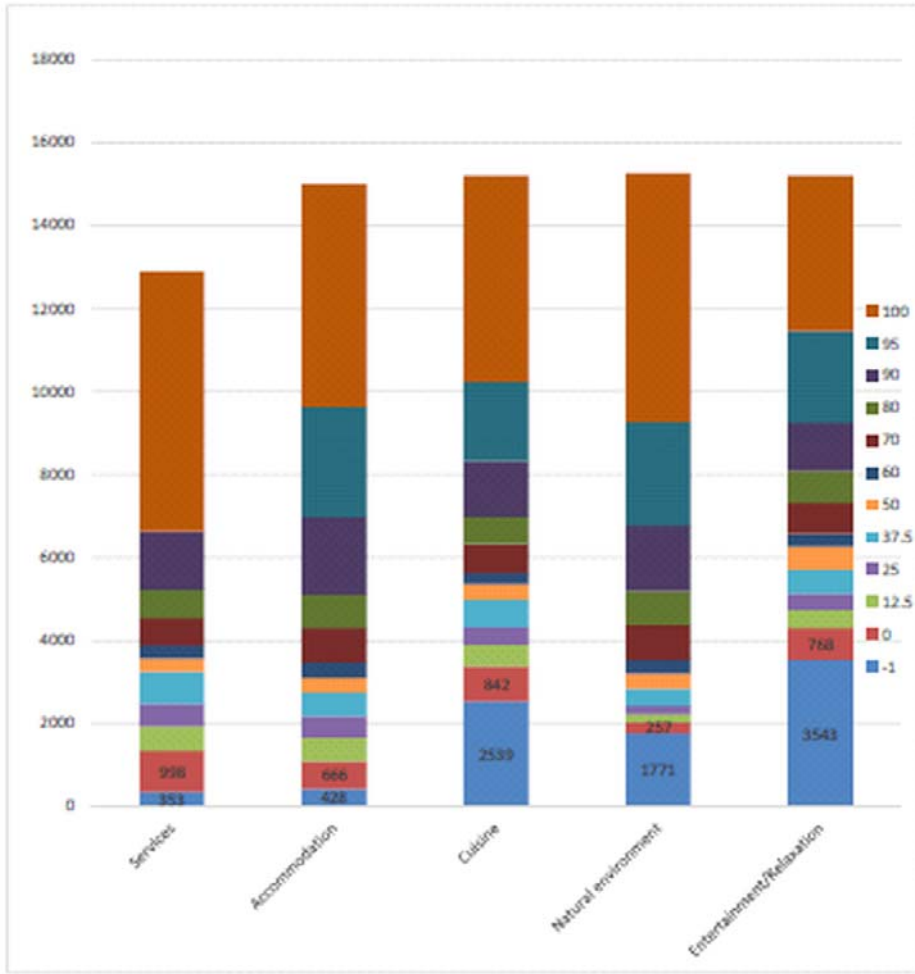

Figure 1. *Distribution of points on evaluation characteristics, for the 15,200 evaluations*

The analysis of evaluations on each quality characteristic shows a significant percentage of non-evaluations (marked by „- 1") or the awarding of „0" points. For example for the "Entertainment / Relaxing" characteristic, out of the 15,200 evaluations, 28.36% were of this type (Figure 1). We can deduce that for these respondents those characyeristic evaluated with „0" points or for which no score was specified are not generally important or not important in the context of that service package. Certainly, the quality characteristics have different importance in the individual evaluations and therefore should have different weights in the calculated average.

The difficulty of evaluating the reviews posted is related to the large number of them (which is a tremendous advantage of the websites) or 15,200 only for the evaluation of accommodation structures in Romania on the AmFostAcolo.ro website. Qualitative research is carried out on small samples because they involve semantic, text, correspondence analysis etc. New technologies have dramatically improved the samples and have found methods of operation and processing of these quantities of data.

In this paper, we propose a method capable of operating with large amounts of data, provide results easy to interpret and useful for the decisions of managers in tourism.

Table 1. Characteristics of study population

Characteristics	Absolute values	Relative values
Accommodation structures in Romania reviewed on the website	3,755	100%
Reviewers for the structures in Romania	by id: 9418 by name: 9417	100%
Reviews for Romania	20,883	100%
Reviews for locations in Romania	5,683	27.21%
Accommodation reviews for Romania	15,200	72.79%
Age groups for reviewers for Romania		
<16 years old	32	0.34%
16-20	98	1.04%
20-30	2,081	22.10%
30-40	4,709	50.00%
40-50	1,913	20.31%
50-60	441	4.68%
>60	88	0.93%
Type of travel of reviewers for Romania		
Single	182	1.93%
Childless couple	2973	31.57%
Families with children	5038	53.49%
Any	6	0.06%
Friends	1000	10.62%
Team-building	216	2.30%
Colleagues (there is this possibility in the file)	3	0.03%

A fairly common problem in today's world is the translation of data to information. This is the case with narratives of any sort as they abound, consequently leading to the difficult mission of quantifing them (Zou, Jin, Yang & Du, 2013). The methods used may vary from classic statistical analysis to various data mining tehniques, depending on the context and the information sought. If data interaction is a point of interest, complex networks may be used. This is significant, in the light of Gershenson (2013), which stressed the point that the individual words interaction is of great consequence in natural language processing. Fundamentally,

a complex network is a graph, it contains nodes that interact with each-other through edges, that are directed or not. A key difference between a complex network and a graph is the fact that a complex network models the interactions in a real-life system, while a graph is not constrained by such modeling issues. Therefore, complex networks should be seen as a graph-based analysis method of real-life systems that can highlight connections and connectors, as well as interconnected groups of connectors (Barabasi, 2012).

Studies have shown that most real-life systems models have certain mathematical properties that transcend study fields: for example, the same nodes weight distribution power-law pattern characterizes a network describing litterary works and social connections (Grabska-Gradzinska, 2012).

An advantage of using complex network analysis on text is the posibility of uncovering previously unidentified knowledge as no human (linguistic) intuition is employed (Colobert, 2011). Some studies (Biemann, 2006; Choudhury, 2009) point towards clustering and/or degree as two pertinent evaluation parameters. While degree reflects the frequency of connection to other nodes, clustering, "also known as transitivity is a typical property of acquaintance networks, where two individuals with a common friend are likely to know each other" (Bocalletti, 2006). This leads to the concept of communities "defined as groups of nodes such that there is a higher density of edges within groups than between them" (Bocalletti, 2006).

Stages of research and results

The application of text analysis using neural networks involved some laborious data processing stages, based on the data in raw form on the AmFostAcolo website. The reviews posted on the website were uneven, inhomogeneous, far from standard language, without diacritics mainly, with speech or grammatical mistakes. The challenge was to exploit as well as possible the posted text and the overcoming of numeric barrier (15,200 reviews).

Stage 1: Data selection. Starting from the database of the website, the first stage consisted in selecting the column showing the users' comments, the column describing the area for which the comments were made, the column with the scores given to the location and the PMA column.

Stage 2: Data preparation. (1) *The column describing the area*: the columns describing the area were grouped on national development regions (according to official data from the government). To achieve this classification, there were selected unique instances from the column that shows the location of the database that totaled 236 entries and for each entry was determined the belonging to a county, which then was framed into a development region. For example: Avrig belongs to Sibiu county, which belongs to the central region. There were also

problems because some locations present general guidelines pertaining to several counties. In the end, this process enables data sharing so that regional specificity be more easily determined; (2) *The reviews column*: The reviews are the corpus to be evaluated. From the start, due to the significant number of entries (15200) and because of the stylistic variation of entries (8912 distinct users) there were obvious processing difficulties such as the use of diacritics or not or the use of colloquial expressions, the use of abbreviations, vocalizations and so on. Therefore we created a mini-corpus which can be excluded, that included: linking words, pronouns (all conjugations of auxiliary verbs to have and to be, regionalisms, onomatopoeia / interjections, numbers, certain abbreviations and misspelled words. Some abbreviations have been replaced with the literary forms of the words. In total the filtering corpus had over 480 distinct entries which totaled 4,011,135 replacements in the corpus to the processed. In the end, this allowed the uniformization of the corpus to be processed.

Stage 3 Building word interaction networks. For each review passed through the filtering stage a mini-network was generated as follows: (1) each distincy word will form a node; (2) two adjoined words form a ridge with equally weighted score calculated in step 2. Example: Suppose we have as a comment the following text: "very good hostel highly recommend similar facilities star hotel turkey meal rich buffet good quality quantity great view large landscaped garden arranged grill special place". The mini-networks were then combined to obtain the neyworks of the development regions.

Stage 4: Extracting national characteristics. The intersection of the networks of all development regions represents national characteristics. The national network has 5296 nodes and 507,024 edges and rendered with no specific layout looks like in Fig. 4. Despite its extremely complex aspect, it is relevant to mention that, mathematically, this network presents only 8 communities of different sizes individually reflecting different characteristic aspects (ex. road, surroundings, people, and food). It is important to note that the division into communities was not made manually but automatically according to the algorithm from the paper Blondel *et al.* (2008), implemented in Gephi. Also, it should be noted that a crucial parameter in the calculation of communities is the percentage of edges. Further, the communities numbered from 0 to 7 will be also analyzed individually.

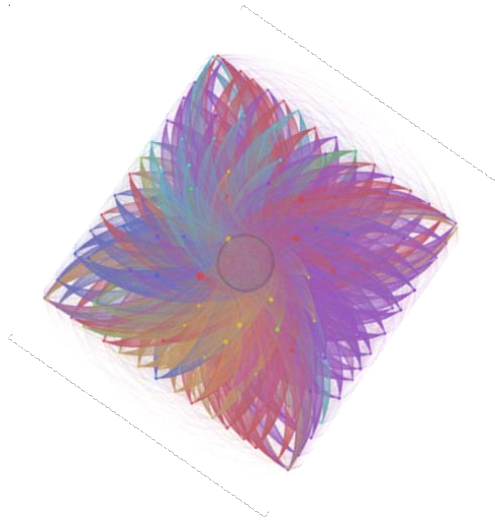

Figure 2. *National network obtained, with 5296 nodes and 507024 edges*

Basically, by determining the communities and especially the theme (the hidden meaning) of the communities will determine the answer to research topic 1: the characteristics of the evaluation system - which matters for the review writers. The size of the community tells us how much it matters. In turn, each community can be decomposed into mini-communities which have sub-features or otherwise expressed, the evaluation sub-criteria can be determined. The graphical representation of networks was done using an Atlas Force layout, node scaling after connecting nodes, edge scaling to the same size and different coloring of the communities through Gephi. There were obtained eight communities / clusters (*Figure 4*).

Table 2. *Example: Components of Community/Cluster 0-Road*

Place	Percentage	Representative key words	Name community
1	26.12%	evening, hour, morning, arrival, night, late	Beginning
2	22.47%	departure, day, itinerary, finished, tiredness	End
3	12.92%	days, holiday, nights, weekend	Duration
4	8.71%	past, year, week, August, month, end, July, December	Time horizons
5	8.71%	period, time, day, journey, left, nights, itinerary, weeks	Duration personal
6	8.15%	day, date, first, before, last, time, leave, weeks	Duration impersonal
7	7.58%	surprise, experience, music, pleasant, stay, band, live, unpleasant, dance, concert, singing, show	Music ambient destination
8	2.81%	tourists, necessary, majority, comfort, clients	Undefined
9	1.12%	week, end, of, the	Undefined/English
10	0.84%	around, lunch, hour	Undefined
11	0.56%	smile, lips	Undefined

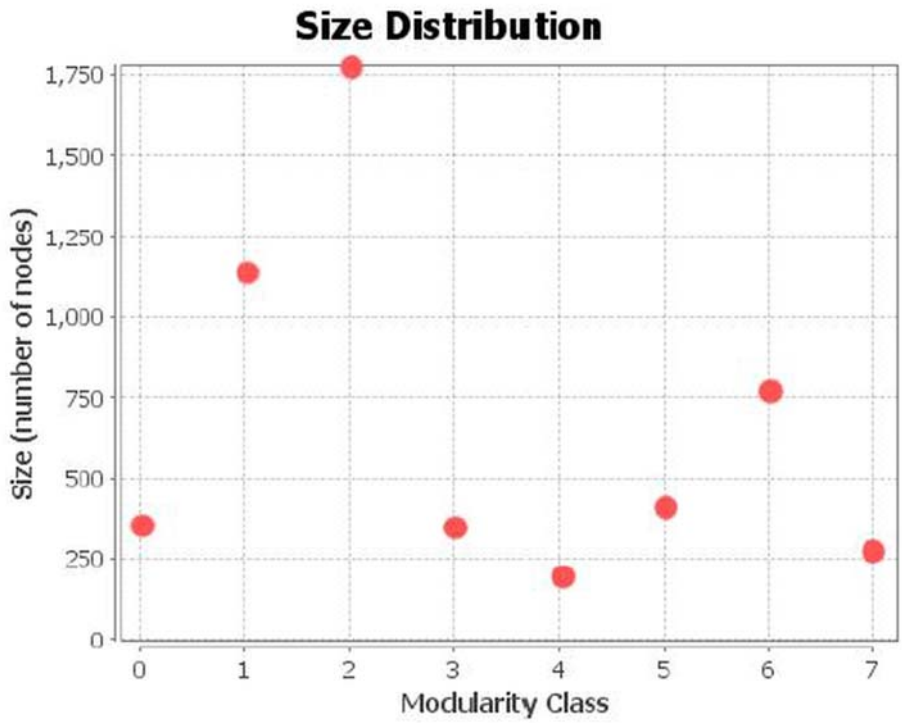

Figure 3. *Reprezentation of the 8 clusters with afferent dimensions*

The first community / community cluster identified is Community 0 – "Road". This community has 356 nodes, 5665 edges, organized into 11 distinct sub-communities. The degree of interconnection is 0.089 with a graph diameter 4 (Figure 3).

Cluster 1: *Description of accommodation compared with the destination area* has 1146 nodes, 39571 edges, organized into 11 distinct sub-communities. The degree of interconnection is 0.06 with a graph diameter 4.

Cluster 2: *Interaction during accommodation* has 1776 nodes, 69,282 edges, organized into 11 distinct sub-communities. The degree of interconnection is 0.044 with a graph diameter 4.

Cluster 3: *General conditions of accommodation* has 349 nodes, 4569 edges, organized into 11 distinct sub-communities. The degree of interconnection is 0.075 with a graph diameter 4.

Cluster 4: *Ambient* has 201 nodes, 1683 edges, organized into 11 distinct sub-communities. The degree of interconnection is 0.084 with a graph

diameter 5 (the only community with a diameter other than 4). This implies greater value dispersion.

Cluster 5: *Food* has 413 nodes, 7314 edges, organized into 11 distinct sub-communities. The degree of interconnection is 0.086 with a graph diameter 4. The common theme of the majority of nodes in the community 5 is the food.

Cluster 6: *Subjective evaluation* has 776 nodes, 19741 edges, organized into 9 distinct sub-communities. The degree of interconnection is 0.075 with a graph diameter 4.

Cluster 7: *Room facilities* has 279 nodes, 3083 edges, organized into 11 distinct sub-communities. The degree of interconnection is 0.079 with a graph diameter 4.

Summarizing the tables within each class we can create the following situation in which communities / clusters and sub-communities are obvious, in which hierarchies can be made according to the frequency of mention (Table 3). Although we analyzed only the reviews on accommodation (not travel) we observe that the considerations refer to the road to the destination and the environment where the location is placed. All these contribute to the overall journey satisfaction, although they can not be influenced by the supplier of accommodation services.

Table 3. *Communities/clusters and subclusters and the frequency of mention*

Cluster	Percentage	Percentage in categroy	Name community	Compound percentage
Road	6.72%	26.12%	Beginning	1.76%
		22.47%	End	1.51%
		12.92%	Duration	0.87%
		8.71%	Time horizons	0.59%
		8.71%	Duration personal	0.59%
		8.15%	Duration impersonal	0.55%
		7.58%	Music ambient destination	0.51%
		2.81%	N	0.19%
		1.12%	N / English	0.08%
		0.84%	N	0.06%
		0.56%	N	0.04%
Framing into the area	21.64%	18.45%	framing into the area	3.99%
		17.14%	room location	3.71%
		14.67%	dining room location	3.17%
		12.83%	description of visible landmarks in the accommodation	2.78%
		11.42%	how to get to the lodging	2.47%
		5.71%	entrance cottage	1.24%
		5.27%	description area and landmarks	1.14%
		5.01%	architectural details	1.08%
		3.95%	cottage area	0.85%
		3.78%	view	0.82%
		1.76%	landmarks in the county	0.38%

Interactions during accommodation	33.53%	17.51%	interaction with people belongong to the accommodation	5.87%
		15.54%	motivation trip	5.21%
		11.15%	possible activities	3.74%
		9.91%	personal information on the interaction important to share	3.32%
		8.16%	corelations with the time context	2.74%
		8.05%	positive personal impressions	2.70%
		7.94%	positive subjective impression on the interaction with people belonging to the accommodation	2.66%
		7.09%	comparison with information existing on-line	2.38%
		5.69%	negative subjective impression on the interaction with people belonging to the accommodation	1.91%
		4.73%	ways to express the review	1.59%
		2.98%	help got/offered	1.00%
		1.24%	negative personal impressions	0.42%
General conditions of accommodation	6.59%	28.62%	special offers	1.89%
		26.42%	payment	1.74%
		23.27%	the way the dining table looks like	1.53%
		11.01%	what can be organized	0.73%
		4.40%	breakfast/ n	0.29%
		3.46%	extra / n	0.23%
		2.20%	society games	0.14%
Ambient	3.80%	28.35%	group travellers without children	1.08%
		18.56%	groups with children	0.71%
		12.89%	purpose trip	0.49%
		9.28%	metaphorical natural environment	0.35%
		7.22%	boosted circumstances	0.27%
		7.22%	playground outside	0.27%
		6.70%	timing / n	0.25%
		4.64%	green space/ n	0.18%
		3.09%	night lighting	0.12%
		1.03%	n	0.04%

Subjective evaluation	14.65%	36.81%	scores	5.39%
		18.07%	dimesnional look of the accommodation	2.65%
		16.06%	aesthetic look of the accommodation	2.35%
		11.78%	room care	1.73%
		5.76%	negative ambient aspects	0.84%
		4.55%	perceived quality/price ratio	0.67%
		3.61%	kitchen care	0.53%
		1.34%	perception of food serving	0.20%
		1.34%	N	0.20%
		0.67%	nationalist attributes	0.10%
Room facilities	5.27%	27.24%	electronic appliances in the room	1.44%
		17.20%	operating conditions in the bathroom	0.91%
		13.62%	room furniture	0.72%
		13.26%	bathroom equipment	0.70%
		8.24%	negative aspecys in the bathroom	0.43%
		6.09%	n	0.32%
		5.02%	heating	0.26%
		3.58%	bathroom consumables	0.19%
		2.87%	N	0.15%
		2.15%	N	0.11%
		0.72%	N	0.04%

Note: *N, n - undefined*

They identified eight distinct clusters of characteristics that reviewers make reference to, which have different percentages in total. The frequency of mention gives importance in the reviewers' evaluation. Thus, the clusters and percentages in the total are the following: (1) interactions during accommodation - 33.53%; (2) framing into the area – 21.64%; (3) subjective evaluation – 14.65%; (4) food – 7.80%; (5) road – 6.72%; (6) general conditions of accommodation - 6.59%; (7) room facilities – 5.27%; (8) ambient – 3.80%. It is worth discussing the undefined categories. They account for 1.63% of the total of impressions, so even if they are undefined as clustering has not considered an ineffable parameter, the error caused by their presence is relatively small, therefore tolerable. The explicitly positive categories total 5.36% of the total of opinions, and those explicitly negative total 4.07% as a result we have more positive experiences than negative.

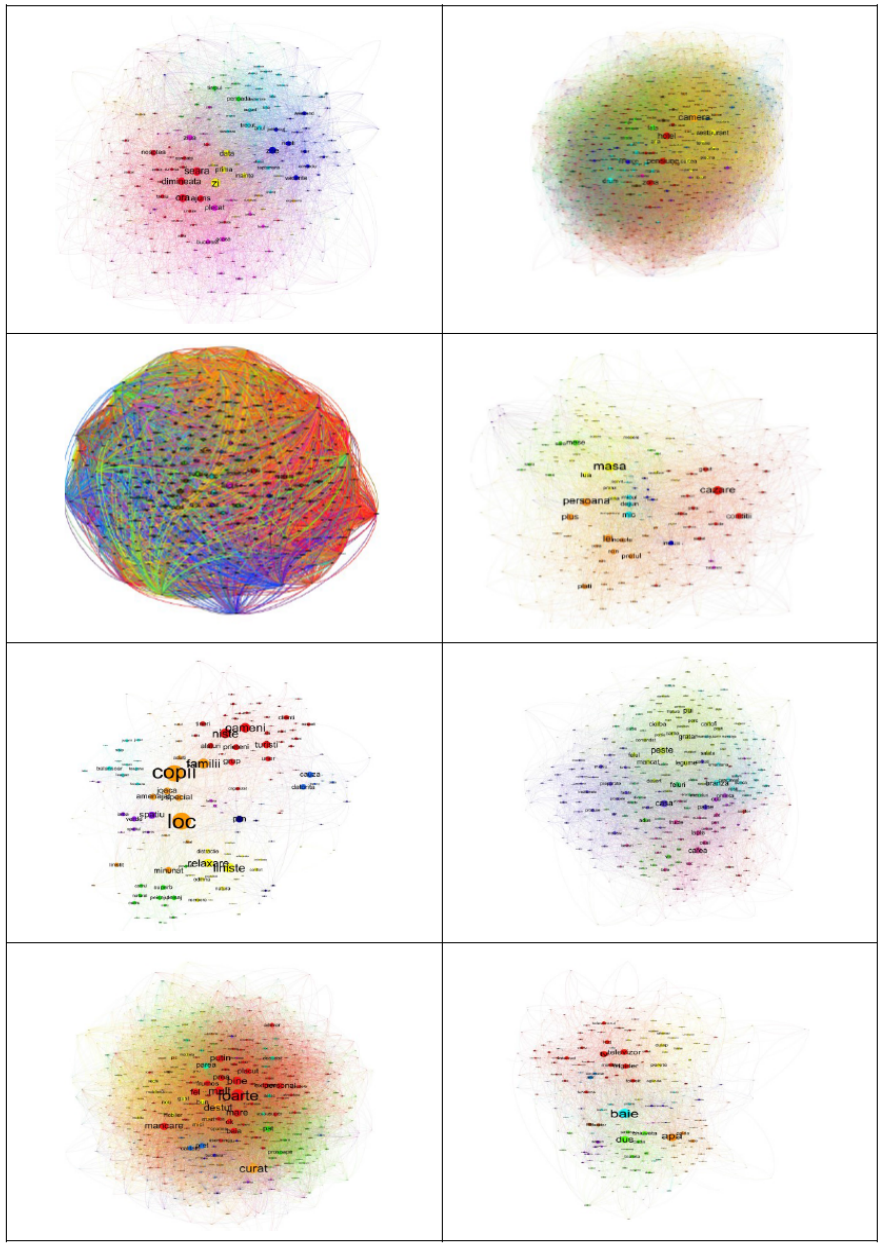

Figure 4. *Communities/Clusters (from left to right) a. 0-Road, b. 1-Description of accommodation compared to the destination area, c. 2- Interaction during accommodation, d. 3- General conditions of accommodation, e. 4-Ambient, f. 5- Food, g. 6-Subjective evaluation, h. 7- Room facilities.*

Conclusions, limitations and further research

The analysis of the 15,200 reviews posted highlighted the aspects / characteristics important for reviewer, heir importance being given by the frequency of mention. Although we are talking about a very large number of reviews difficult to analyze, in which the correlations between words, the meaning of which have been used have been verified etc. certain concrete characterising aspects of the accommodation or stay could be noted. They have been grouped into clusters and the clusters have been named after characterizing the content. Example of characteristics identified (with high percentages in total): the interaction with the people belonging to the accommodation, possible activities, accommodation location in the area, room location, etc. These clusters, in order of importance of mention are: interactions during accommodation, framing into the area; subjective evaluation; food, road; general conditions of accommodation; room amenities and ambient. Getting a reasonable number of clusters increases their usefulness. Although the characteristics directly related to the accommodation service provider are under 20% of total (food, room amenities, general conditions of accommodation) other characteristics that have proven to be priorities for the consumer should be considered. For example, the distinct category - interactions during accommodation- is the most mentioned by the consumers and can be influenced by the accommodation services provider.

The comparison of clusters of characteristics resulted and the evaluation characteristics on theAm Fost Acolo website show a partial overlapping. Admitting that we naming a new category could lead to differences, we verified the correspondence between characteristics and sub/ communities in clusters (with the highest percentages). Thus, we accepted the correlation between general conditions of accommodations and room facilities with the accommodation characteristic. However, the cluster with the highest percentage in the total of reviews - interactions during accommodation (interactions with people belonging to the accommodation) finds no correspondence between characteristics on the website. Although it is a subjective characteristic, seemingly independent from the provider of accommodation services, its high percentage in the ratings draws the attention. The accommodation service provider could create the conditions for positive interaction during stay.

The limitations of the research relate only to the analysis of a single travel website. However, the findings obtained are interesting and support the existence of some evaluation characteristics of accommodation services different in content or importance, which can be drawn from the reviews posted by neural network method. We have in view a research of the comments according to tourist areas (sea, mountains etc.) to check if there are differences in what concerns the characteristics mentioned or their percentage in total.

Acknowledgements

We would like to express our gratitude for the support in this research to Mr. Cornel Bociort, the website developer (amfostacolo.ro). All authors have equal contribution in this research.

References

Barabasi, A.L. (2012). The network takeover. *Nature Physics*, *8*, 14-16.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E. 2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *10* P10008.

Biemann, Ch. (2006). Chinese Whispers: An Efficient Graph Clustering Algorithm and Its Application to Natural Language Processing Problems. *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing*, 73-80.

Blal, I. & Sturman, M.C. (2014). The Differential Effects of the Quality and Quantity of Online Reviews on Hotel Room Sales. *Cornell Hospitality Quarterly*, *55*(4) 365-375.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, *424*(4), 175-308.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*, 345-354.

Choudhury, M. & Mukherjee, A. (2009). The Structure and Dynamics of Linguistic Networks. Dynamics On and Of Complex Networks Modeling and Simulation in Science. *Engineering and Technology*, pp. 1-22., http://research.microsoft.com/pubs/81246/review_of_linguistic_networks.pdf

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, P. & Kuksa, P. (2011). Natural Language Processing (almost) from Scratch. *Journal of Machine Learning Research*, *12*, 2493-2537.

Filieri, R., McLeay, F. (2013). E-WOM and Accommodation: An Analysis of the Factors That Influence Travelers' Adoption of Information from Online Reviews. *Journal of Travel Research*, *53*(1) 44-57.

Gershenson, C. & Niazi, M. A. (2013). Multidisciplinary applications of complex networks modeling, simulation, visualization, and analysis. *Complex Adaptive Systems Modeling, 1*(1), 1-4.

Grabska-Gradzinska, I., Kulig, A., Kwapien, J., & Drozdz, S. (2012). Complex network analysis of literary and scientific texts. *International Journal of Modern Physics*, *23*(7), 1-15.

Hu, N., Liu, L., & Zhang, J. J. (2008). Do online reviews affect product sales? *The role of Hospitality Management*, *2*, 165-177.

Levy, S., W. Duan, & Boo, S. (2013). An analysis of one-star online reviews and responses in the Washington, DC, lodging market. *Cornell Hospitality Quarterly*, *54* (1), 49-63.

Melián-González, S. Bulchand-Giduma1, J. & López-Valcárcel, B.G. (2013). Online Customer Reviews of Hotels: As Participation Increases, Better Evaluation Is Obtained. *Cornell Hospitality Quarterly*, *54*(3), 274-283.

Min, H., Lim, Y. & Magnini, V.P. (2015). Factors Affecting Customer Satisfaction in Responses to Negative Online Hotel Reviews: The Impact of Empathy, Paraphrasing, and Speed. *Cornell Hospitality Quarterly, 56*(2), 223-231.

Mudambi, S., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly*, *34*, 185-200.

Park, S.Y. & Allen, J.P. (2013). Responding to Online Reviews: Problem Solving and Engagement in Hotels. *Cornell Hospitality Quarterly*, *54*(1) 64-73.

Pekar, V & Ou, S. (2008). Discovery of subjective evaluations of product features in hotel reviews. *Journal of Vacation Marketing*, *14*(2), 145-155.

Popescu, A. M. and Etzioni, O. (2005). Extracting Product Features and Opinion from Reviews, *Proceedings of the 2005 Human Language Technology Conference/ Conference on Empirical Methods in Natural Language Processing.*

Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management*, *30*, 123-127.

Ye, Q., Law, R. & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management,* 28, 180-182.

Ye, Q. Li, H., Wang, Z., & Law, R. (2014). The Influence of Hotel Price on Perceived Service Quality and Value in E-Tourism: An Empirical Investigation Based on Online Traveler Reviews. *Journal of Hospitality & Tourism Research*, *38*(1), 23-39.

Yi, J., N., Bunescu, T., R. & Niblack, W. (2003). Sentiment Analyzer: Extracting Sentiments about a Given Topic Using Natural Language Processing Techniques. *Proceedings of the 3rd IEEE International Conference on Data Mining*, pp. 423-434.

Zhang, Z.Q., Ye, Q., Law, R., & Li, Y.J. (2010). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management, 29*, 694-700.

Zou, S.R., Jin, Y., Yang, X.F., Du, Z.J. (2013). Text Research Based on Complex Network, *Proceedings of the 10th Web Information System and Application Conference*, November, pp. 33-37.