Identifying Salient Attributes of Peer-to-Peer Accommodation Experience*

Iis P. Tussyadiah
School of Hospitality Business Management
Carson College of Business
Washington State University Vancouver
Email: iis.tussyadiah@wsu.edu

and

Florian Zach
School of Hospitality Business Management
Carson College of Business
Washington State University Vancouver
Email: florian.zach@wsu.edu

Abstract

This study explores key content and themes from online reviews to explain major service attributes of peer-to-peer (P2P) accommodation sought by guests. The results from lexical analyses indicate that attributes frequently mentioned in guest reviews are associated with location (proximity to point of interest and characteristics of neighborhood), host (service and hospitality), and property (facilities and atmosphere). Reviews focusing on location and feeling welcome are consistently linked with higher rating scores, including accuracy, cleanliness, check-in, communication, value, and overall ratings. This confirms that P2P accommodation appeal to consumers who are driven by experiential and social motivations. Marketing implications are provided.

Keywords: *sharing economy, collaborative consumption, business analytics, accommodation, consumer review, text mining*

---

* **How to cite:** Tussyadiah, I.P., Zach, F. (2016). Identifying salient attributes of peer-to-peer accommodation experience. *Journal of Travel & Tourism Marketing*, Ahead of Print. DOI: 10.1080/10548408.2016.1209153

1. Introduction

In a relatively short period of time, a new wave of businesses utilizing the concept of sharing economy, also known as collaborative consumption (Belk 2014; Botsman and Rogers 2011), has entered the tourism and hospitality marketplace with the introduction of peer-to-peer (P2P) accommodation services. The sharing economy is a socioeconomic system where connected individuals organize the distribution of excess capacity or resources sitting idle in exchange for a fee or other compensation among each other (Belk, 2014). P2P accommodation services such as Airbnb and Roomorama create a platform that enables "regular people" (i.e., as opposed to business entities) to rent out their spare rooms or unoccupied houses and apartments and serve tourists. P2P accommodation services continue to grow at a phenomenal rate. In the summer of 2015, Airbnb served about 17 million guests worldwide, which is a 350% increase from 2010 (Airbnb 2015a). At about 113% year-over-year, the revenue growth of Airbnb is far higher than publicly traded hotel companies, such as Marriott and Wyndham at 8% and 6%, respectively (CB Insights, 2015). Furthermore, Airbnb has more rooms than any branded chain hotels (Freitag and Haywood 2015), making it a formidable competition to hotels. Indeed, based on a study in the state of Texas, US, Zervas, Proservio, and Byers (2015a) found that an increase in Airbnb listing causes a decrease in hotel revenues, with budget hotels and hotels not catering for business travelers being the most affected.

In order to conceptualize and assess the competitive advantages of P2P accommodation in comparison with the conventional accommodation services such as hotels, it is important to explore the aspects of P2P accommodation experience that really matter for guests. Previous studies extract key content from consumer reviews to explain the important attributes of accommodation services and how these attributes contribute to guest satisfaction (e.g., Xiang, Schwartz, Gerdes, and Uysal 2015; Zhou, Ye, Pearce, and Wu 2014). However, it is largely unknown in tourism and hospitality management literature if consumers expect the same aspects of service from P2P accommodation. Also, previous research has suggested various drivers of consumer participation in the sharing economy, including value (i.e., cost-saving), social relationship and sense of community, and authentic experience in non-tourist areas (e.g., Botsman and Rogers, 2011; Guttentag, 2015; Tussyadiah 2016; Tussyadiah and Pesonen, 2015), indicating that what guests seek in P2P accommodation experiences may be different from that of a hotel stay. To that end, the goal of this study is to explore key service characteristics of P2P accommodation emerging from consumer reviews online. Knowing which attributes are important to guests will inform hosts with what areas to pay more attention to in order to attract and satisfy more guests.

In many cities, the growth of P2P accommodation is faced with a variety of legal and regulatory challenges, mostly associated with restrictions for short term rental of residential dwelling units. For example, The City of Berlin in Germany poses hefty fines for users who rent out entire properties on Airbnb in an attempt to safeguard affordable housing for its residents (Kim 2016). Similarly, it has been shown that most Airbnb rentals in New York City, United States (US) violate the short-term leasing law in the State of New York (Gonen and Sutherland 2015). On the contrary, The City of Portland, Oregon, US was among the first to introduce regulations for P2P accommodation rentals, including requirements for business permit and registration, adherence with zoning law, taxes for short term rentals and transient lodging, and property inspection (City of Portland 2015). Due to this major regulatory undertaking, Portland provides a unique context for this study because it “levels the playing field” for commercial and P2P accommodation providers, allowing its residents to serve the tourism market alongside the conventional lodging companies and, at the same time, providing infrastructure to protect both its residents and its visitors. In order to capture potential differences in terms of attributes considered important by guests in different areas, this study also analyzed samples of reviews from listings in New York City, New York, US and London, United Kingdom (UK). The findings from these analyses are included in supplemental materials.

2. Conceptual Framework

2.1 Attributes of Accommodation Services

Identifying operational and marketing dimensions that capture and retain guests and, in turn, inform management decisions to increase profits is critical in hospitality management. A better understanding of the relative importance of various attributes in driving value for guests will allow service providers to allocate resources more efficiently (Mattila 1999), as well as tailor and develop offerings to achieve and maintain the highest possible occupancy (Lockyer 2004). Indeed, various studies suggest that there are different features that guests evaluate and use as decision criteria in hotel selection process (e.g., Clow, Garretson, and Kurtz 1994; Lockyer 2004). Different attributes influence guest satisfaction and post-purchase behavior associated with hotel stay, such as loyalty and electronic word of mouth (eWOM) behavior (e.g., Albayrak and Caber 2014; Xiang et al. 2015; Yen and Tang 2015). Guest decision making, which includes hotel selection, satisfaction, and post-purchase behavior, is a result of cognitive and affective response to hotel attributes (Westbrook 1987) and the overall evaluation of accommodation services is a combination of guest judgments about these different attributes and benefits (Mattila 1999).

Previous studies have identified and proposed different salient hotel dimensions in various research contexts, although common attributes were found in these studies. For example, Knutson (1988) found that business and leisure travelers consider factors of cleanliness and comfort, convenience of location, promptness and courtesy of service, safety and security, and friendliness of employees to select hotels for a first or repeat visit. Rivers, Toh and Alaoui (1991) suggested that convenience of location and overall services are stated as the most important attributes for travelers to select hotels. Examining the factors of guest satisfaction and repeat patronage in Hong Kong, Choi and Chu (2001) suggested factors of staff service quality, room quality, general amenities, business services, value, and security. Studying the topic with participants in New Zealand, Lockyer (2005) identified price, location, cleanliness, facilities and other. A meta-analysis of 21 studies by Dolnicar and Otter (2003) provides a comprehensive look on what guests consider important by grouping 173 identified hotel attributes into the following categories: image, value/price, hotel, room, services, marketing, food and beverage, security, and location. They also identified top attributes from these groups with convenience of location being the most important criterion, followed by service quality, reputation, friendliness of staff, price, room cleanliness, and value for money (Dolnicar and Otter 2003).

For hotel selection, previous studies identified various decision criteria (e.g., Ananth et al. 1992; Atkinson 1988; Lockyer 2004; 2005; Stringam, Gerders, and Vanleeuwen 2010) that Alpert (1971) termed as "determinant attributes". These are features that directly influence purchase intention and differentiate service offerings from competitors. Since intangible cues are very difficult to evaluate prior to patronage, consumers turn their attention to more tangible cues to make purchase decisions and to evaluate past performances (Bitner 1990; Clow et al. 1994). For example, Clow et al. (1994) show that in order to evaluate service quality, consumers refer to their own experiences, staff behavior, price structure, word-of-mouth, and appearance of hotel facility. Additionally, Saleh and Ryan (1992) identified that guests consider tangible, physical components (particularly hotel appearance) as determinant factors in hotel selection.

Although the dimensions included in these studies are varied, attributes for hotel selection and evaluation are well researched. However, knowledge on attributes to evaluate P2P accommodation is extremely limited. While human interactions (i.e., staff recognition, friendliness, attentiveness) have been considered an important hotel attribute, the different roles between hosts and hotel staff as well as the intimacy attached to the sharing practice (i.e., staying at someone's home) highlight the importance of social interactions in P2P accommodation stay. From this point of view, P2P accommodation to some extent can be compared to a type of bed and breakfast (B&B) service, where (commercial) hosts (i.e., innkeepers) offer accommodation with homelike settings. Previous studies identified a high level of satisfaction among guests in B&Bs (e.g., Felix, Broad, and Griffiths 2008; Scarinci and Richins 2008; Zane 1997) and suggested the top qualities of B&B services that guests consider important. These qualities (i.e., service attributes), while include facilities such as availability of private bathrooms, room size, and choice of bed sizes, are centered on hosts who go out of their way to make guests feel comfortable, give recommendation on attractions and restaurants, create homelike atmosphere, and are non-intrusive and respect guests' privacy (Felix, Broad, and Griffiths 2008; Zane 1997). Therefore, it can be suggested that when cohabitation occurs (i.e., hosts and guests stay in the same property) in homelike settings, guests of P2P accommodation would consider the friendliness and attentiveness of hosts as important.

Even though the basic services of P2P accommodation are comparable to hotels and B&Bs (i.e., room and board), P2P accommodation is characterized by a lack of standards (i.e., absence of star ratings or quality classification). Guests can choose three types of accommodation through Airbnb: an entire house/apartment, a private room (often with shared facilities), or a shared room. The features of these listings vary greatly (e.g., shared or private bathroom, kitchen, internet access, and other room amenities). Therefore, it is important to explore which features really matter for guests when evaluating P2P accommodation in order to better understand the factors that differentiate P2P accommodation from more established accommodation offerings, including hotels and B&Bs. While still in its infancy, recent studies on collaborative consumption in hospitality have suggested cost-savings (i.e., value for money) and social motivations (e.g., desire for community and social interactions) to drive the use of P2P accommodation (Guttentag 2015; Tussyadiah 2015; Tussyadiah and Pesonen 2015). Based on a questionnaire distributed to Airbnb users (including both hosts and guests), Möhlmann (2015) identified cost-savings, familiarity (with the system), trust (amongst users and toward the system), and utility (quality of service as compared to other alternatives) to be the significant factors of satisfaction in P2P accommodation marketplaces. In terms of providers, Karlsson and Dolnicar (2016) suggested three motivations for participation in P2P accommodation: income (80%), social interaction (31%), and sharing (14%). Finally, based on a questionnaire responded by P2P accommodation guests, Tussyadiah (2016) identified the factors of enjoyment, amenities, and cost-savings as positively influence satisfaction (in order of significance), with social benefits only found significant among those who rent private rooms. However, location, which is one of the most important hotel attributes (e.g., Lockyer 2005; Rivers, Toh and Alaoui 1991), was not significant in influencing guests' satisfaction or behavioral intention to use P2P accommodation (Tussyadiah 2016). The limited literature on evaluation criteria for P2P accommodation is one of the motivations for this research.

2.2 User-Generated Content and Text Analytics

Previous studies apply different methodologies to assess the relative importance of hotel attributes among consumers, many focusing on importance ratings of different attributes through interviews with and questionnaires distributed to consumers (e.g., Clow et al., 1994). More recently, through development in mobile devices and social network technologies consumers leave traces of their consumption patterns online through pictures, check-ins, statuses, and reviews. Lipsman (2007) suggests that more than 87% of consumers rely on online user-generated content (UGC) to make purchase decisions for hotels. UGC, when appropriately managed and analyzed, mounts to significant consumer intelligence valuable for tourism and hospitality businesses. Indeed, business intelligence and analytics, and the related field of big data analytics, are considered critical in providing market insights and competitive analyses to assist business managers in making timely decisions (Chen, Chiang, and Storey 2012). Therefore, UGC provides opportunities for tourism and hospitality decision makers to gain actionable insights on factors of guest experiences and satisfaction.

As a form of electronic word of mouth (eWOM), online hotel reviews are valuable in predicting booking intention and guest satisfaction (e.g., Stringam et al., 2010; Tsao, Hsieh, Shih, and Lin 2015; Xiang et al. 2015; Zhou et al. 2014). By extracting frequently discussed attributes in online reviews, UGC can reveal the influence of different dimensions of hotel services on purchase decisions and evaluation. Several studies indicate the usefulness of analyzing UGC to create knowledge and recognize patterns. Based on TripAdvisor reviews on hotels in Hong Kong, Li, Law, Vu, and Rong (2013) suggested six hotel selection criteria (value, location, sleep, room, cleanliness, and service) and demonstrated that evaluation differs based on travel type and traveler origin. Ramanathan and Ramanathan (2011) used online reviews to examine UK hotel performance and identified customer service, room quality, and quality of food as dissatisfiers. Most recently, analyzing 60,648 hotel reviews, Xiang et al. (2015)

identified six dimensions in hotel reviews (hybrid, deals, amenities, family friendliness, core product, and staff) with varying degrees of influence on satisfaction, as measured through star ratings.

Key to analyzing UGC is to extract valuable nuggets of information and patterns from relatively large, highly unstructured human-authored text data. Manually scanning and analyzing such data is considered impractical for business decisions due to high computational burden. Advances in computer science, especially in machine learning and natural language processing (NLP) resulted in text mining techniques (also known as text analytics or knowledge discovery from textual database) that effectively extract knowledge from natural (human) language text documents. The overarching goal of text mining is to turn (unstructured, often messy) text into (structured, organized, labeled) data for analysis through application of NLP and other statistical models so high quality (interesting, novel, relevant, and non-trivial) information can be extracted from the data. It typically involves such tasks as text categorization, text clustering, concept extraction, and document summarization. Text analytics target an automatic extraction of features from single documents and analyze feature distribution over the collection of documents to detect patterns and trends (Dörre, Gerstl, and Seiffert 1999).

In tourism and hospitality marketing and management, text mining techniques can be valuable in handling voluminous text available online from online reviews, blogs, tweets, discussion forums, etc. As travel consumers leave their traces online during various travel experiences from information search to reflection (e.g., consumer feedback), a large amount of knowledge about tourist behavior and perception is available for tourism destinations and hospitality businesses (Fuchs, Höpken, and Lexhagen, 2014). Text mining techniques are applicable to extract important features (e.g., accommodation attributes, customer satisfaction factors) and surrounding patterns and trends in order to better understand what consumers want as well as the strengths and weaknesses of competitors.

3. Method

This study analyzes online reviews of Airbnb listings to extract salient attributes of P2P accommodation. The underlying assumption in this study is that attributes most frequently mentioned in guest reviews are indicative of guest satisfaction factors. Data were obtained from a third-party website, *insideairbnb.com* (Inside Airbnb, 2015), published under a Creative Commons Zero CC0 1.0 Universal (CC0 1.0) “Public Domain Dedication” (Creative Commons, n.d.). The dataset contains information regarding property listings in Portland, Oregon, which was sourced by Inside Airbnb from publicly available information on Airbnb.com website on May 12, 2015. After eliminating cases with missing information and reviews written in languages other than English, 41,560 reviews from 1,617 property listings were included in the study. On average, each review contains five sentences (i.e., a total of 215,497 sentences in the dataset). Data management and analyses followed several steps, which include preprocessing, lexical analysis, and visualization, facilitated by text mining software called *KH Coder* (Higuchi, 2015).

3.1 Data Preprocessing and Descriptive Statistics

To prepare the textual data for further analysis, the dataset was preprocessed following these procedure (see an example in Exhibit 1): (1) tokenization (i.e., breaking a stream of text into words, phrases, symbols, and other meaningful elements called tokens), (2) eliminating stop words (i.e., removing frequently occurring non-context-bearing, common words, such as definite or indefinite articles and auxiliary verbs, including “a,” “an,” “and,” “the,” etc.), (3) part-of-speech (POS) tagging (i.e., assigning parts of speech to each word, such as noun, verb, adjective, etc., based on both its definition and its context), and (4) lemmatization (i.e., conflating tokens to their root form, such as “staying” and “stayed” into “stay”). The preprocessing was conducted using *Stanford POS Tagger* program, a Java implementation of the log-linear POS tagging approach as described in Toutanova, Klein, Manning, and Singer (2003). The dataset contains 3,530,597 tokens and 33,059 word types. After exclusion of stop words, 1,473,197 tokens and 21,561 word types (i.e., representative terms) remained for analysis. Figure 1 illustrates the distribution of the term frequency (TF, number of occurrences of words) in the dataset. The mean TF is 45.57 (i.e., words appear 45 times on average) with a standard deviation of 484.58. As represented by the long tail in the distribution plot, about 99% of words appear less than 1,000 times in the dataset.

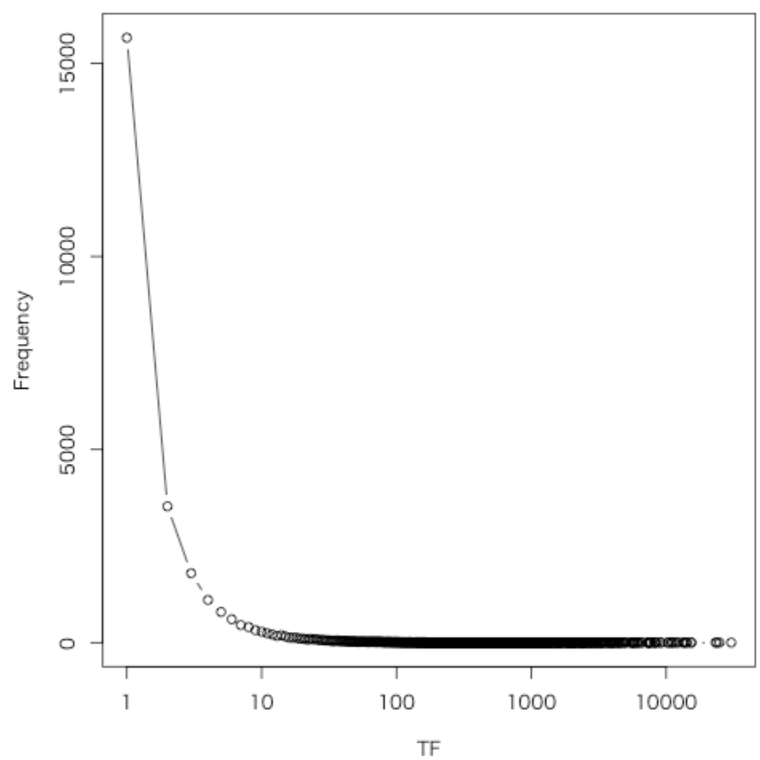


Figure 1. Term Frequency (TF) Distribution

3.2 Automated Term Recognition

To obtain terms that are most relevant to P2P accommodation experience from the dataset, important compound words (*N*-grams), which could be a two-word combination (bi-gram), a three-word combination (tri-gram), etc., were identified using an automated term recognition (ATR) approach as explained in Nakagawa (2000) and Nakagawa and Mori (2002), facilitated by *TermExtract* module in *KH Coder* program. It applies a termhood-based approach to measure the extent to which a candidate term *(ct)* is related to a domain-specific context (Korkontzelos, Klapaftis, and Manandhar 2008), assuming that terms with complex structure are made of simple terms (Nakagawa 2000; Nakagawa and Mori 2002). Therefore, it measures the termhood of single tokens first and then uses it to measure the termhood of complex terms. Let *R(N)* and *L(N)* be two functions that calculate the number of distinct words that adjoin *N* or *N* adjoins, respectively. For each candidate term $ct = N_1, N_2 \dots N_k$, an importance score (*IMP*) is calculated by:

$$IMP(ct) = \left(\prod_{i=1}^{k}\left((R(N_i)+1)*(L(N_i)+1)\right)\right)^{1/2k} \quad (1)$$

It can be interpreted that the higher the importance score of a term in this dataset, the more relevant it is to characterize P2P accommodation experience.

3.3 Word Co-occurrence Network

The distribution of high frequency keywords in the dataset was examined by developing word co-occurrence networks to identify how words were used in connection with each other in one review. Each network consists of nodes and edges that connect the nodes. Nodes are frequently mentioned words. Edges are determined by the Jaccard Coefficient of word pairs. Jaccard Coefficient is a statistical measure to compare the similarity between finite sample sets, which is defined as the size of the intersection divided by the union of the sample sets. The Jaccard Coefficient of a word pair *A* and *B* is:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

The layout of the networks is determined by the Fruchterman–Reingold's (1991) algorithm, which uses a force-based graph drawing technique to present networks in an aesthetically pleasing way. The co-occurrence networks were developed using the *igraph* package in the *R* statistical program.

3.4 Cluster Analysis

Agglomerative hierarchical cluster analysis was conducted to partition the dataset into meaningful and coherent groups of similar reviews. Words with higher probability of appearance in a specific cluster (as indicated by its conditional probability score) give the cluster its distinctiveness from the rest. Reviews containing similar sets of words typically discuss the same topic. Therefore, in this study, cluster analysis assists in identifying groups of reviews that discuss similar topics, representing attributes of P2P accommodation. In order to consider whether reviews are similar or different, this study uses Jaccard Distance as a distance measure, which compares the sum weight of shared terms to the sum weight of terms that are present in either of the two documents (i.e., reviews) but are not the shared terms. Jaccard Similarity (SIM_J) of two reviews d_a and d_b, represented by their term vectors $\overrightarrow{t_a}$ and $\overrightarrow{t_b}$, is:

$$SIM_J\left(\overrightarrow{t_a}, \overrightarrow{t_b}\right) = \frac{\overrightarrow{t_a} \cdot \overrightarrow{t_b}}{\left|\overrightarrow{t_a}\right|^2 + \left|\overrightarrow{t_b}\right|^2 - \overrightarrow{t_a} \cdot \overrightarrow{t_b}} \tag{3}$$

Jaccard Distance (D_J) or dissimilarity between the two reviews is:

$$D_J(\overrightarrow{t_a}, \overrightarrow{t_b}) = 1 - SIM_J(\overrightarrow{t_a}, \overrightarrow{t_b}) \tag{4}$$

Each review is initially assigned to its own cluster, producing a set of singleton clusters, and then the algorithm proceeds iteratively by merging the two most similar clusters at each stage until there is just a single cluster. The agglomerative method is Ward's (1963) linkage criterion (i.e., Ward's minimum variance method), which uses the error sum of squares to merge the pair of clusters. For optimum inclusion of data and interpretability of results, this study used words that appear at least 4,000 times in the dataset, generating a matrix with 51 columns and 41,560 rows. The cluster analysis was conducted using the *hclust* package in the *R* statistical program, resulting in each review being a member of a specific cluster.

3.5 Influences of Reviews on Ratings

Regression analyses were conducted using the SPSS statistical program to identify if different attributes of P2P accommodation mentioned in reviews influence rating scores. It is important to note that, unlike other online review platforms where individual rating scores are made visible for each review (i.e., a direct association between review and rating score can be made), only aggregate rating scores for each property are made available on Airbnb website. Hence, no direct association between each review and rating score given by the reviewer could be made. Therefore, this study uses the share of cluster membership (i.e., the number of reviews belonging to Cluster 1, Cluster 2, etc.) at the property level as independent variables and rating scores of the property as dependent variables in the regression models.

4. Results and Discussion

4.1 High Frequency Keywords

The 20 most frequent keywords for nouns (representing attributes), adjectives (representing assessment), and verbs (representing activities) were extracted from the dataset (see Table 1). Top nouns include general descriptors of the listing (e.g., "place," "house," "home") and host, specific attributes of the property (e.g., “location,” “room,” “bed”), and descriptors of overall experience (e.g., “stay,” “time,” “experience”). Top adjectives represent a positive evaluation directed towards physical property and room (e.g., “clean,” “comfortable,” “cozy”), host (e.g., “friendly,” “helpful”), and overall experience (e.g., “great,” “wonderful”). Top verbs represent guest actions (e.g., “stay,” “enjoy,” “arrive”) and host actions (e.g., “welcome,” “recommend,” “accommodate”).

Table 1. Word Frequency Lists (Top 20)

No.	Nouns	Freq.	Adjectives	Freq.	Verbs	Freq.
1	place	23770	great	30387	stay	24827
2	house	15400	comfortable	13946	make	11364
3	host	15255	clean	13277	recommend	10348
4	stay	14259	nice	10222	feel	8542
5	home	13603	easy	8133	need	8242
6	neighborhood	12222	perfect	7639	walk	7268
7	location	11865	wonderful	7357	enjoy	6583
8	room	10973	beautiful	6082	love	6576
9	time	10918	quiet	5921	welcome	5585
10	apartment	9141	lovely	5691	come	4481
11	space	7642	good	5628	accommodate	4005
12	bed	7556	friendly	5589	look	3928
13	restaurant	7378	helpful	5054	provide	3851
14	experience	6228	cozy	4724	visit	3810
15	area	6099	little	4433	want	3749
16	downtown	5808	super	4405	locate	3452
17	night	5431	amazing	3630	leave	3380
18	lot	4959	spacious	3094	use	3268
19	day	4828	close	3071	meet	3230
20	coffee	4600	warm	3069	arrive	2808

The top 50 compound words (i.e., domain specific terminologies) in the dataset ranked by their importance score (see Equation 1) are presented in Table 2. The list is dominated by positive assessment of property (e.g., “great place,” “beautiful home”), host (e.g., “great host(s),” “wonderful hosts”), and experience (e.g., “great time,” “great stay,” “wonderful experience”). Additionally, specific attributes emerged from the list, such as location (e.g., “great location,” “downtown portland,” “quiet neighborhood”), convenience (i.e., “easy access,” “public transportation," "short walk"), facilities in the neighborhood (i.e., "great restaurants," "coffee shops"), and amenities (i.e., "comfy bed," "living room"). All but one (#15, "first Airbnb experience") of the top 50 compound words are bigrams, mostly in the forms of adjective–noun and noun–noun combinations.

5 Table 2. Top 50 Compound Words

No	Compound Words	No	Compound Words
1	great location	26	nice neighborhood
2	great host	27	nice place
3	great place	28	lovely home
4	great time	29	living room
5	great neighborhood	30	wonderful place
6	great stay	31	good location
7	downtown portland	32	great recommendations
8	great experience	33	one night
9	next time	34	coffee shops
10	great hosts	35	perfect place
11	great restaurants	36	great area
12	first time	37	wonderful hosts
13	wonderful host	38	short walk
14	quiet neighborhood	39	great coffee
15	first airbnb experience	40	perfect location
16	portland area	41	wonderful experience
17	wonderful stay	42	easy access
18	great food	43	excellent host
19	wonderful time	44	first experience
20	beautiful home	45	great places
21	alberta street	46	comfy bed
22	public transportation	47	street parking
23	great space	48	great spot
24	great house	49	beautiful neighborhood
25	comfortable bed	50	beautiful house

5.1 Word Co-Occurrence Network

The top 500 pairs of high frequency words based on their Jaccard Coefficients (as a measure of similarity) are presented in a word co-occurrence network (Figure 2). The minimum term frequency (TF) was set to 4,000, allowing the network to be developed from a 51 (words) x 41,560 (reviews) matrix. Size of nodes indicates word frequency; thickness of edges indicates strength of connections (i.e., extent of similarity) between word pairs; color indicates word communities in the network (i.e., densely connected sub-graphs), which are detected using random walk method (Pons and Latapy 2005). The result shows that 48 of the selected 51 words make the top 500 word pairs, indicating that they are not only important because they are frequently mentioned individually, but also in combination with each other. The network density is .443, which means that 44.3% of all possible edges are present in the network.

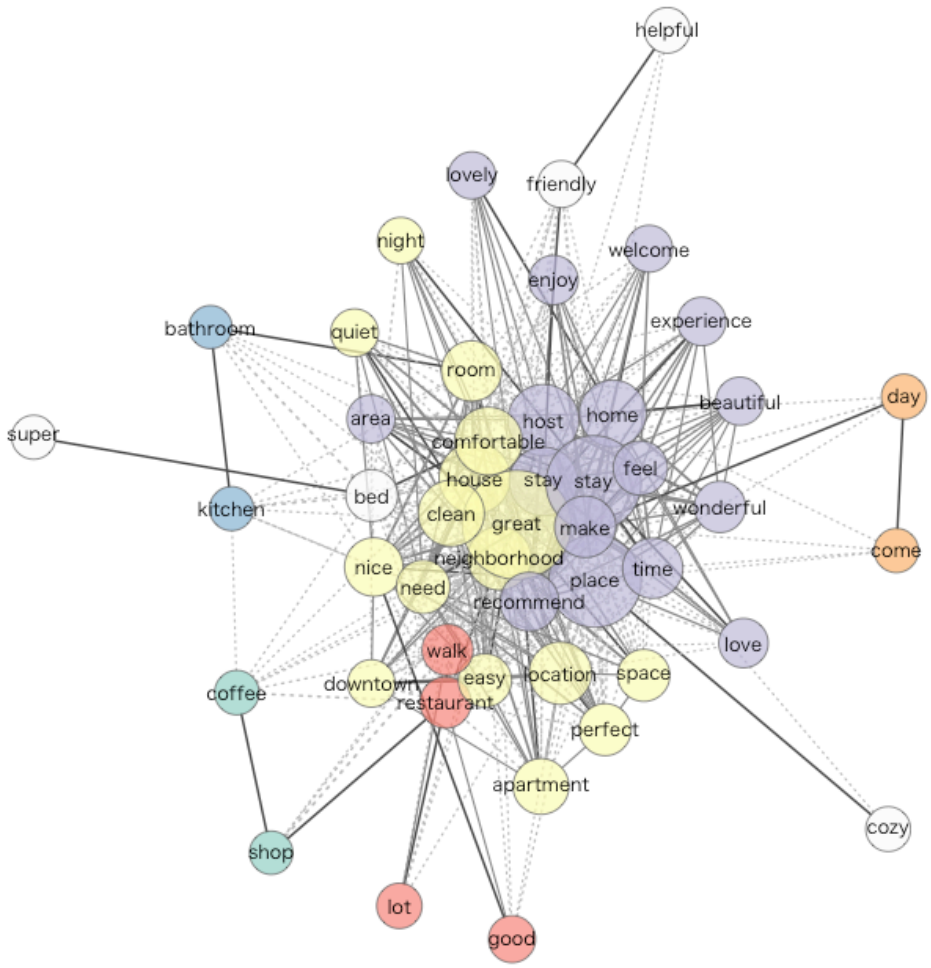


Figure 2. Word Co-Occurrence Network

Two large word communities dominate the core of the network: one centered on “stay” and “place” (representing overall staying experience) and another on “house” and "neighborhood" (representing specific assessment regarding the property and its location). Both are strongly connected with the word "great," the largest node (i.e., the most frequently mentioned word) in the network. These two word communities suggest that location and (interactions with) host are considered important attributes of P2P accommodation experiences. These attributes (and overall evaluation) are common to most reviews, as represented by the large node sizes (most frequently mentioned) and thick edges (frequently mentioned together). The smaller word communities capture more specific attributes, such as facilities and amenities, which are not linked to each other (i.e., these attributes appear together less frequently in the reviews). This indicates that while the majority of reviews contain general evaluation and locational attributes, they emphasize different features or characteristics that are most relevant to the guest experience. To better understand these different attributes, it is important to identify groups of reviews that share similar topics.

5.2 Review Clusters

Hierarchical cluster analysis produced five review clusters that identify different themes: Cluster 1: Service (19,463 reviews), Cluster 2: Facility (4,353 reviews), Cluster 3: Location (6.890 reviews), Cluster 4: Feel Welcome (3,451 reviews) and Cluster 5: Comfort of a Home (7,263 reviews). Only a small fraction of reviews in the document (less than one percent of the data) does not belong to any of the clusters. As presented in Table 3, a majority of reviews in Service and Location (65% and 60%, respectively) were written for experiences staying in an entire home/apartment, while reviews in Facility, Feel Welcome, and Comfort of a Home were evenly distributed for experiences in an entire home/apartment and a private room. Reviews for experiences in a shared room are extremely small (one to two percent of all reviews).

Table 3. Review Cluster and Room Type

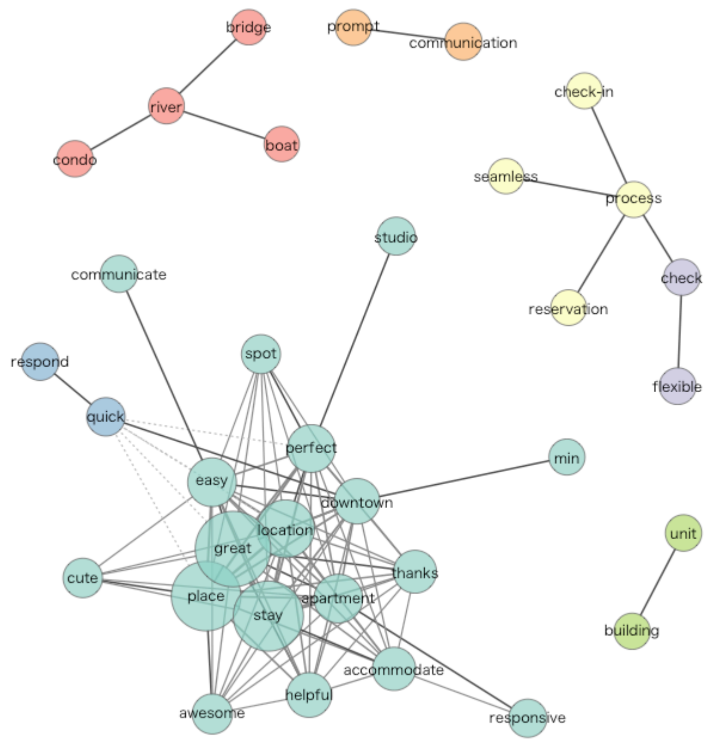
Cluster	Entire Home/Apt		Private Room		Shared Room		Total
	Count	Percent	Count	Percent	Count	Percent	
Service	12615	65%	6608	34%	240	1%	19463
Facility	1768	41%	2542	58%	43	1%	4353
Location	4142	60%	2699	39%	49	1%	6890
Feel Welcome	1621	47%	1774	51%	56	2%	3451
Comfort of a Home	3729	51%	3441	47%	93	1%	7263
Total	23875	58%	17064	41%	481	1%	41420

Service. Reviews in this cluster contain words that describe hosts' service characteristics. In particular, as can be seen in Table 4 and Figure 3 (top 100 word pairs), the descriptors of this review group include communication (i.e., easy communication with hosts), responsiveness (i.e., hosts respond quickly to inquiries, prompt communication), reservation process, and check-in and check-out time. In the word co-occurrence network, the largest word community also includes assessment of overall experience and location. Important differentiators are communities linked to the core as well as isolate communities around responsiveness (i.e., prompt communication) and convenience (i.e., reservation, flexible check-in).

Table 4. Top 20 Keywords in Service Cluster

No	Word	Jaccard[1]	Cond. Prob.[2]	No	Words	Jaccard[1]	Cond. Prob.[2]
1	great	.34	.53	11	thanks	.09	.10
2	place	.32	.45	12	awesome	.06	.07
3	stay	.30	.46	13	cute	.05	.06
4	location	.23	.30	14	spot	.05	.05
5	apartment	.16	.18	15	quick	.04	.04
6	easy	.15	.18	16	responsive	.04	.04
7	perfect	.14	.17	17	studio	.04	.04
8	downtown	.12	.14	18	respond	.03	.03
9	helpful	.11	.13	19	communication	.03	.03
10	accommodate	.10	.11	20	communicate	.03	.03

Note: [1]*Jaccard Coefficient (similarity) with this cluster*
[2]*Conditional probability of word to belong to this cluster*



Nodes: 34; Edges: 100; Density: .178

Figure 3. Word Co-Occurrence Network in Service Cluster

In previous research on hotel attributes, service quality is consistently proposed as an important attribute of guest evaluation and typically attached to staff encounters (e.g., Choi and Chu 2001; Knutson 1988; Li et al., 2013; Shergill and Sun 2004). Hotel guests appreciate promptness and courtesy of service from hotels (Knutson 1988) and quality standards at check-in and check-out (Dubé and Renaghan 1999), which is consistent with the theme in this cluster. In P2P accommodation, however, there is a role shift in service delivery from hotel employees to individual hosts. While the contact platform (e.g., Airbnb) can set quality standards for some service processes (e.g., ease of booking), hosts have a crucial role in delivering the service quality that meets guest expectation (i.e., promptness and flexibility).

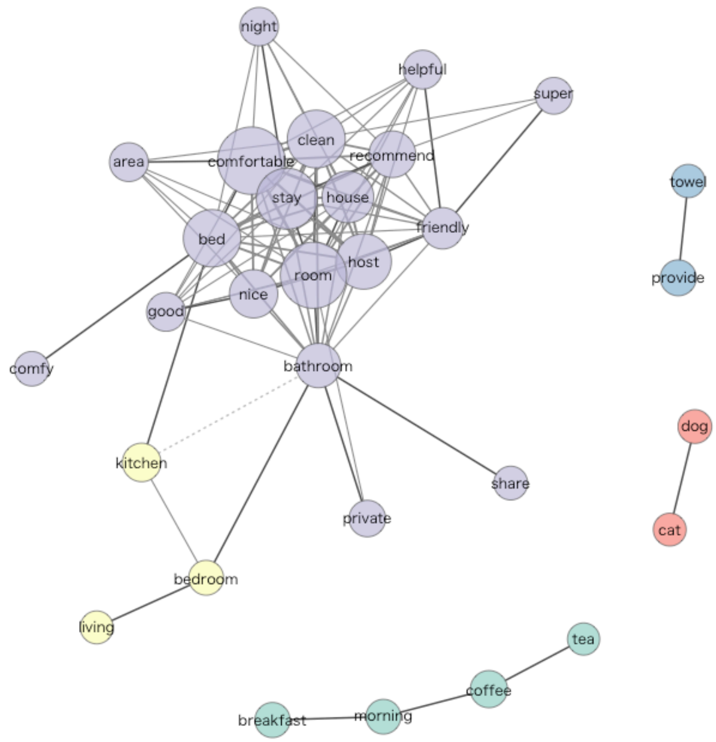
Facility. Descriptors in this cluster focus on physical aspects of the property, including space (e.g., "room", "bathroom", "kitchen"), supporting goods (e.g., "bed", "towel"), and other amenities. As presented in Table 5 and Figure 4 (top 100 word pairs), frequently mentioned terms include clean room, comfortable bed, and private (or shared) bathroom, which are located in the largest word community within the network. A few isolate word communities revolve around breakfast (i.e., "coffee," "tea") and pet (i.e., "dog," "cat").

Table 5. Top 20 Keywords in Facility Cluster

No	Word	Jaccard[1]	Cond. Prob.[2]	No	Words	Jaccard[1]	Cond. Prob.[2]
1	room	.22	.55	11	recommend	.08	.25
2	bed	.19	.42	12	kitchen	.07	.13
3	comfortable	.17	.57	13	helpful	.07	.13
4	bathroom	.13	.22	14	good	.07	.13
5	clean	.12	.43	15	night	.06	.13
6	nice	.10	.28	16	area	.06	.13
7	house	.10	.33	17	private	.06	.10
8	host	.10	.39	18	coffee	.06	.11
9	stay	.10	.48	19	spacious	.06	.09
10	friendly	.08	.17	20	super	.06	.10

Note: [1] *Jaccard Coefficient (similarity) with this cluster*
[2] *Conditional probability of word to belong to this cluster*

Previous studies on hotel evaluation also cited physical property and facilities aspects (Lockyer 2005), room (Choi and Chu 2001; Li et al. 2013), cleanliness (Knutson 1988; Li et al. 2013; Saleh and Ryan 1992), and comfort (Knutson 1988) as selection criteria influential for guest satisfaction and return intention. For hotel services studies differentiate between core (basic) services, general amenities, and add-on services (e.g., deals, freebies) (e.g., Saleh and Ryan 1992; Xiang et al. 2015). Due to the shared and residential nature of most P2P accommodations, this cluster revealed that guests value clean rooms and comfortable beds and appreciate add-ons stemming from typical hotel experiences: private bathroom, access to (use) kitchen, hosts providing towels and toiletries and tea or coffee for breakfast. Compared to findings from previous studies on expectations of B&B guests with regards to amenities, private bathroom was stated as most important (Felix, Broad and Griffiths 2008; Zane 1997). Although, B&B guests also expect other amenities such as fireplace, TV, cooked meals for breakfast, etc. that are perceived as influential to satisfaction (Felix, Broad and Griffiths 2008; Scarinci and Richins 2008; Zane 1997).



Nodes: 30; Edges: 100; Density: .230

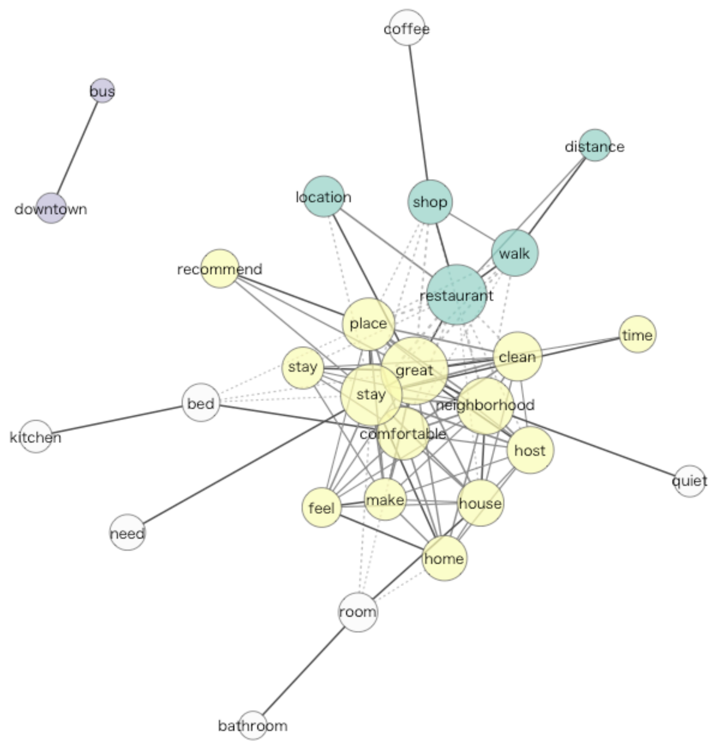
Figure 4. Word Co-Occurrence Network in Facility Cluster

Location. This cluster is dominated by descriptions of location and characteristics of the neighborhoods where the properties are located. Reviews in this cluster also highlight locational advantages in terms of proximity to other points of interest (e.g., distances to shops and restaurants) and transportation convenience (i.e., walking distance, access to public transit) (see Table 6). Figure 5 displays the word co-occurrence network within this cluster (top 100 word pairs). The two largest word communities in the network explain the property (i.e., “house,” “home,” “place”) as part of a nice neighborhood and the amenities that the neighborhood provides (e.g., “restaurants”, “shops”). An isolated community refers to convenience of having public transportation (“bus”) to explore downtown.

Table 6. Top 20 Keywords in Location Cluster

No	Word	Jaccard[1]	Cond. Prob.[2]	No	Words	Jaccard[1]	Cond. Prob.[2]
1	restaurant	.34	.51	11	stay	.16	.53
2	shop	.29	.35	12	distance	.16	.22
3	walk	.24	.37	13	kitchen	.16	.23
4	neighborhood	.23	.48	14	bed	.16	.28
5	coffee	.19	.26	15	clean	.16	.40
6	comfortable	.18	.44	16	make	.16	.33
7	home	.17	.36	17	room	.15	.30
8	great	.17	.59	18	easy	.15	.26
9	feel	.17	.30	19	need	.15	.26
10	house	.17	.37	20	nice	.15	.28

Note: [1]*Jaccard Coefficient (similarity) with this cluster*
[2] *Conditional probability of word to belong to this cluster*



Nodes: 28; Edges: 100; Density: .265

Figure 5. Word Co-Occurrence Network in Location Cluster

Convenience of location has been consistently reported as one of the most important attributes for hotels (e.g., Ananth et al. 1991; Dolnicar and Otter 2003; Knutson 1988; Lockyer 2005; Rivers, Toh, and Alaoui 1991) as well as B&Bs (e.g., Scarinci and Richins 2008). In these studies, hotel location is often assessed as distances to city center, central business district, or main attractions. However, many P2P accommodation listings are located outside of traditional tourist (hotel) areas (Airbnb 2015b). This implies the crucial role of neighborhood characteristics in addition to convenience in P2P accommodation evaluation. That is, since most P2P accommodations are not located in tourist areas, the vitality of the neighborhoods where these properties are located (abundance of restaurants and coffee shops within walking distance) becomes important for guest satisfaction. This also confirms previous studies suggesting that staying at P2P accommodation offers the experiential value of being in authentic, non-tourist settings (Guttentag 2015; Möhlmann 2015).

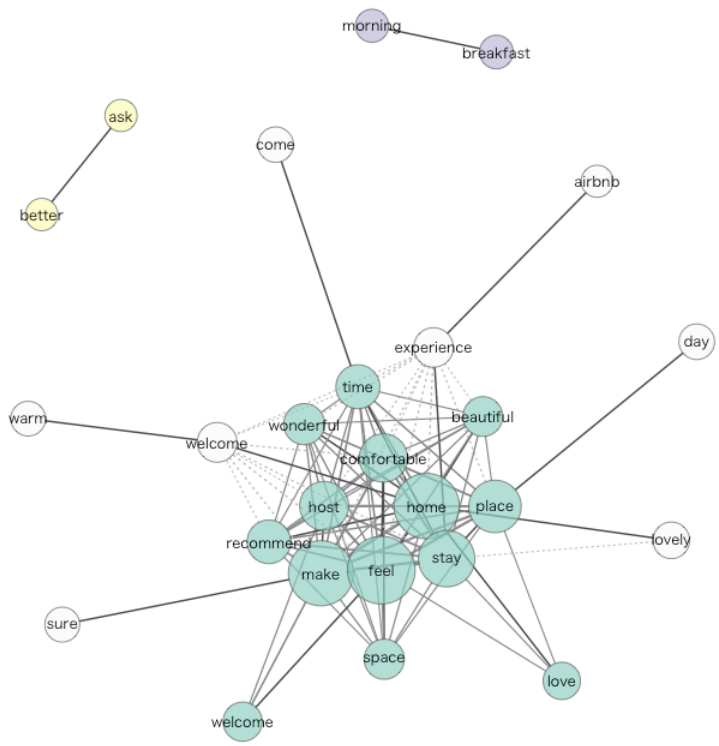
Feel Welcome. The main descriptors in this cluster relate to feeling "at home" while staying in a P2P accommodation. Reviews in this cluster emphasize social interactions with hosts and hosts' efforts to welcome and accommodate guests (see Table 7). Figure 6 illustrates the word co-occurrence network for this review cluster (top 100 word pairs). The main word community describes how hosts make guests feel welcome. Noteworthy adjectives (e.g., "lovely," "warm," "wonderful") support the positive emotions expressed by guests to describe their experience.

Table 7. Top 20 Keywords in Feel Welcome Cluster

No	Word	Jaccard[1]	Cond. Prob.[2]	No	Words	Jaccard[1]	Cond. Prob.[2]
1	feel	.29	.70	11	place	.08	.42
2	make	.21	.65	12	space	.07	.20
3	home	.20	.66	13	experience	.07	.18
4	welcome	.13	.18	14	host	.07	.36
5	stay	.08	.49	15	recommend	.07	.26
6	comfortable	.08	.35	16	sure	.06	.10
7	wonderful	.08	.21	17	love	.06	.14

No	Word	Jaccard[1]	Cond. Prob.[2]	No	Words	Jaccard[1]	Cond. Prob.[2]
8	time	.08	.26	18	trip	.05	.10
9	welcome	.08	.19	19	city	.05	.10
10	beautiful	.08	.19	20	lovely	.05	.13

Note: [1]*Jaccard Coefficient (similarity) with this cluster*
[2] *Conditional probability of word to belong to this cluster*



Nodes: 26; Edges: 100; Density: .308

Figure 6. Word Co-Occurrence Network in Feel Welcome Cluster

Studies found that hotel guests value courteous, friendly, and helpful staff or personnel as important satisfaction criteria (Clow, Garretson, and Kurtz 1994; Dolnicar and Otter 2003; Knutson 1988). Similar to the role of individual hosts in delivering services in the Service cluster, hosts play a significant role in fulfilling customer expectation for courteous services prior to arrival, during stay, and after departure. Reviews in this cluster particularly focus on direct guest – host interactions during the stay. That is, the nature of P2P accommodation stays (51% in this cluster involve co-habitation), requires hosts to go beyond being courteous and helpful in a professional sense (i.e., staff – customer relations) by maintaining positive interactions in a social sense (i.e., host – guest relations). This is consistent with the expectation of B&B guests for friendly and welcoming hosts who create the atmosphere that is inviting (Felix, Broad and Griffiths 2008; Scarinci and Richins 2008; Zane 1997). This also supports the stipulation that P2P accommodation appeals to customers due to its social dimensions (i.e., making friends, developing social relationships with hosts) (Guttentag 2015; Tussyadiah 2015; Tussyadiah and Pesonen 2015).

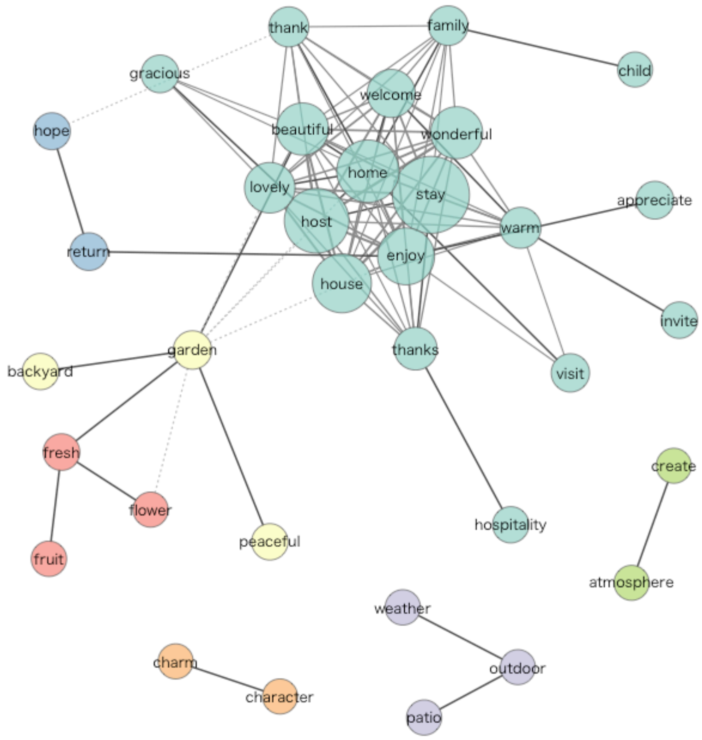
Comfort of a Home. The last cluster contains reviews emphasizing the uniqueness of enjoying the comfort of a home, including a homey and peaceful atmosphere and the hospitality of the hosts and their family. While the aspect of comfort is also discussed in reviews within the Facility cluster (e.g., comfortable bed), reviews in this cluster describe the general ambiance of staying at a residential property in a residential neighborhood. As presented in Table 8 and Figure 7 (top 100 word pairs), reviewers highlight the charm of a beautiful home, full of characters, with a backyard, garden, and outdoor patio. Adjectives used in this cluster (“peaceful,” “lovely,” “warm”) capture the uniqueness of the servicescape that P2P accommodation provide and guests appreciate from this type of accommodation (i.e., “appreciate,” “thank”). Finally, a word community connected to the core describes guests’ intention to return to the property (i.e., “hope,” “return”), making home atmosphere an important factor of return patronage for P2P properties.

Table 8. Top 20 Keywords in Comfort of a Home Cluster

No	Word	Jaccard[1]	Cond. Prob.[2]	No	Words	Jaccard[1]	Cond. Prob.[2]
1	stay	.23	.50	11	warm	.06	.08
2	enjoy	.17	.27	12	thank	.05	.07
3	home	.16	.33	13	family	.05	.06
4	host	.14	.36	14	garden	.04	.05

5	beautiful	.13	.21	15	visit	.04	.05
6	house	.13	.29	16	gracious	.03	.04
7	lovely	.12	.19	17	appreciate	.03	.04
8	wonderful	.12	.21	18	return	.03	.04
9	welcome	.10	.16	19	pleasant	.03	.03
10	thanks	.07	.10	20	cat	.03	.03

Note: [1]*Jaccard Coefficient (similarity) with this cluster*
[2] *Conditional probability of word to belong to this cluster*



Nodes: 34; Edges: 100; Density: .178

Figure 7. Word Co-Occurrence Network in Comfort of a Home Cluster

While this is consistent with the studies on B&Bs (e.g., Felix, Broad and Griffiths 2008; Scarinci and Richins 2008), the theme surrounding the general atmosphere of a hotel has not been the subject of many hotel selection and satisfaction criteria studies, with some exception (e.g., Dubé and Renaghan 1999). Notably, Wilkins, Merrilees and Herington (2007) suggested a dimension called stylish comfort, which includes stylish atmosphere, relaxing ambiance, artifacts and size of facilities that suggest grandeur, as integral to the total quality of hotels. Instead of making references to style and/or grandeur, P2P accommodation reviews in this cluster use similar components (i.e., “backyard”, “garden”, “patio”) to signal a relaxing, peaceful atmosphere of a home. These add to the overall image of P2P properties and were suggested as important hotel attributes (Dolnicar and Otter 2003).

5.3 Review Clusters and Ratings

In order to better understand how these different attributes contribute to guest satisfaction with P2P accommodation, the association between review clusters and rating scores was analyzed. The property ratings are presented in Table 9, which include ratings for overall experience, accuracy (i.e., how online description accurately represents real condition of listings), check-in process, cleanliness (of property), communication (i.e., interaction with host prior to arrival and during stay), location, and value (for money). As suggested in previous research (e.g., Zervas, Proserpio, and Byers 2015b), rating scores on Airbnb platform are positively skewed compared to other online review platforms. Therefore, it is important to note that the contribution of review clusters revealed in this study is based on small variations in the rating scores (i.e., note the relatively low standard deviations). However, since Airbnb employed a two-way review system (i.e., hosts and guests can review and rate each other) at the time of data collection, giving a negative review can be a disincentive for guests due to the inherent risk of possible retaliation that potentially damages his/her own reputation. Consequently, even a small deviation from 5-point rating score can be an indication of lack of satisfaction. Therefore, the outcomes provide useful insights on aspects of P2P accommodation stay that contribute positively or negatively to guest satisfaction.

Table 9. Property Ratings (*N* = 1613)

Ratings	Minimum	Maximum	Mean	Std. Deviation
Overall	2.00	5.00	4.769	.279
Accuracy	2.00	5.00	4.830	.339
Cleanliness	2.00	5.00	4.793	.379
Check-in	2.00	5.00	4.927	.250
Communication	2.00	5.00	4.922	.224
Location	3.00	5.00	4.757	.330
Value	2.00	5.00	4.737	.349

Regression analyses were performed using property with at least one review belonging to the five clusters as the unit of analysis (*N* = 1,613). The analyses were conducted separately for overall, accuracy, cleanliness, check-in, communication, location, and value ratings. As presented in Table 10, the results show that higher proportion of Feel Welcome reviews (i.e., reviews belonging to Feel Welcome cluster) contribute positively to higher scores in all ratings, except for Location rating. On the other hand, higher proportion of Service reviews contributes negatively to all ratings. In order to better understand the negative impacts of the prevalence of Service-related reviews on rating scores, reviews were carefully examined to identify negative terms associated with services that might indicate lower performance among P2P accommodation hosts. Instead, a pattern emerged that guests who are not fully satisfied with their stay due to the listing conditions (e.g., cleanliness, noise, lack of amenities) or problems during stay (e.g., disturbances, miscommunication, etc.) tend to emphasize how hosts were quick to offer solutions to these problems (i.e., respond quickly to inquiries and demand). Therefore, it can be suggested that Service reviews are associated with lower scores not necessarily because hosts perform poorly in servicing the guests (i.e., occurrences of negative reviews are negligible), but many guests highlight the positive aspects of host services (especially responsiveness) when writing reviews to offset their dissatisfaction toward cleanliness, values, and, thus, overall stay. Since both clusters, Service and Feel Welcome, are associated with host – guest interactions, it can be suggested that staying with hosts who go beyond making service processes easier leading up to the stay (e.g., being responsive to inquiries, keeping guests well-informed), but make guests feel at home during the stay is key to a satisfying stay in P2P accommodation. This signifies the importance of positive social relationships between hosts and guests during the stay.

Table 10. Regression Models: Review Clusters on Rating Scores

	Overall Rating	**Accuracy**	**Cleanli-ness**	**Check-in**	**Commu-nication**	**Location**	**Value**
Model							
R^2	.102	.038	.042	.042	.047	.098	.075
F (sig.)	7.838 (.00)	5.438 (.00)	6.077 (.00)	6.483 (.00)	6.779 (.00)	14.879 (.00)	11.218 (.00)
Independent Variables – Beta (sig.)							
Service	-.309 (.00)	-.133 (.02)	-.224 (.00)	-.191 (.00)	-.144 (.01)	-.153 (.01)	-.227 (.00)
Facility	n.s.	n.s.	n.s.	n.s.	n.s.	-.202 (.00)	n.s.
Location	.215 (.00)	.188 (.00)	.207 (.00)	.133 (.02)	.125 (.02)	.408 (.00)	.123 (.02)
Welcome	.298 (.00)	.113 (.02)	.108 (.03)	.153 (.00)	.184 (.00)	n.s.	.226 (.00)
Comfort	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

The results also demonstrate the positive contribution of Location reviews on all ratings. Notably, Location reviews have a significantly large positive effect on Location rating, while Facility reviews contribute negatively to Location rating. It can be suggested that guests who are satisfied with the location of the P2P accommodation property (i.e., giving a higher rating for Location) specifically emphasize the locational advantages in their reviews. On the other hand, guests staying at properties in less advantageous locations highlight other aspects (i.e., facilities) reflecting the benefits offered by the P2P accommodation. The analyses revealed no significant effect of Comfort of a Home reviews on ratings.

5. Conclusion and Recommendation

Critical for accommodation providers is to understand guest needs and attributes that contribute to guest satisfaction. Knowing what matters enables management to allocate resources effectively to achieve the best return on investment, often in the form of repeat visitation or positive (electronic) word of mouth. This study analyzed P2P accommodation reviews to contribute to the limited literature in this rapidly growing segment of the hospitality industry. Based on a cluster analysis on Airbnb reviews in Portland, OR, five accommodation attributes were identified: Service, Facility, Location, Feel Welcome, and Comfort of a Home. Some of these attributes are comparable with those identified in previous studies on hotels and B&Bs. For example, P2P accommodation and traditional hotel guests expect service quality (e.g., Knutson 1988; Dubé and Renaghan 1999). However, with P2P accommodation, a greater emphasis was given on hosts making guests feel welcome in their homes. This indicates the focus on direct guest – host relations in social settings (i.e., beyond customer – staff relations in professional/commercial settings), confirming the social appeal of the sharing economy systems (Guttentag 2015; Tussyadiah and Pesonen 2015). Based on the association of these attributes and guest evaluation, feeling welcome was consistently linked to higher rating scores, signifying its importance on guest satisfaction and positive eWOM. Similarly, convenience of location was identified as important for hotel and B&B guests (e.g., Ananth et al. 1991; Dolnicar and Otter 2003; Knutson 1988). However, P2P accommodation reviews emphasize the experiential characteristics of the neighborhoods, confirming that P2P accommodation guests appreciate staying in authentic (non-tourist) settings (Guttentag 2015; Tussyadiah 2015; Tussyadiah and Pesonen 2015).

This study contributes to the literature in tourism and hospitality on key service characteristics that differentiate P2P accommodation, a new entrant to the competitive landscape of hospitality industry, from other accommodation types. The peer-to-peer contexts present conceptual challenges due to the shift in service delivery from professional service providers to individuals. To date, there is limited research to recognize if this shift may cause an adjustment in customer expectation and evaluation of accommodation services. Additionally, from marketing and management points of view, the new business model poses some issues in conceptualizing business strategies for (non-business) individuals (i.e., hosts). Therefore, the findings in this study enrich the literature in two ways: (1) clarifying the difference between P2P accommodation and other types of accommodation by identifying the key service attributes and their relative importance for guests, and (2) explicating the roles of (individual) providers in peer-to-peer service contexts so better marketing and management strategies can be formulated.

These findings provide several implications for academics and practitioners. First, this study confirms the benefits of extracting market and competitive intelligence from large, unstructured human-authored text data to support marketing and management decisions in tourism and hospitality. The explosive growth of P2P accommodation users contributes to the massive online content through its built-in reputation systems (i.e., allowing for a large amount of information from host and guest evaluation), emphasizing the increasing importance of big data analytics in hospitality management (e.g., as previously suggested by Xiang et al., 2015). For researchers, this study demonstrates the importance to apply different methodologies to summarize and interpret salient information emerging from large textual data to better understand and explain consumer behavior in service contexts that occur both online (i.e., reservation) and offline (i.e., during stay). For practitioners, the utilization of text analytics and visualization of publicly available dataset (i.e., online consumer reviews) provides effective yet reasonable alternatives for or complimentary to other market research tools such as customer surveys or guest comment cards to aid in better understanding the determinants of guest satisfaction and the strengths of their services relative to competitors.

Secondly, the results from this study inform both P2P accommodation hosts and service providers (commercial sharing platforms) with a greater understanding of what guests appreciate about their stay. More specifically, areas of continued investment to drive guest satisfaction and positive eWOM were identified: location (neighborhood characteristics) and feeling welcome. While some hosts would reap the benefits from the locational advantages of their property, those whose properties are in less advantageous locations (e.g., far from tourist attractions, less neighborhood amenities, limited access of public transportation) should provide accurate descriptions and communicate clearly with prospective guests to shape their expectations (e.g., targeting those who prefer quieter neighborhoods or bring their own cars). Alternatively, hosts can offer additional services, such as airport pick-ups and drop-offs, to add value for convenience. Most importantly, regardless of locations and facilities, making guests feel welcome is imperative to ensure satisfaction. For P2P accommodation platforms such as Airbnb, it is important to provide hosts with easy to use online tools to estimate competitive pricing and recommendation of value-added services based on locations (i.e., to compensate for locational values) as well as education and training for host competence.

Despite the aforementioned contribution, this study has several limitations stemming from the data. First, while applying text analytics has proven to be beneficial for attribute and pattern extraction, there are concerns that valuable, often nuanced, information is lost in the data. Hence, a qualitative analysis of select samples of reviews will be beneficial to contribute to the accuracy of big data analysis. Furthermore, the size of the dataset may have an influence on the results of regression analyses, making insignificant relationships significant. However, the significant relationships identified in this study are relevant to the aforementioned practical implications. Second, the bias towards positive reviews on Airbnb (consistent with Zervas, Proserpio, and Byers 2015b) allowed for identification of P2P accommodation attributes that are valued by guests. However, the lack of negative reviews did not allow identifying common areas that require immediate improvements for P2P accommodation hosts. Third, since Airbnb does not link individual reviews with their ratings on its website, it was not possible to make direct association between reviews and ratings (aggregate ratings were used in this study). Future research should identify different approaches to pinpoint keywords that add to or subtract from guest evaluation.

Appendix

The main purpose of comparing data across multiple locations is to diagnose if new attributes emerge from different datasets and to test the consistency of the identified themes in other contexts. Two datasets consisting of randomly selected 10,000 reviews (using a random number generator) from property listings in New York City, New York, US and London, United Kingdom (UK), available on *insideairbnb.com* were analyzed. As they both are large metropolitan areas and top international destinations, they serve as a contrast to Portland, Oregon, a smaller tourism destination. Following the same procedure, reviews written in languages other than English were eliminated first and preprocessing was applied to produce sentences, tokens, word types, and term frequency. Table A1 presents descriptive statistics of the datasets.

Table A1. Descriptive Statistics of Main Dataset and Comparison Samples

	Portland, OR (US) (Main Dataset)	**New York City, NY (US)** (Sample)	**London (UK)** (Sample)
Reviews	41,560	9,999	8,831
Sentences	215,497	47,710	36,674
Tokens	1,473,197	302,776	252,371
Word Types	21,561	12,150	11,245
Term Frequency: Mean (St. D)	45.57 (484.58)	22.28 (156.09)	20.20 (133.51)

New York City. The top 10 nouns, adjectives, verbs and compound words for New York City revolve around similar themes as in the main dataset (Table A2). These are general assessment of listings (in particular, New York City reviews frequently mentioned “apartment” instead of “home”), hosts (e.g., “great host,” “wonderful host”), location (e.g., “great location,” “central park,” “times square”), and overall experience (e.g., “great stay,” “great time”). New terms include “subway” and “subway station” (substituting “public transportation” and “bus”), which point to the importance of convenience and mobility to explore the big city.

Table A2. New York City: Top 10 Words and Compound Words

No	Nouns	Freq.	Adjectives	Freq.	Verbs	Freq.	Compound Words
1	apartment	7866	great	6662	have	8523	great host
2	place	5977	clean	3944	stay	4942	great location
3	host	3796	nice	3414	recommend	2816	great time
4	location	3481	comfortable	2476	do	2728	great place
5	stay	3249	good	2275	get	2444	great stay
6	room	3117	easy	1838	make	2385	great experience
7	time	2967	perfect	1833	need	2022	central park
8	subway	2769	helpful	1704	feel	1939	subway station
9	everything	2325	friendly	1445	go	1611	great apartment
10	neighborhood	2225	wonderful	1103	walk	1345	first time

Cluster analysis for New York City reviews produced four clusters: City Life (1,356 reviews), Service (1,181 reviews), Location and Facility (3,195 reviews), and Feel Welcome (4,194 reviews). The latter three clusters match the four clusters in Portland, with Location and Facility merging into one. Due to the nature of New York City and its listings (i.e., more apartments rather than individual homes), Portland's Comfort of a Home cluster is replaced with experiencing the ambiance of big city living. Nevertheless, both clusters are the key aspects that represent the overall stay at the destination. As presented in Table A3, the theme within City Life focuses on ambience (e.g., "luxurious," "fancy," "cool"), property features (e.g., "loft," "rooftop," "balcony"), and easy access and neighborhood. The merging of Location and Facility indicates that the two aspects are frequently mentioned together, suggesting that they are considered equally valuable for P2P accommodation guests in New York City. Importantly, despite locational differences between New York City and Portland, the themes around Service and Feeling Welcome are consistent. Reviews in Service emphasize host responsiveness and flexibility, as well as smooth, seamless, uncomplicated transaction and booking process. Reviews in Feel Welcome highlight how hosts provide everything that guests need and make them feel at home.

Table A3. New York City: Top 10 Keywords in Review Clusters

No	City Life			Service			Location & Facility			Feel Welcome		
	Words	Freq.	JC[1]	Words	Freq.	JC[1]	Words	Freq.	JC[1]	Words	Freq.	JC[1]
1	place	840	.18	location	566	.15	have	1876	.30	have	2793	.43
2	recommend	435	.12	host	594	.14	great	1644	.26	place	2210	.36
3	neighborhood	368	.10	apartment	630	.12	apartment	1583	.25	stay	2101	.36
4	nice	315	.10	helpful	254	.10	stay	1078	.22	apartment	2260	.33
5	perfect	248	.09	friendly	171	.07	clean	1238	.22	great	2087	.31
6	easy	238	.09	accommodate	117	.06	host	1200	.22	everything	1427	.30
7	love	113	.05	describe	64	.04	location	1091	.20	need	1330	.29
8	super	96	.05	responsive	50	.03	comfortable	903	.20	make	1367	.29
9	access	70	.04	flexible	48	.03	subway	901	.18	get	1336	.28
10	anyone	66	.04	listing	24	.02	time	861	.18	feel	1277	.28

Note: [1]*Jaccard Coefficient*

London. Similarly, based on the top 10 words and compound words for P2P accommodation in London (see Table A4), reviews are dominated by the assessment of overall experience (e.g., "great time," "great experience"), host (e.g., "helpful," "great host"), and location (e.g., "central london," "center of london," "perfect location"). As for New York City, many reviews focus on staying in an apartment (i.e., "flat," "apartment") rather than a home. Key terms for convenience and access are also present but again adjust to the destination (e.g., "tube station", "bus").

Table A4. London: Top 10 Words and Compound Words

No	Nouns	Freq.	Adjectives	Freq.	Verbs	Freq.	Compound Words
1	flat	4033	great	4588	have	6804	great location
2	place	3816	clean	3174	stay	4048	great host
3	apartment	3767	nice	3167	recommend	2744	central london
4	host	3385	comfortable	2208	do	1891	tube station
5	location	3124	good	2192	make	1861	great time
6	room	3086	helpful	1742	get	1779	great place
7	stay	2984	perfect	1573	need	1633	great stay
8	time	2155	lovely	1566	feel	1553	good location
9	station	1765	easy	1396	go	1265	center of london
10	house	1631	friendly	1297	walk	1215	next time

Four clusters were identified: Facility (935 reviews), Service (2,447 reviews), Feel Welcome (637 reviews), and Location (4,732 reviews). They match the four clusters in Portland (see Table A5), except for Comfort of a Home. It is noteworthy that the majority of London reviews belong to Location (54%) and Service (28%), while Facility and Feel Welcome account for about 18% of the reviews. That is because descriptors in Location also include terms that represent broad categories of accommodation attributes. However, reviews in Service are consistent with those in Portland and New York City in representing hosts' service quality (e.g., "attentive," "helpful," "advice"). The presence of Service and Feel Welcome clusters confirms the importance of host-guest interactions prior to and during the stay.

Table A5. London: Top 10 Keywords in Review Clusters

	Location			Service			Feel Welcome			Facility		
No	**Words**	**Freq.**	**JC[1]**	**Words**	**Freq.**	**JC[1]**	**Words**	**Freq.**	**JC[1]**	**Words**	**Freq.**	**JC[1]**
1	flat	698	.26	host	926	.20	feel	518	.35	have	2873	.48
2	have	454	.12	location	878	.20	home	448	.32	great	2051	.34
3	stay	445	.10	apartment	730	.18	make	435	.24	stay (N)	1928	.33
4	great	405	.10	nice	729	.17	welcome (A)	278	.14	stay (V)	1866	.32
5	host	366	.10	good	508	.14	comfortable	274	.09	clean	1830	.32
6	location	359	.10	helpful	475	.13	lovely	264	.08	place	1765	.30
7	clean	334	.09	friendly	381	.11	welcome (V)	215	.08	room	1628	.29
8	recommend	286	.09	describe	109	.04	host	206	.08	recommend	1516	.26
9	time	223	.09	kind	88	.03	stay	206	.08	comfortable	1455	.25
10	helpful	202	.08	accommodation	73	.03	wonderful	205	.07	nice	1385	.25

Note: [1]Jaccard Coefficient, (A) = adjective, (V) = verb, (N) = noun

The results from locational comparison with samples from New York City and London confirm that: (1) location of P2P properties is consistently included on the list of top keywords in guest reviews, regardless of locational contexts; (2) host – guest interactions are considered important, both for service delivery (e.g., facilitating reservation, responsive to requests, flexibility with check-in time, etc.) and for creating the positive feeling of being welcome; (3) locational differences shape the themes around P2P property characteristics, including facility and ambience. In particular, as shown from the main dataset (Portland) and the sample from New York City, even though details of mentioned property features are different (i.e., charming home living in Portland versus loft living in New York City), guests value the experiences from staying with and embracing the lifestyle of local residents. It is important to note that while Portland is not considered a popular destination for international tourists, New York City and London are. Therefore, it is expected that more diverse international guests, most likely with different expectations on accommodation experiences, write reviews for P2P accommodation in New York City and London. Consequently, identifying consistent attributes among these different destinations signifies their salience.

References

Airbnb, 2015a. *Airbnb summer travel report: 2015*. http://blog.airbnb.com/wp-content/uploads/2015/09/Airbnb-Summer-Travel-Report-1.pdf

Airbnb, 2015b. *Airbnb Economic Impact*. http://blog.airbnb.com/economic-impact-airbnb/

Albayrak, T., Caber, M., 2014. Prioritisation of the hotel attributes according to their influence on satisfaction: A comparison of two techniques. *Tourism Management* 46, 43–50.

Alpert, M. I., 1971. Identification of determinant attributes: A comparison of models. *Journal of Marketing Research* 8, 184–191.

Ananth, M., DeMicco, F.J., Moreo, P.J., Howey, R.M., 1992. Marketplace lodging needs of mature travellers. *The Cornell Hotel and Restaurant Quarterly* 33 (4), 12–24.

Atkinson, A., 1988. Answering the eternal question: What does the customer want? *The Cornell Hotel and Restaurant Administration Quarterly* 29 (2), 12–14.

Balmer, S., Baum, T., 1993. Applying Herzberg's hygiene factors to the changing accommodation environment. *International Journal of Contemporary Hospitality Management* 5, 32–35.

Belk, R., 2014. You are what you can access: Sharing and collaborative consumption online. *Journal of Business Research* 67 (8), 1595–1600.

Bitner, M. J., 1990. Evaluating service encounters: The effects of physical surroundings and employee responses. *Journal of Marketing* 54, 69–82.

Botsman, R., Rogers, R., 2011. *What's Mine is Yours: The Rise of Collaborative Consumption.* New York: Harper Business.

Cadotte, E. R., Turgeon, N., 1988. Key factors in guest satisfaction. *The Cornell Hotel and Restaurant Administration Quarterly* 28 (4), 45–51.

CB Insights, 2015. *Why That Crazy-High AirBnB Valuation Is Fair*. https://www.cbinsights.com/blog/airbnb-hospitality-industry-valuation-breakdown/

Chan, J. K. L., Baum, T., 2006. Determination of satisfiers and dissatisfiers using Herzberg's motivator and hygiene factor theory: An exploratory study. *Tourism, Culture and Communication* 7 (2), 117–131.

Chen, H., Chiang, R. H. L., Storey, V. C., 2012. Business intelligence and analytics: From big data to big impact. *MIS Quarterly* 56 (4), 1165–1188.

Choi, T. Y., Chu, R., 2001. Determinants of hotel guests' satisfaction and repeat patronage in the Hong Kong hotel industry. *International Journal of Hospitality Management*, 20, 277–297.

City of Portland, The, 2015. Accessory short term rental permits. *Development Services: Permits*. http://www.portlandoregon.gov/bds/65603.

Clow, K. E., Garretson, J. A., Kurtz, D. L., 1994. An exploratory study into the purchase 19 decision process used by leisure travellers in hotel selection. *Journal of Hospitality and Leisure Marketing*, 4, 53–71.

Creative Commons, n.d.. *CC0 1.0 Universal (CC0 1.0). Public Domain Dedication*. https://creativecommons.org/publicdomain/zero/1.0/

Dolnicar, S., Otter, T., 2003. Which hotel attributes matter? A review of previous and a framework for future research. In Griffn, T. and R. Harris (Eds.), Proceedings of the 9th Annual Conference of the Asia Pacific Tourism Association (APTA), University of Technology Sydney, 2003, 1, 176–188.

Dörre, J., Gerstl, P., Seiffert, R., 1999. Text mining: Finding nuggets in mountains of textual data. In Proceedings of the 5th ACM International Conference on Knowledge Discovery and Data Mining (KDD-99), 398–401, San Diego, US. ACM Press, New York, US.

Dubé, L., Renaghan, L.M., 1999. Building customer loyalty. *The Cornell Hotel and Restaurant Administration Quarterly*, 78-88.

Felix, D., Broad, S., Griffiths, M., 2008. The bed-and-breakfast experience: An analysis of hosts' and guests' expectations. In Jauhari, V. (Ed.), *Global Cases on Hospitality Ind*ustry (pp. 55-77). New York – London: Routledge.

Freitag, J.D., Haywood, J., 2015. Digging to the core of Airbnb Big Apple data. http://www.hotelnewsnow.com/Article/16395/Digging-to-the-core-of-Airbnbs-Big-Apple-data

Fruchterman, T.M.J., Reingold, E.M., 1991. Graph drawing by force-directed placement. *Software - Practice and Experience* 21 (11), 1129–1164.

Fuchs, M., Höpken, W., Lexhagen, M., 2014. Big data analytics for knowledge generation in tourism destinations – A case from Sweden. *Journal of Destination Marketing & Management* 3 (4), 198-209.

Gonen, Y., Sutherland, A., 2015. Most Airbnb rentals violate the state's short-term leasing law. *New York Post*. http://nypost.com/2015/12/02/most-airbnb-rentals-violate-the-states-short-term-leasing-law/

Guttentag, D., 2015. Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism 18*(2), 1192-1217.

Herzberg, F. B., Mausner, B., Snyderman, B., 1962. *The Motivation to Work.* 2nd Edition. New York: John Wiley and Sons, Inc.

Higuchi, K., 2015. *KH Coder 2.x Reference Manual*. http://khc.sourceforge.net/en/manual_en_v2.pdf

Inside Airbnb, 2015. *Get the Data.* http://insideairbnb.com/get-the-data.html.

Karlsson, L., Dolnicar, S., 2016. Someone's been sleeping in my bed. *Annals of Tourism Research* 58, 159-162.

Kim, S., 2016. Berlin bans thousands of Airbnb properties. *The Telegraph*. http://www.telegraph.co.uk/travel/destinations/europe/germany/berlin/articles/airbnb-listings-plummet-in-berlin-as-ban-comes-into-force/

Knutson, B., 1988. Frequent travelers: making them happy and bringing them back. *The Cornell Hotel and Restaurant Administration Quarterly* 29 (1), 83–87.

Korkontzelos, I., Klapaftis, I. P., Manandhar, S. 2008. Reviewing and evaluating automatic term recognition techniques. In Nordstrom, B., and A. Ranta (Eds.), *Advances in Natural Language Processing* (pp. 248-259). Springer, Berlin – Heidelberg.

Li, G., Law, R., Vu, H. Q., Rong, J., 2013. Discovering the hotel selection preferences of Hong Kong inbound travelers using the Choquet Integral. *Tourism Management* 36, 321-330.

Lipsman, A., 2007. Online consumer-generated reviews have significant impact on offline purchase behavior. *comScore*. Retrieved from http://www.comscore.com/Insights/Press-Releases/2007/11/Online-Consumer-Reviews-Impact-Offline-Purchasing-Behavior

Lockyer, T., 2004. Weekend Accommodation - The Challenge: What are the Guests Looking for?,*Journal of Hospitality and Tourism Management*, 11, 1, 1-12.

Lockyer, T., 2005. The perceived importance of price as one hotel selection dimension. *Tourism Management* 26 (4), 529-537.

Mattila, A., 1999. Consumers' value judgments: How business travelers evaluate luxury-hotel services. *The Cornell Hotel and Restaurant Administration Quarterly*, 40-46.

Möhlmann. M., 2015. Collaborative consumption: Determinants of satisfaction and the likelihood of using a sharing economy option again. *Journal of Consumer Behaviour*. DOI: 10.1002/cb.1512.

Nakagawa, H., 2000. Automatic term recognition based on statistics of compound nouns. *Terminology* 6 (2), 195-210.

Nakagawa, H., Mori, T., 2002. A simple but powerful automatic term extraction method. *In* Proceedings of the Second International Workshop on Computational Terminology (COMPUTERM 2002). Stroudsburg, PA: Association of Computational Linguistics.

Owyang, J., 2013. The collaborative economy: products, services and market relationships have changed as sharing startups impact business models. A market definition report. *Altimeter*. https://www.slideshare.net/Altimeter/the-collaborative-economy

Pons, P., Latapy, M., 2005. Computing communities in large networks using random walks. http://arxiv.org/pdf/physics/0512106v1.pdf.

Ramanathan, U., Ramanathan, R., 2011. Guests' perceptions on factors influencing customer loyalty: an analysis for UK hotels. *International Journal of Contemporary Hospitality Management* 23 (1), 7–25.

Rivers, M.J., Toh, R.S., Alaoui, M., 1991. Frequent-stayer programs: the demographic, behavioral, and attitudinal characteristics of hotel steady sleepers. *Journal of Travel Research* 30 (2), 41–45.

Saleh, F., Ryan, C., 1992. Client perceptions of hotels: A multi-attribute approach. *Tourism Management* 13 (2), 163-168.

Scarinci, J., Richins, H., 2008. Specialist lodging in the USA: Motivations of bed and breakfast accommodation guests. *Tourism: An International Interdisciplinary Journal*, 56(3), 271-282.

Shergill, G. S., Sun, W., 2004. Tourists' perceptions towards hotel services in New Zealand. *International Journal of Hospitality and Tourism Administration* 5 (4), 1-29.

Stringam, B. B., Gerdes, J., Vanleeuwen, D. M., 2010. Assessing the importance and relationships of ratings on user-generated traveler reviews. *Journal of Quality Assurance in Hospitality and Tourism* 11 (2), 73-92.

Toutanova, K., Klein, D., Manning, C., Singer, Y., 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. *In* Proceedings of HLT-NAACL 2003 (pp. 252-259).

Tsao, W.-C., Hsieh, M.-T., Shih, L.-W., Lin, T. M. Y., 2015. Compliance with eWOM: The influence of hotel reviews on booking intention from the perspective of consumer conformity. *International Journal of Hospitality Management* 46, 99-111.

Tussyadiah, I. P., 2015. An exploratory study on drivers and deterrents of collaborative consumption in travel. In Tussyadiah, I., Inversini, A. (Eds.), *Information & Communication Technologies in Tourism 2015*. Switzerland: Springer, pp. 819-32.

Tussyadiah, I. P., 2016. Factors of satisfaction and intention to use peer-to-peer accommodation. *International Journal of Hospitality Management 55,* 70-80.

Tussyadiah, I. P., Pesonen, J., 2015. Impacts of Peer-to-Peer Accommodation Use on Travel Patterns. *Journal of Travel Research.* OnlineFirst. DOI: 10.1177/0047287515608505.

Ward, J. H., Jr., 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236-244.

Westbrook, R. A., 1987. Product/consumption-based affective responses and post-purchase processes. *Journal of Marketing Research* 24 (3), 258-270.

Wilkins, H., Merrilees B., Herington, C., 2007. Towards an understanding of total service quality in hotels. *International Journal of Hospitality Management* 26, 840-853.

Xiang, Z., Schwartz, Z., Gerdes, J.H., Uysal, M., 2015. What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management* 44, 120-130.

Yen, C.-L., Tang, C.-H., 2015. Hotel attribute performance, eWOM motivations, and media choice. *International Journal of Hospitality Management* 46, 79-88.

Zane, B., 1997. The B&B guests: A comprehensive view. *Cornell Hotel and Restaurant Administration Quarterly*, 38(4), 69-75.

Zervas, G., Proserpio, D., Byers, J., 2015a. The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. Boston U. School of Management Research Paper No. 2013-16. Available at SSRN: http://ssrn.com/abstract=2366898

Zervas, G., Proserpio, D., Byers, J., 2015b. A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average. Available at SSRN: http://ssrn.com/abstract=2554500

Zhou, L., Ye, S., Pearce, P. L., Wu, M.-Y. 2014. Refreshing hotel satisfaction studies by reconfiguring customer review data. *International Journal of Hospitality Management*, 38, 1 – 20.