Hotels vs. Peer-to-Peer Accommodation Rentals: Text Analytics of Consumer Reviews in Portland, Oregon*

Iis P. Tussyadiah
School of Hospitality Business Management
Carson College of Business
Washington State University, USA

and

Florian Zach
School of Hospitality Business Management
Carson College of Business
Washington State University, USA

ABSTRACT

Peer-to-peer (P2P) accommodation rentals continue to grow at a phenomenal rate. Examining how this business model affects the competitive landscape of accommodation services is of strategic importance to hotels and tourism destinations. This study explores the competitive edge of P2P accommodation in comparison to hotels by extracting key content and themes from online reviews to explain the key service attributes sought by guests. The results from text analytics using terminology extraction and word co-occurrence networks indicate that even though guests expect similar core services such as clean rooms and comfortable beds, different attributes support the competitive advantage of hotels and P2P rentals. While conveniences offered by hotels are unparalleled by P2P accommodation, the latter appeal to consumers driven by experiential and social motivations. Managerial implications for hotels and P2P accommodation are provided.

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

INTRODUCTION

The sharing economy has penetrated the tourism and hospitality marketplace. Facilitated by online social networking platforms, consumers coordinate the acquisition and distribution of access to accommodation among their peers through services such as Airbnb and 9flats, a phenomenon known as collaborative consumption (Belk, 2014). Revenues generated from peer-to-peer (P2P) accommodation have surpassed US$3.5 billion in 2013 with growth exceeding 25%, making it a disruptive economic force (Geron 2013). The rapid rise of peer-to-peer accommodation presents opportunities (e.g., generates local income, provides alternative employment) and challenges (e.g., regulatory issues) for tourism destinations (Geron 2012; 2013). Critically, P2P accommodation rentals affect the competitive landscape of accommodation services as "regular people" host tourists and, by so doing, take consumers away from hotels. For example, Zervas, Proservio, and Byers (2014) estimate that 1% increase in Airbnb listing causes .05% decrease in hotel revenues in the State of Texas. Therefore, it is

*In Proceedings of the Travel and Tourism Research Association (TTRA) 46th Annual International Conference, Portland, OR, June 15 – 17, 2015. Link: http://scholarworks.umass.edu/ttra/ttra2015/Academic_Papers_Oral/2/
Contact: iis.tussyadiah@wsu.edu

important to explore the competitive advantage of P2P accommodation in comparison to hotels for both parties to better strategize their services.

Gaining actionable insights from consumer intelligence available online is fundamental in today's hospitality business analytics. Studies extracting key content and themes from consumer reviews to explain the important attributes of accommodation services have emerged (e.g., Xiang, Schwartz, Gerdes, and Uysal 2015; Zhou, Ye, Pearce, and Wu 2014). These studies identified different dimensions in hotel reviews that carry varying weights to guest satisfaction, informing hotel management with essential factors of service to direct their attention to. However, little is known if consumers would evaluate P2P accommodation in the same manners, or for the same aspects, as they do hotels. Guttentag (2013) suggests that consumers use P2P accommodation because of its economic and experiential values. Tussyadiah (2015) identified three major factors that motivate the use of P2P accommodation: sustainability (i.e., social and environmental responsibility), community (i.e., social interactions), and economic benefits (i.e., lower cost). Aspects typically tied to hotel selection factors, such as location and amenities, were not identified in her study, indicating that there may be differences in terms of the service dimensions that people seek from alternative accommodation. To this end, online consumer reviews may offer important intelligence to clarify these issues. Hence, the goals of this study are twofold: (1) to explore and compare the key service characteristics of hotels and P2P accommodation emerging from consumer reviews and (2) to recommend how to turn these insights into management actions to achieve competitive advantage.

CONCEPTUAL FRAMEWORK

Accommodation Attributes

In tourism and hospitality marketing and management, identifying various accommodation attributes that influence hotel selection and guest satisfaction is considered important due to its practical relevance in attracting new guests and retaining current patrons. Indeed, various studies suggest that there are different hotel features that guests evaluate and use as decision criteria in the hotel selection process (e.g., Clow, Garretson, and Kurtz 1994; Dolnicar 2002). Studies also demonstrate that different hotel attributes influence satisfaction and post-purchase behavior associated with hotel stay, such as loyalty and electronic word of mouth (eWOM) behavior, to a varying degree (e.g., Xiang et al. 2015; Yen and Tang 2015). It is suggested that 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).

In the context of patronage decision, previous studies include tangible and intangible dimensions in hotel selection criteria. It is argued that 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 (e.g., Bitner 1990; Clow, Garretson, and Kurtz 1994). Clow et al. (1994) show that, for example, in order to evaluate the quality of service, consumers refer to own personal experiences, staff behavior, price structure, word-of-mouth, and the appearance of the hotel facility. Herzberg's Two-Factor Theory (Herzberg, Mausner, and Snyderman 1962) has been used to explain the different hotel attributes that contribute to satisfaction (e.g., Balmer and Baum 1993). The theory suggests the following conditions: (1) hygiene (maintenance) factors, whose absence would lead to conditions of dissatisfaction, and (2) motivators (true satisfiers) factors, whose presence would lead to conditions of satisfaction. According to Chan and Baum (2006), satisfiers are often derived from experiential dimensions (intangible attributes) that result in affective responses, while hygiene factors are typically derived from utilitarian values (tangible attributes) that result in cognitive responses. A study by Dolnicar and Otter (2003) provides a comprehensive look on literature discussing hotel attributes that guests consider important. Based on a meta-analysis of 21 studies, they identified 173 hotel attributes that are grouped further into the following categories: "Image," "Value/Price," "Hotel," "Room," "Services," "Marketing," "Food and Beverage," "Security," and "Location." They also identified top attributes with convenience of location being the most important criterion, followed by service quality, reputation, friendliness of staff, price, room cleanliness, value for money, etc. (Dolnicar and Otter, 2003).

Although the dimensions included in these studies are varied, attributes for hotel selection and evaluation are well-researched. However, the knowledge on the dimensions used to evaluate P2P accommodation is extremely limited. While the basic services of P2P accommodation are comparable to hotels (i.e., room and board), P2P accommodation is characterized by a lack of standards. Guests can choose three types of accommodation listings 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, etc.). Therefore, it is important to explore which features really matter for guests when evaluating their stay at P2P accommodation. Cost savings, value for money, and a drive for community are confirmed as motivators for the use of P2P accommodation (Guttentag 2013; Möhlmann 2015; Owyang 2013; Tussyadiah 2015). 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 stays. Considering the rapid growth of this collaborative consumption model, it is important to identify the attributes of P2P accommodation stays that guests consider important and verify their similarities and differences with hotel attributes.

Analytics of Online Reviews

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, Garretson, and Kurtz 1994). More recently, the development in consumer devices and social network technologies allows consumers to leave traces of their consumption patterns online through pictures, check-ins, statuses, reviews, etc. 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, mount 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 intelligence and competitive analysis to assist business managers with making timely decisions (Chen, Chiang, and Storey 2012). Therefore, UGC provides opportunities for tourism and hospitality managers to gain actionable insights to the factors of guest experiences and satisfaction.

The interest in extracting consumer opinion from UGC continues to grow among academics and business practitioners. Hotel reviews are identified as a vehicle for eWOM, which is valuable in predicting booking intention and guest satisfaction (e.g., Tsao, Hsieh, Shih, and Lin 2015; Xiang et al. 2015; Zhou et al. 2014). Importantly, an analysis of UGC data can reveal the influence of different dimensions of hotel services, by extracting attributes that are frequently discussed by consumers, on purchase decisions and evaluation. Zhou et al. (2014) identified 17 attributes that are classified into satisfiers, dissatisfiers, bidirectional forces, and neutrals, based on their impacts on guest satisfaction. Most recently, Xiang et al. (2015) identified six dimensions in hotel reviews (i.e., "Hybrid," "Deals," "Amenities," "Family friendliness," "Core product," and "Staff") with varying degrees of influence on satisfaction, which is measured through star ratings. These studies indicate the usefulness of analyzing UGC in creating knowledge and recognizing patterns to better understand the factors that matter most for guest experiences.

The challenge in gaining actionable insights from UGC is to extract valuable nuggets of information and patterns from relatively large, highly unstructured (often messy) text data, written in natural language (human-authored). Manually scanning and analyzing such data is considered impractical for decision making due to high computational burden. Therefore, efforts have been made to create and apply effective automatic knowledge extraction through text mining techniques, integrating approaches from machine learning and natural language processing (NLP). Rooted in information retrieval, text mining (or text analytics) is a set of techniques used to discover new knowledge by automatically extract information from free-text documents, which include extraction of features from single documents and the analysis of the feature distribution over the collection of documents to detect interesting patterns and trends (Dörre, Gerstl, and Seiffert 1999). The advancement in NLP technologies allows for these processes in text mining: information extraction (i.e., identifying key phrases within text), topic tracking, summarization (i.e., reducing document length while retaining its main points), categorization (i.e., identifying main themes or "bag of words"), clustering, concept linkage (i.e., connecting documents with shared concepts), information visualization, and question answering (Fan, Wallace, Rich, and Zhang 2006). For tourism and hospitality businesses, text mining techniques can be valuable in handling voluminous online review data to extract important features (e.g., accommodation attributes) and detect patterns and trends to better understand their consumers and competition.

METHODOLOGY

This study analyzes and compares the competitive advantages of hotels and P2P accommodation by extracting important attributes from consumer reviews using text mining techniques. Portland, Oregon was selected as a context for this study due to its major regulatory undertaking for P2P accommodation businesses (e.g., requirements for business permit and registration, adherence with zoning law, short term rentals and transient lodging taxes, room inspection, etc.), making it the most Airbnb-friendly city in the US (Plautz 2014). The city provides a unique context for this study not only because the regulation allows P2P rentals to serve the tourism market alongside hotels, several quality standards for P2P accommodation listings are put in place to protect consumers and hosts (see City of Portland 2015). Hotel reviews were extracted in November 2014 from a major travel review website by crawling up to 50 pages of reviews for all hotel properties with at least one review, resulting in 18,166 reviews. The same procedure was applied to all available listings in a major P2P rental website, resulting in 2,130 reviews.

The first step of the text analysis is preprocessing the data using *Stanford POS Tagger,* a *Java* implementation of the log-linear part-of-speech (POS) tagging approach described in Toutanova, Klein, Manning, and Singer (2003). Preprocessing includes sentence splitting, tokenization (i.e., breaking a stream of text into tokens), eliminating stop words, POS tagging (i.e., categorization of words with similar grammatical properties into noun, verb, adjective, etc.), and lemmatization (i.e., grouping together the different inflected forms of a word). The preprocessed hotel data consist of 2,609,196 tokens and 33,474 word types, while the P2P accommodation data consist of 151,992 tokens and 5,994 word types as target analysis.

In order to identify major terms and themes that represent important attributes of hotels and P2P accommodation in consumer reviews, the documents were analyzed using lexical analysis, association statistics, and data visualizations. To identify important terminologies used in the reviews, the top keywords from each review corpus are extracted based on the frequency of occurrence of each word (i.e., in its basic lemma), also called term frequency (TF). To extract important compound words (e.g., bigram, trigram, etc.), this study utilizes an automated term recognition (ATR) program called *TermExtract*, a *Perl* implementation to the ATR approach explained in Nakagawa (2000) and Nakagawa and Mori (2002), which obtains domain specific terminologies from documents. The program applies termhood-based approach, which measures the extent to which a candidate term is related to a domain-specific context (Korkontzelos, Klapaftis, and Manandhar 2008), under the assumption that terms with complex structure are made of existing simple terms (Nakagawa 2000; Nakagawa and Mori 2002). Therefore, it measures the termhood of single words 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 term candidate $ct = N_1, N_2 \ldots N_k$, an importance score (*IMP*) is calculated by:

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

In order to incorporate the frequency of independent occurrences of candidate terms, *IMP* is multiplied by the marginal frequency ($MF(ct)$), which is the number of independent occurrences of ct, to obtain the statistical barrier (SB) of ct:

$$SB(ct) = IMP(ct) * MF(ct) \quad (2)$$

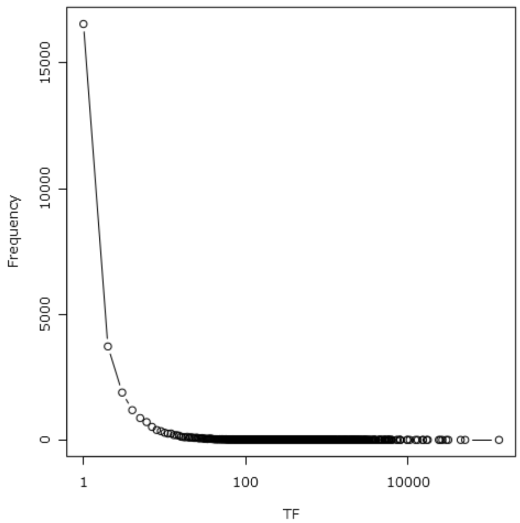
Finally, to examine the distribution of the high frequency words in the documents (i.e., how they are used in connection with each other in one review), word co-occurrence networks are developed using the *igraph* package in *R* statistical program. The nodes of the networks are the high frequency words. The edges of the network are determined by Jaccard Coefficient (Romesburg 1984) of the word pairs. Jaccard Coefficient is a statistical measure used to compare the similarity between finite sample sets (i.e., words), 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:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

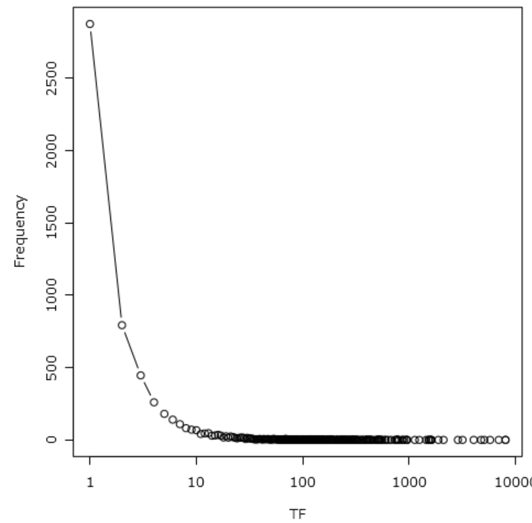
The layout of the networks is determined by the Fruchterman-Reingold's (1991) algorithm, which uses a force-based graph drawing technique to present networks in an aesthetically pleasing way. In order to facilitate further analyses on specific words of interest, co-occurrence networks of associated words around a specific word are also developed following the same approaches.

RESULTS AND DISCUSSION

To capture potential aggregate patterns related to host evaluation in the analyses, all host names (pronouns) were replaced by the word "*pname*". Figure 1 represents the distribution of the term frequency (TF) in both documents. The mean of TF is 49.04 in hotel reviews (words appear 49 times on average) and 25.36 in P2P accommodation reviews. As represented by the long tails in the distribution plots, about 93% of words appear less than 50 times in hotel reviews and 92% appear less than 30 times in P2P accommodation reviews. These indicate that the top keywords (i.e., the most frequently discussed terms) comprise less than 10% of the total word types in the documents.



a. Hotel Reviews
(*Mean* of TF = 49.04, *s.d.* of TF = 959.96)



b. P2P Accommodation Reviews
(*Mean* of TF = 25.36, *s.d.* of TF = 238.32)

Figure 1. Term Frequency (TF) Distribution: Hotel vs. P2P Accommodation Reviews

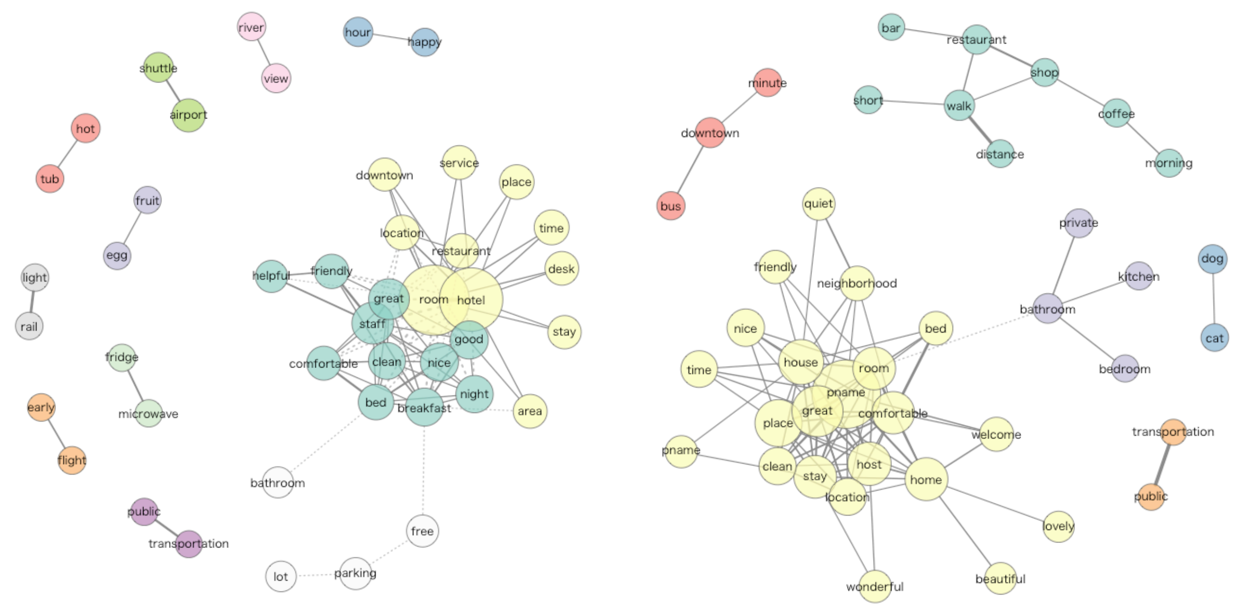
Based on the results of POS tagging, the frequency lists for top nouns (representing attributes) and adjectives (representing assessment) from hotel and P2P accommodation reviews were compared (See Table A1 in Appendix). Unique high frequency nouns in P2P accommodation reviews include *pname*, *place*, *home*, *house*, *neighborhood*, and *experience*; unique nouns in hotel reviews include *staff, breakfast*, *service*, *airport*, *restaurant,* and *parking*. This is an early indication that besides the basic features (*room* and *bed*), attributes related to homes and hosts are the central terms in P2P accommodation reviews as amenities and services in hotel. Additionally, in terms of adjectives, hotel reviews present more factual evaluation terms (e.g., *small*, *big*, *hot*, *old*, *free*), while P2P accommodation reviews also include more emotional evaluation terms (e.g., *cozy*, *warm*, *cute*, *lovely*, *sweet*).

To provide a better understanding on the important topics used in these reviews, the top compound nouns (i.e., word clusters) based on their importance as domain specific keywords in the documents were extracted using the *TermExtract* module (see Table A2 in Appendix). The results demonstrate that for hotel guests, staff and services are among the most important attributes for hotel evaluation as presented by the top word clusters: *front desk staff*, *hotel staff, room service,* and *great service*. Hotel amenities and values (i.e., freebies) are also important terms in the reviews as represented by: *free breakfast*, *free parking*, *breakfast room*, *parking lot*, and *shuttle service*. Finally, the location of the hotels seems to be of importance to hotel guests as it was referenced in: *downtown portland*, *great location, portland airport, airport hotel*, and *portland area*.

The results from term extraction also confirm that P2P accommodation reviews concentrate on the homes and the hosts, as represented in the word clusters: *great host*, *pname place*, *pname house*, *pname home*, *great place*, and *wonderful host*. Further, guests at P2P accommodation also highlighted their general staying experience: *great time*, *great stay*, and *great experience*. Similar to hotel reviews, guests also emphasize the location as important attributes for P2P accommodation, as represented in *great location*, *walk distance*, *downtown portland*, *great neighborhood*, and *quiet neighborhood*. It is important to note that while hotel reviews focus on hotel location in a proximity to the general areas of attractions or facilities (i.e., downtown, airport), the P2P accommodation reviews also pay attention to the neighborhoods where the accommodation is located.

Co-Occurrence Networks

While counting the term frequency alone can be useful in knowledge extraction, examining the distribution of the keywords in the documents in relation to other keywords can be more powerful to understand the contexts in which the keywords are used. The top 100 pairs of high frequency keywords (based on their Jaccard Coefficients) that appear together in one review from hotel and P2P accommodation data are presented in Figure 2. For hotel reviews, the minimum TF was set into 500, allowing the network to be developed from a 269 (words) x 18,166 (reviews) matrix. For P2P accommodation reviews, the minimum TF was set to 50, allowing the network to be developed from a 152 (words) x 2,130 (reviews) matrix. The size of the bubbles (nodes) indicates word frequency; thickness of edges indicates strength of connections (Jaccard coefficient); color indicates word communities in the network (i.e., densely connected subgraphs) detected using random walk method as described in Pons and Latapy (2005).



a. Hotel Reviews (Nodes: 44; Edges: 100; Density: .106; Min TF = 500)

b. P2P Accommodation Reviews (Nodes: 41; Edges: 100; Density: .122; Min TF = 50)

Figure 2. Co-occurrence Networks: Hotel vs. P2P Accommodation Reviews

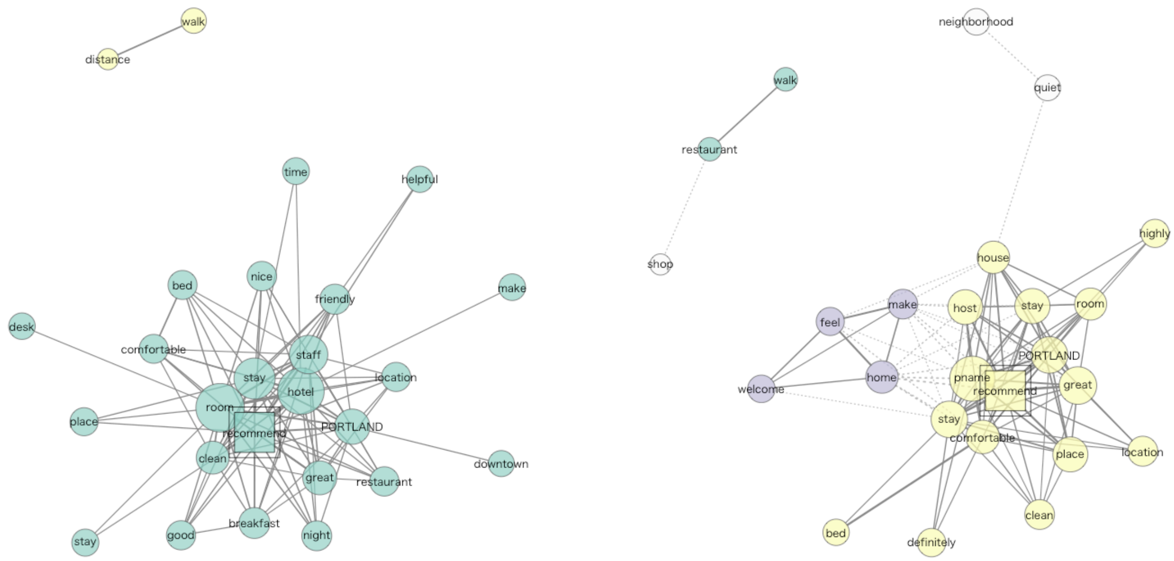
In hotel reviews, two word communities are at the core of the network, one representing services and location (*hotel*, *room*, *service*, *location*, *downtown*, *area,* etc.) and the other representing amenities and staff (*comfortable*, *bed*, *friendly*, *helpful*, *staff*, *breakfast*, etc.). Connected to the core is a theme on deals/freebies (*free*, *parking*, *lot*), representing important complements to the core services. Other topics, isolated from the core, represent added values for guests, including transportation facilities (*public, transportation*, *light, rail*, *airport, shuttle, early, flight*), room features (*hot*, *tub*, *fridge*, *microwave*, *river*, *view*), and food and beverage (*happy*, *hour*, *fruit*, *egg*). These word communities represent tangible cues (e.g., location, physical facilities, amenities, etc.) used to evaluate hotel stay and guest experiences. It is also noteworthy that evaluations on clean room and comfortable bed are connected with those on hotel staff behavior in a dense word community, indicating that these two factors are of equal importance. This is further confirmed with the high degree centrality and betweenness centrality of the words *hotel*, *room*, and *staff*, indicating importance of these words in the network and in bridging between other nodes.

In P2P accommodation reviews, there is one community at the core of the network, which is centered on *pname* (the hosts), capturing the host (*friendly*, *host*), the home (*welcome*, *home*, *lovely*, *wonderful*), core services (*clean*, *comfortable, bed*, *room*), and location (*location*, *quiet*, *neighborhood*). This core word community indicates not only that these attributes are of equal importance (i.e., often discussed together), but also the main criteria used by guests to evaluate P2P accommodation. The words *pname* and *home* have the highest degree and betweenness centrality, indicating that they are important nodes that bridge other nodes in the network. A word community with direct connection to the core represents room amenities (*private*, *bathroom*, *kitchen*, *and bedroom*). Other word communities are around convenience, including short walking distance to shop and restaurants for morning coffee, minutes to downtown by bus, and public transportation. Compared to the hotel reviews, more positive emotional expressions are used in P2P accommodation reviews.

The "Recommend" Network

In order to identify which attributes of hotels and P2P accommodation have strong connections with post-purchase behavior, co-occurrence networks of associated words around the word *recommend* are developed. The underlying assumption is that the themes that are highly connected with the word *recommend* can be used to predict the willingness of guests to recommend the hotel or P2P accommodation listing to others. First, the lists of associated words are consulted and ranked based on Jaccard Coefficients. Then, networks were developed using top 100 connections between the word *recommend* and its associated words.

Figure 3 illustrates the *recommend* networks in hotel and P2P accommodation reviews that include top 100 connections. In hotel reviews one word community is at the core, with attributes representing the core services (*hotel*, *clean*, *room*, *stay, bed, comfortable*), location (*location*, *Portland*, *downtown*), staff (*friendly*, *helpful*, *staff*) and added services (*good*, *breakfast*). One community in the periphery represents convenience (*walk*, *distance*), which reinforces hotel location. In P2P accommodation reviews, the core community is around host, home, basic services (*comfortable*, *bed*), and location. Connected to the core is an important community of peer hospitality (*make*, *feel*, *welcome*, *home*), occurring when hosts make their guests feel welcome. Additionally, communities around quiet neighborhood and convenience to restaurants and shops indicate the importance of location of the listings. These attributes are believed to have a great impact on guests' willingness to recommend the accommodation listings to others.



a. Hotel Reviews
(Nodes: 26; Edges: 100; Density: .308)

b. P2P Accommodation Reviews
(Nodes: 25; Edges: 100; Density: .333)

Figure 3. Co-occurrence Networks around *Recommend*

CONCLUSION AND RECOMMENDATION

Based on the analyses, P2P rental reviews put more emphasis on the hospitality of the hosts (i.e., the experience of being welcome in someone's home) and the locale, with guests highlighting the quiet neighborhood within short walking distances to local restaurants and shops as well as within minutes by bus to downtown. This indicates that P2P rental is a desirable option for guests driven by desire for community and social interactions (Guttentag 2013; Owyang 2013; Tussyadiah 2015). On the other hand, hotel offerings unparalleled by P2P rentals include conveniences such as airport shuttle services for guests with early morning flights, free parking, good breakfast options, and in-room services. Therefore, it can be concluded that even though consumers expect similar core services (i.e., nice, clean room and comfortable bed), reviews suggest different attributes supporting the competitive advantage among hotels and P2P rentals. The results from this study inform both hotels and P2P accommodation with their advantages and disadvantages based on guest evaluations and, as a consequence, insights on how to augment their advantages and overcome their weaknesses and threats from each other. In order to appeal to consumers' social and environmental motivations, hotels could make guest experiences more personal (e.g., staff meaningfully interact with guests) and socially responsible. On the other hand, P2P rentals could offer additional conveniences for guests (e.g., morning coffee, private bathroom, insider tips for local attractions and facilities, etc.).

This study confirms the different key attributes that mark the competitive edge of P2P rentals and hotels from analyzing consumer intelligence obtained from extracting UGC. This confirms the potentials of applying text analytics to larger unstructured text data to gain valuable market and competitive intelligence to support marketing and management decisions for tourism and hospitality. Future studies should include reviews in other tourism destinations to confirm the generalizability of these findings. Also, in order to elucidate the differences among hotels in different locations and with different quality standards, it would be beneficial for future studies to analyze the attributes of hotels located in downtown and airport, to compare upscale, mid-scale and budget hotels, as well as P2P listings with different categories (i.e., entire home, private room, shared room) and price structures.

References

Balmer, S., and T. Baum (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.

Chan, J. K. L., and T. Baum (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., R. H. L. Chiang, and V. C. Storey (2012). "Business intelligence and analytics: From big data to big impact." *MIS Quarterly*, 56(4), 1165-1188.

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

Clow, K.E., J. A. Garretson, and D. L. Kurtz (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.

Dolnicar, S. (2002). Business travelers' hotel expectations and disappointments: A different perspective to hotel attribute importance investigation. *Asia Pacific Journal of Tourism Research*, 7(1), 29-35.

Dolnicar, S. and T. Otter (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., P. Gerstl, and R. Seiffert (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.

Fan, W., L. Wallace, S. Rich, and Z. Zhang (2006). "Tapping the power of text mining." *Communications of the ACM*, 49(9), 77-82.

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

Geron, T. (2012, November). "Airbnb had $56 million impact on San Francisco: Study." *Forbes*. http://www.forbes.com/sites/tomiogeron/2012/11/09/study-airbnb-had-56-million-impact-on-san-francisco/.

Geron, T. (2013, January). "Airbnb and the unstoppable rise of the share economy." *Forbes*. http://www.forbes.com/sites/tomiogeron/2013/01/23/airbnb-and-the-unstoppable-rise-of-the-share-economy/.

Guttentag, D. (2013). “Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector.” *Current Issues in Tourism*, ahead of print, 1-26.

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

Korkontzelos, I., I. P. Klapaftis, and S. Manandhar (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.

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

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., and T. Mori (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

Plautz, J. (2014, July). “Portland Could Soon Be the Most Airbnb-Friendly City in the U.S.” *Mashable*. http://mashable.com/2014/07/23/portland-airbnb/

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

Romesburg, H. C. (1984). *Cluster Analysis for Researchers*. Belmont, CA: Lifetime Learning Publications.

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

Tsao, W.-C., M.-T. Hsieh, L.-W. Shih, and T. M. Y. Lin (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 *Information & Communication Technologies in Tourism 2015*, edited by I. Tussyadiah, and A. Inversini. Switzerland: Springer, pp. 819-32.

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

Xiang, Z., Z. Schwartz, J.H. Gerdes, and M. Uysal (2015). “How 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., and C.-H. Tang (2015). “Hotel attribute performance, eWOM motivations, and media choice.” *International Journal of Hospitality Management*, 46, 79-88.

Zervas, G., D. Proserpio, and J. Byers (2014). “The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry.” *Boston University School of Management Research Paper* No. 2013-16. http://ssrn.com/abstract=2366898.

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

APPENDIX

Table A1. Word Frequency Lists: Hotel vs. P2P Accommodation Reviews

	Hotel Reviews				P2P Accommodation Reviews			
	Noun	Freq.	Adjective	Freq.	Noun	Freq.	Adjective	Freq.
1	room	29624	great	9765	*pname*[*]	1883	great	1130
2	hotel	24995	good	8019	place	959	comfortable	755
3	staff	10493	nice	8018	house	871	nice	545
4	breakfast	7917	clean	7211	host	804	clean	539
5	night	7471	friendly	5465	home	778	welcome	399
6	bed	6297	comfortable	5249	stay	772	friendly	369
7	location	5863	helpful	3989	room	771	quiet	327
8	time	5829	free	3918	location	530	wonderful	324
9	restaurant	5659	other	3785	time	510	beautiful	300
10	place	5648	front	3508	neighborhood	426	easy	291
11	service	5194	next	2882	bed	385	lovely	272
12	area	5154	small	2855	experience	299	good	256
13	desk	5051	little	2478	area	272	perfect	255
14	stay	4799	excellent	2384	everything	266	helpful	227
15	airport	4629	large	2208	night	258	super	183
16	downtown	4620	many	2157	walk	239	first	182
17	day	4298	quiet	2138	day	234	warm	173
18	parking	3909	best	2031	downtown	225	private	171
19	bathroom	3235	first	2029	bathroom	211	cozy	163
20	food	3172	easy	1950	space	207	amazing	155
21	lot	2937	few	1882	restaurant	193	little	154
22	floor	2755	hot	1746	friend	180	close	136
23	lobby	2661	wonderful	1733	dog	169	next	134
24	morning	2628	more	1667	guest	159	awesome	133
25	price	2543	convenient	1653	thanks	159	short	131
26	door	2511	close	1645	breakfast	153	convenient	129
27	shuttle	2419	old	1624	bus	148	best	125
28	pool	2418	only	1578	coffee	147	excellent	123
29	bar	2413	spacious	1514	city	146	few	123
30	coffee	2385	big	1340	minute	140	able	115
31	car	2379	better	1293	kitchen	137	cool	111
32	minute	2225	sure	1282	morning	134	more	111
33	thing	2191	perfect	1254	shop	133	sure	102
34	trip	2147	available	1206	apartment	131	fantastic	99
35	noise	2112	happy	1203	thing	128	many	94
36	hour	2085	new	1202	distance	126	spacious	93
37	way	2071	last	1181	people	121	cute	87
38	street	2055	several	1170	fun	120	safe	86
39	people	2023	able	1132	cat	119	sweet	81
40	everything	2014	full	1121	town	119	gracious	80

**pname* = host names

Table A2. Word Clusters: Hotel vs. P2P Rental Reviews

	Hotel Reviews		**P2P Accommodation Reviews**	
	Word Cluster	Score (*SB*)	Word Cluster	Score (*SB*)
1	front desk	3852053.21	great host	16705.65
2	downtown portland	1884261.79	*pname** place	9538.94
3	front desk staff	966341.59	great location	8254.43
4	hotel staff	900713.58	*pname** house	8234.29
5	room service	767249.36	*pname** home	7929.20
6	great location	747242.17	great place	6960.59
7	hotel room	682530.19	great time	5954.00
8	free breakfast	589750.52	great stay	5447.29
9	great hotel	571924.91	wonderful host	5206.74
10	free parking	486449.85	great experience	3910.32
11	portland airport	478939.05	walk distance	3662.80
12	nice hotel	416276.66	downtown portland	3420.83
13	breakfast room	377691.76	first **** experience	3007.00
14	great place	335365.09	great neighborhood	2996.38
15	parking lot	327771.34	next time	2862.54
16	hotel restaurant	301566.02	comfortable bed	2844.03
17	portland area	294135.99	first time	2793.28
18	continental breakfast	283150.28	quiet neighborhood	2042.32
19	next time	267269.38	beautiful home	1989.28
20	shuttle service	253755.11	short walk	1955.34
21	great service	245843.31	portland area	1782.30
22	next morning	242371.64	wonderful stay	1642.86
23	airport hotel	240925.45	minute walk	1559.37
24	next door	235571.95	alberta street	1479.22
25	friendly staff	228902.38	great restaurant	1432.35
26	good location	226417.89	excellent host	1407.71
27	downtown area	209266.54	lovely home	1393.98
28	valet parking	199129.12	lovely host	1313.82
29	great staff	194382.78	friendly host	1309.35
30	customer service	192874.10	gracious host	1261.83
31	other hotels	191481.62	perfect location	1170.04
32	light rail	188538.56	perfect host	1133.49
33	complimentary breakfast	187471.86	one night	1130.10
34	desk staff	176508.81	guest house	1104.69
35	nice room	172512.36	wonderful experience	1056.59
36	downtown hotel	170153.59	beautiful house	1031.82
37	breakfast buffet	168749.39	alberta arts district	983.45
38	great stay	168635.33	comfortable home	968.82
39	free shuttle	168381.43	**** experience	968.39
40	good hotel	166381.03	private bathroom	965.28

**pname* = host names