

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

## Social Interactions or Business Transactions? What customer reviews disclose about Airbnb marketplace.

### This is the author's manuscript

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/1730546> since 2024-04-03T08:06:52Z

*Publisher:*

International World Wide Web Conferences Steering Committee

*Published version:*

DOI:10.1145/3366423.3380225

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

# Social Interactions or Business Transactions?

## What customer reviews disclose about Airbnb marketplace

Giovanni Quattrone  
g.quattrone@mdx.ac.uk  
Middlesex University, UK  
University of Turin, Italy

Licia Capra  
l.capra@ucl.ac.uk  
University College London, UK

Antonino Nocera  
antonino.nocera@unipv.it  
University of Pavia, Italy

Daniele Quercia  
daniele.quercia@kcl.ac.uk  
King's College, UK  
Nokia Bell Labs, UK

### ABSTRACT

Airbnb is one of the most successful examples of sharing economy marketplaces. With rapid and global market penetration, understanding its attractiveness and evolving growth opportunities is key to plan business decision making. There is an ongoing debate, for example, about whether Airbnb is a hospitality service that fosters social exchanges between hosts and guests, as the sharing economy manifesto originally stated, or whether it is (or is evolving into being) a purely business transaction platform, the way hotels have traditionally operated. To answer these questions, we propose a novel market analysis approach that exploits customers' reviews. Key to the approach is a method that combines thematic analysis and machine learning to inductively develop a custom dictionary for guests' reviews. Based on this dictionary, we then use quantitative linguistic analysis on a corpus of 3.2 million reviews collected in 6 different cities, and illustrate how to answer a variety of market research questions, at fine levels of temporal, thematic, user and spatial granularity, such as (i) how the business vs social dichotomy is evolving over the years, (ii) what exact words within such top-level categories are evolving, (iii) whether such trends vary across different user segments and (iv) in different neighbourhoods.

### CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**; • **Applied computing** → *Law, social and behavioral sciences*.

### KEYWORDS

Sharing economy, Airbnb, linguistic analysis, market analysis

#### ACM Reference Format:

Giovanni Quattrone, Antonino Nocera, Licia Capra, and Daniele Quercia. 2020. Social Interactions or Business Transactions? What customer reviews disclose about Airbnb marketplace. In *Proceedings of The Web Conference 2020 (WWW '20)*, April 20–24, 2020, Taipei, Taiwan. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3366423.3380225>

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '20, April 20–24, 2020, Taipei, Taiwan

© 2020 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.

ACM ISBN 978-1-4503-7023-3/20/04.

<https://doi.org/10.1145/3366423.3380225>

### 1 INTRODUCTION

The sharing economy, also known as peer-to-peer or collaborative economy, is an economic model based on a distributed network of individuals, directly accessing each other underused assets. Airbnb is one of the most successful examples of such model, with hosts renting out their unused rooms or entire properties by directly engaging in computer-mediated transactions with potential guests. Since its creation in 2008, Airbnb has been experiencing exponential growth, which continues to date. According to recent statistics,<sup>1</sup> the company is currently operating in more than 65,000 cities worldwide, with over 6M listings to choose from, and serving over 2M people on any given night. Airbnb marketplace is not only growing but also very rapidly evolving: for example, while millennials still make up the largest portion of user share at 60%, in the last two years the fastest growing host demographic has been in senior hosts over 60, with a growth rate of 102%. On top of demographic diversification, Airbnb has been experiencing geographic habit diversification, too: for example, the average Berlin guest stays for 6.3 nights, as opposed to the average Amsterdam guest who stays for 3.9 nights only.

One of the challenges that companies like Airbnb face is to understand their attractiveness, as well as their evolving market opportunities, in the face of such rapid and diversifying growth rates. Traditional market research techniques, based on customer surveys and focus groups, offer very detailed insights that can help inform business decision making, but require substantial financial and time investments. As a result, their use in the sharing economy context is limited, due to the fast-evolving and global nature of most such markets. In this setting, more agile techniques are needed to allow companies to strategise promptly. For example, there is an ongoing debate about whether Airbnb is a hospitality service that fosters social exchanges between hosts and guests, as the sharing economy manifesto<sup>2</sup> originally stated, or whether it is (or is evolving into being) a purely business transaction platform, the way hotels have traditionally operated. Being able to assess to what extent Airbnb customers value social interactions vs. business transactions has important implications for how the company may decide to operate, and compete, in the hospitality service. Given

<sup>1</sup><https://ipropertymanagement.com/airbnb-statistics>

<sup>2</sup><http://www.thepeoplewhoshare.com/>

Airbnb different usage patterns in different cities, such market analysis needs to be performed separately in each geographic context the company operates; furthermore, because of the rapidly evolving demographics of its customers, the analysis needs to be repeated frequently, to capture varying trends.

In this paper, we propose a scalable market analysis approach, to complement and enrich traditional ones. Instead of collecting primary data via interviews, focus groups and surveys, our approach exploits ready-available secondary data that most sharing economy platforms like Airbnb possess: a continuous stream of reviews that peers leave upon completion of a service exchange. Key to our approach is a new semi-supervised method to inductively develop platform-specific dictionaries starting from peers' reviews. The method combines qualitative thematic analysis with quantitative machine learning techniques in a novel way, and enables the construction of a dictionary that captures topics disclosed in customers' reviews at different levels of granularity. Based on this purpose-built dictionary, we then define robust topic-adoption metrics that enable us to explore a variety of market research questions, at fine levels of thematic, temporal and spatial granularity.

We specifically illustrate our proposed market analysis approach using the case of Airbnb, and while doing so, we make the following two main contributions:

(1) *Dictionary construction.* We gather 3.2M Airbnb guest reviews about 176K distinct listings, spread across 6 different cities (London, Manchester, New York, San Francisco, Melbourne, Sydney), written between 2010 and 2019 (Section 3). These cities have been chosen so to span different continents (America, Europe, Oceania), later affording us the ability to explore whether trends are geographically bounded or not. Note that, at this stage, we are focusing on reviews written in English only; these represent 90% of all reviews left for properties in these cities. We then analyse these reviews using a combination of thematic analysis and machine learning, and build a dictionary that is capable of classifying words (unigrams) at three levels of granularity: two top-level categories (i.e., 'social' interactions vs. 'business' transactions), four distinct sub-categories and 13 subsub-categories (Section 4).

(2) *Market analysis.* We illustrate how to use the purpose-built dictionary, in combination with robust topic-adoption metrics, to understand to what extent Airbnb guests discuss the social aspect vs. the business aspect of their hospitality experience (Section 5). We do this by exploring four different market research questions that illustrate the ability of our dictionary and analytical approach to address questions at varying levels of detail, while also scaling easily over time and geographic location. We find that, across the 6 cities analysed, business aspects are increasingly being discussed in guests' reviews, while social aspects are steadily declining (Section 5.1). This trend is happening not just at the top-level categories (business vs. social), but across all words in our lexicon (Section 5.2). We then segment Airbnb hosts according to the time they joined the platform, and discover that those who joined at the very beginning (i.e., the so called 'innovators' [36]), are those receiving guests' reviews that most dwell on the social aspects of their hospitality experience, and they remain so over the years. On the contrary, hosts who joined the platform later ('early adopters' and 'early majority'), consistently receive more business-dominated reviews

across all cities (Section 5.3). Finally, we zoom in within each city, to understand whether there is market diversification in different neighbourhoods, and discover that properties in areas of low Airbnb penetration (less tourist areas) receive reviews that discuss social aspects of the experience significantly more than those in areas with higher Airbnb penetration (more tourist areas). Once again, this pattern is consistent in all cities analysed, despite them being located in different countries/continents (Section 5.4).

We conclude this paper with a discussion about practical uses of the proposed method, its current limitations, and possible future developments (Section 6).

## 2 RELATED WORK

Sharing economy platforms like Airbnb have been extensively studied in the past, following two broad lines of inquiries.

A first line of inquiry has analysed the relationship between *sharing economy services and society*, specifically at the level of cities [10, 18, 29, 35, 37, 45, 46]. Several studies have looked into the relationship between these novel services and their traditional counterparts, with findings that often varied depending on geographic location: some scholars found that these new services only marginally disrupt their established counterparts (e.g., Uber vs. taxis, Airbnb vs. hotels) [45]. As an example, in London, the geographical overlap between Airbnb properties and hotels was found to be marginal [35]; furthermore, sharing economy services were found to bring positive effects to the broad tourism industry [10]. Other scholars found opposite results instead: a study performed in Budapest showed that Airbnb and hotels were located in the same central areas, causing fierce competition between the two [4]. Other studies have looked at the relationship between Airbnb and the housing/rental market [39, 42], with findings suggesting that Airbnb is accelerating an ongoing processes of gentrification in London.

A complementary line of inquiry has focused on the relationship between *sharing economy services and people*. Several studies have looked into motivational factors for user participation in such platforms. Using online surveys and host/guest interviews, these investigations have revealed that financial benefits are an important factor for Airbnb hosts to join such platforms, but they do not represent the only factor, as business (financial) reasons and social reasons are intertwined with one other [3, 15, 19, 25, 28, 38]. Whether this is changing over time, and in different locations, is hard to answer, since the primary data used to perform such studies (e.g., survey data) is very costly to obtain (both financially and in terms of time). Other studies have used ready available data from within these online platforms instead, primarily to study user satisfaction with the service provided. An analysis of Airbnb ratings has revealed that 95% of properties in Airbnb boast an average user-generated rating above 4.5 stars [6, 44]; this is in sharp contrast with platforms like TripAdvisor, where the average star rating is 3.8 [44]. Sentiment analysis conducted on reviews seemed to corroborate this finding [1, 13, 26, 30], although the authors caution against a phenomenon of "socially induced reciprocity" which may occur when peers interact socially with one another, leading to negative information being omitted from reviews. Scholars have used sentiment analysis on user reviews to shed light on price dynamics too,

revealing that the price of Airbnb properties is greatly influenced by their review score, after controlling for characteristics of the room and features of the neighbourhood [26].

Recently, reviews have increasingly been used as main data source in sharing economy platform studies [7, 21, 27, 31, 32], not only because they are readily available, but also because, with over 70% of guests writing a review after a stay [12], they can offer very good coverage of peers’ experiences in such platforms. For example, in [22] researchers collected a sample of hosts’ profiles and guests’ reviews in AirBnB and Couchsurfing; after manually labelling and analysing them, they found initial evidence that the primary shared asset in AirBnB is the house (i.e., its facilities, location, neighbourhood), while in Couchsurfing it is the human relationship (i.e., host-guest interaction, experience, self-description, motivation). This finding is corroborated by another study that used interviews as primary data source instead: in [24], 17 users who had participated in both Airbnb and Couchsurfing were interviewed, revealing that Airbnb peers require higher quality services, and put more emphasis on places over people. The same study [24] also analysed 5k random reviews from Couchsurfing and Airbnb using the general-purpose LIWC dictionary [34]. Once again, results confirmed that Airbnb reviews are more business oriented, whereas Couchsurfing reviews are more person-oriented; since the LIWC dictionary is platform-independent, it is not possible to delve deeper into this business vs. social dichotomy. To zoom in further, recent studies have taken an orthogonal approach, mining reviews in an unsupervised fashion, and analysing platform-specific emerging topics: for example, in [7] topics such as ‘location’, ‘amenities’ and ‘host’ appear to automatically emerge; in [31], the five most common aspects of Airbnb reviews that emerge seem to be the communication between guest and host, the experience of the rental, the location of the property, the service offered, and the value of the property. Both studies suggest once again that the nature of Airbnb is mainly about accessing assets rather than sharing them.

In this paper, we further expand on this latter line of inquiry, and propose a mixed-method approach that combines thematic analysis of guest reviews with unsupervised machine learning techniques, to inductively build a dictionary that enables fine-grained and scalable market analysis of platforms such as Airbnb. Unlike unsupervised topic detection techniques (e.g., Latent Dirichlet Allocation model [20]), our approach does not suffer from the problem of over-fitting that is common when text length is short, as it is often the case in reviews (e.g., [8]). Furthermore, unlike approaches that rely on general-purpose dictionaries such as LIWC [34], our approach affords the exploration of platform-specific market research questions (rather than platform-agnostic explorations about, for example, sentiment analysis and mood detection [40]). Before presenting our proposed method, we briefly introduce the dataset we collected.

### 3 DATASET

We gathered Airbnb data from the “Inside Airbnb” organisation (<http://insideairbnb.com/>), containing snapshots of Airbnb listings and reviews around the world collected at regular time intervals (typically, at least once per quarter from 2015, and more often in the last couple of years).

City	# Reviews	Year	# Reviews
Greater Manchester	91,967	2010	1,805
London	992,638	2011	7,398
Melbourne	469,906	2012	20,091
New York City	883,280	...	...
San Francisco	286,592	2017	706,556
Sydney	438,491	2018	1,101,528
		2019	500,834

Table 1: Reviews by city and by year

On June 3rd 2019, we gathered all the listings and reviews associated with six different cities: Greater Manchester (U.K.), London (U.K.), Melbourne (Australia), New York City (U.S.), San Francisco (U.S.), Sydney (Australia). We selected these cities for the following two reasons. First, we did not want to add the inherent noise incurred when performing language translation; we thus favoured cities in English-speaking countries, for which we expected the vast majority of reviews to be written in English. Second, within this constraint, we wanted to consider cities belonging to different countries and continents, so to later explore whether our findings are country/continent bounded or they generalize.

We initially collected 3.9 million Airbnb guest reviews associated with 176 thousand distinct listings. To gain confidence in the validity of the data, we selected 10 random listings, along with their associated reviews in each city and verified their existence on the original Airbnb platform. After this preliminary check, we analysed review length distribution, and removed reviews that were either too short or too long (less than 5 words and more than 175 words – which are about 8% of the original reviews). We further removed reviews automatically generated by the system in case of a cancellation (around 2%); reviews without a year and without comments (less than 1%); reviews generated by power users (i.e., guests who wrote more than 10 reviews) who may bias results (less than 1%), and finally non English reviews (around 5% of reviews removed). We ended up with a dataset comprising 3.2 million guests’ reviews, whose composition by city and by year is shown in Table 1.

## 4 DICTIONARY CONSTRUCTION, ADOPTION AND VALIDATION

In this section, we present a mixed-method approach that combines thematic analysis with machine learning techniques to inductively build a platform-specific (in this case, Airbnb) dictionary that affords us the ability to group the lexicon used in Airbnb guest reviews into categories concerning ‘social interactions’ vs. ‘business transactions’ at different levels of granularity (Section 4.1). We then define metrics to be computed on top of this dictionary (Section 4.2), and report on dictionary and metric validation steps we have conducted (Section 4.3).

### 4.1 Building a Dictionary

We built our dictionary in five steps: first, we developed a coding scheme by performing thematic analysis of a random sample of 100 Airbnb reviews (step 1); second, we refined and validated the coding scheme by means of a crowd-sourcing study conducted on the Crowdfunder<sup>3</sup> platform (step 2), where we asked crowd-workers

<sup>3</sup>Crowdfunder is a crowd-sourced market of online workforce to clean, label and enrich data: <https://www.crowdfunder.com/>.

to label another random set of 100 reviews. Third, we conducted a second study on Crowdfunder, this time asking crowd-workers to label a larger set of 1,500 reviews, using the identified themes (step 3). Using natural language processing techniques, we then defined a lexicon of the words most representative of each such theme (step 4). Finally (step 5), using hierarchical clustering techniques, we grouped together these words into 13 distinct clusters, which represent a finer-grained refinement of the themes manually identified at steps 1 and 2. Our final dictionary comprises two level-1 categories (i.e., business vs social), refined into four level-2 (sub)categories, further refined into thirteen level-3 (subsub)categories, which semantically group together a lexicon of 355 words. We discuss the details of each step next.

*Step 1. Developing a Coding Scheme.* Using stratified sampling to cover all study years and cities, we sampled 100 Airbnb reviews. We broke down each review into its constituting sentences, and performed a thematic analysis over these. In a way similar to [5], two independent annotators coded these resulting sentences by performing three steps: (i) familiarising with the data, (ii) generating the initial codes and searching for themes among codes, and (iii) defining themes. After a first round of coding, the two coders compared their results, and agreed on which themes to maintain, remove, amend, or merge. As a result, they agreed on five main themes named ‘property’, ‘location’, ‘business conduct’, ‘personality’, and ‘social interaction’. The first three are refinements of the theme ‘business’ and the last two of the theme ‘social’.

*Step 2. Validating the Coding Scheme.* To gain confidence in the validity of the coding scheme, we asked crowd-workers to annotate sentences extracted from a new sample of 100 Airbnb reviews using these five themes. In particular, we prepared a Crowdfunder page that consisted of three sections: (i) a list that showed our five themes; (ii) for each theme, actual examples of Airbnb reviews manually labelled by us; and (iii) new Airbnb sentences to be labelled. We paid 0.01\$ per annotation, and each Airbnb sentence was independently annotated by at least four different workers. We computed the Fleiss’ kappa agreement score for the five themes [11], and two of them (i.e., ‘personality’ and ‘social interaction’) had a Fleiss’ kappa score less than 0.5. We merged these two themes into one, resulting in four themes: ‘property’, ‘location’, ‘professional conduct’ and ‘social interaction’. To ascertain the effectiveness of coding with those four themes, we again asked crowd-workers to annotate a new sample of sentences extracted from yet another 100 Airbnb reviews. All four themes resulted in a Fleiss’ kappa score higher than 0.5, suggesting their validity.

*Step 3. Labelling Reviews.* We were then ready to label a larger set of Airbnb reviews using the identified four themes. We used again Crowdfunder to annotate unlabelled sentences extracted from a new set of 1,500 reviews. We gathered 22,975 distinct annotations of 4,062 sentences. We kept those sentences on which at least 75% of annotators agreed – so to have high confidence that the words inferred from these sentences are reliable – and ended up with a set of 1,868 sentences having high agreement. The second column of Table 2 shows the frequency of occurrence of each of the four themes in these sentences. The most popular theme was ‘property’,

followed by ‘location’ and ‘professional conduct’; ‘social interaction’ was the least frequent theme instead.

Theme	Frequency	Initial words	Expanded words
Property	35%	63	77
Location	28%	97	109
Professional Conduct	23%	107	119
Social Interaction	14%	61	68

**Table 2: Inferred four themes along with their frequency, number of words in each theme before and after enrichment**

*Step 4. Building the Dictionary.* To build a dictionary, we needed to identify a lexicon (that is, list of words) that could represent the four themes above. We did so in a data-driven fashion. First, for each theme  $\tau$ , we split the 1,868 annotated sentences into two sets: (i)  $Set_\tau$ , that is the set of sentences labelled with the theme  $\tau$  by at least three quarter of workers; and (ii)  $Set_{\bar{\tau}}$ , that is the set of sentences labelled with the theme  $\tau$  by at most one worker. Second, we extracted all words from  $Set_\tau$  and  $Set_{\bar{\tau}}$ . For each word  $w$ , we computed two measures:  $tf(w, \tau)$  and  $tf(w, \bar{\tau})$ , respectively denoting the term frequency of  $w$  in  $Set_\tau$  and in  $Set_{\bar{\tau}}$ . Finally, we computed  $tf_{gain}(w, \tau) = \frac{tf(w, \tau)}{tf(w, \bar{\tau})}$ .

For each theme  $\tau$ , we then associated all the words  $w$  such that  $tf(w, \tau) \geq tf_{min}$ ,  $tf(w, \tau) \leq tf_{max}$  and  $tf_{gain}(w, \tau) \geq tf_{gain}$ , with  $tf_{min}, tf_{max} \in [0, 1]$  and  $tf_{gain} \in [1, +\infty)$ . The first two thresholds,  $tf_{min}$  and  $tf_{max}$ , allowed us to remove extremely unpopular and extremely popular words respectively. The use of the last threshold  $tf_{gain}$  enabled us to associate to a theme  $t$  only those words that were *comparatively more popular* in  $Set_\tau$  than in  $Set_{\bar{\tau}}$ . Since there is no ground-truth about what a dictionary should look like, automated parameter tuning was not viable. Rather, different thresholds needed to be manually tested and validated. To this purpose, we followed a methodology resembling the Elbow criterion [23]. Specifically, we considered the following threshold values:  $tf_{min} = \{0.001, 0.01, 0.05\}$ ,  $tf_{max} = \{0.15, 0.30, 0.60, 1\}$ ,  $tf_{gain} = \{1.5, 2, 3, 4, 6\}$ . We started with the the most restrictive combination of  $tf_{min}$ ,  $tf_{max}$  and  $tf_{gain}$ ; that is, the combination of parameters generating the smallest dictionary. This combination was  $tf_{min} = 0.05$ ,  $tf_{max} = 0.15$  and  $tf_{gain} = 6$ . We then changed each threshold value iteratively, with each iteration adding a new set of words to the dictionary. We manually validated this added set of words and measured the ratio of noise; that is, the ratio of words that according to our (human) judgement were incorrectly assigned to a particular category. We stopped our search for the best combination of parameters when this ratio was significantly higher than the one identified at the previous step. We ended up with the following manually tuned thresholds:  $tf_{min} = 0.01$ ,  $tf_{max} = 0.15$  and  $tf_{gain} = 3$ . The third column of Table 2 summarises the number of words that each theme contained at this point.

We then used a word embedding machine learning technique (i.e., word2vec [14]) to further enrich our initial lexicon. We started by training the technique on the whole corpus of 3.2M reviews, and mapped each word into a vector having 50 dimensions. For each word already present in our lexicon, we then computed a list of similar words, that is, a list of words having a cosine similarity higher than a threshold  $th_{cos}$ . We included these words as part

of the lexicon of our dictionary if they were not already present. In so doing, we enriched our dictionary with words that are not frequently used in the 1,868 labelled sentences, but still widely used in the whole corpus of reviews (and similar to those previously derived from our labelled corpus). We used a procedure similar to the one described above to manually tune  $th_{cos}$ . The threshold values considered during this step were  $th_{cos} = \{0.6, 0.7, 0.8, 0.9\}$ ; the manually tuned value chosen in the end was  $th_{cos} = 0.7$ . The last column of Table 2 shows the total number of words belonging to each of the four themes after this enrichment step.

*Step 5. Identifying categories at different levels of granularity.* A manual inspection of our expanded lexicon revealed that several sub-themes could be identified within the four main ones that we manually coded at steps 1 and 2. For example, under theme ‘social interaction’, we identified both words that refer to *whom* the peers interacted with (e.g., husband, wife, daughter) as well as *how* (e.g., meals together, talking). In order to offer a more fine-grained taxonomic structure on top of our lexicon, we used a clustering algorithm. For each of the 4 themes in turn, we took all the words associated with them and placed them in a single cluster. We then iteratively increased the number of clusters until the ‘optimal’ number of clusters was found. We chose k-means as clustering algorithm, with the Elbow method [23] applied to find the optimal number of clusters. We ended up with 13 clusters: three clusters were refinements of the ‘property’ theme, four clusters were refinements of the ‘professional conduct’ theme, and a further five of the ‘social interaction’ theme. The ‘location’ theme was mapped to a single cluster, without further refinement.

Table 3 provides an overview of the final dictionary we built. Themes were directly mapped into a 3-tier hierarchical structure consisting of two level 1 categories (that is, ‘business’ and ‘social’), four level 2 categories (that is, ‘property’, ‘location’, ‘professional conduct’, and ‘social interaction’), and thirteen level 3 categories (those automatically inferred by our clustering analysis).<sup>4</sup> An example of lexicon for each category is also provided (the top five words by inverse order of term frequency), together with the intra-cluster similarity computed for each level 3 category (the higher the value, the more cohesive the words inside the cluster). The full dictionary is available for download at <https://figshare.com/s/991c8677e3e9ce013774>.

Quite interestingly, the clusters corresponding to *property* directly matched the property description fields of Airbnb listings – that is, property type (e.g., whether a house or a flat), internal layout (e.g., kitchen, bed, cozy), and facilities (e.g., wifi, tv, fridge). In terms of *professional conduct*, distinct elements have been detected: basic communication (e.g., questions, quick, responded), handling of logistics (e.g., check in, arrival), and provision of advice (e.g., tips, directions). For *social interaction*, five level-3 categories have emerged from clustering, these being ‘people’ (e.g., with whom the guests interact – e.g., husband, wife), what their ‘personality’ is (e.g., friendly, kind, warm), if/what they are ‘sharing’ (e.g., share, stories, experiences), and the how – ‘talking’ (e.g., chat, talking, conversation) over a ‘meal’ (e.g., breakfast, dinner together).

<sup>4</sup>Note that the name of the sub-theme was assigned by us after clustering.

## 4.2 Adopting our dictionary

Having built the dictionary above, our next step is to define metrics operating with its categories (from level 1 to level 3), and its lexicon.

*Metric operating on the dictionary categories.* The first metric we define works at the category level, and it is called *adoption*. As its name suggests, it measures the adoption of a specific category on a given set of reviews. Specifically, let  $R$  be a set of reviews (e.g., reviews left in a given year and/or city), let  $r \in R$  be a specific review belonging to  $R$ , and let  $c$  be the category (of any level, from level 1 to level 3) under consideration. Let us define as  $W$  the set of words contained in  $R$  and as  $C$  the set of words belonging to category  $c$ . For each word  $w \in W$  contained in the review  $r$ , we compute the logarithmically scaled term frequency  $tf(w, r)$ . For each pair  $\langle w, r \rangle$ , we define the percentage of adoption of a category  $c$  associated with the review  $r$  as:

$$\%adp(c, r) = \frac{\sum_{w \in C} tf(w, r)}{\sum_{w \in W} tf(w, r)} \times 100 \quad (1)$$

Finally, to compute the percentage of adoption of a category  $c$  associated with a set of reviews  $R$ , we computed the geometric mean of Eq. 1. Since our data may contain zeros, a constant value  $k$  equal to the minimum adoption excluding zero has been added to each value in the set and later subtracted from the result.

$$\%adp(c) = \left( \prod_{i=1}^{|R|} (\%adp(c, r_i) + k) \right)^{\frac{1}{n}} - k \quad (2)$$

In the above formula,  $|R|$  is the cardinality of the set of reviews  $R$ . We always show results when  $|R| > 1K$  reviews, so to have a percentage error less than 2% with 95% confidence interval [2, 17].

*Metric operating on the dictionary lexicon.* Beside the *adoption* metric defined in Eq. 2, we define another metric called *term frequency gain*, which supports a more fine-grain level of investigation by operating at the lexicon level. Specifically, let  $R_A$  and  $R_B$  be two sets of reviews (e.g., reviews left in two given years and/or cities), let  $r$  be a specific review belonging to  $R_A \cup R_B$ . Let  $W$  be the set of words (unigrams) contained in  $R_A \cup R_B$ . For each word  $w \in W$ , we compute the logarithmically scaled term frequencies  $tf_A(w)$  and  $tf_B(w)$  associated with, respectively,  $R_A$  and  $R_B$ . Finally, we compute the term frequency gain of each word  $w$  as:

$$tf_{gain}^{A/B}(w) = \frac{tf_A(w)}{tf_B(w)} \quad (3)$$

Note that, because each term frequency in Eq. 3 is normalised in  $[0, 1]$ , this metric allows us to detect words that are over-used in  $R_A$  compared to  $R_B$  ( $tf_{gain}^{A/B} > 1$ ), and vice versa ( $tf_{gain}^{A/B} < 1$ ).

## 4.3 Validating our dictionary

To gain confidence in the ability of our dictionary and metrics to genuinely distinguish reviews that semantically belong to different categories, we performed two tests, one using a small set of manually labelled sentences, and one using the whole corpus of 3.2M unlabelled reviews.

*Validation 1 – Labeled reviews.* For the first validation test, we used the 1,868 manually labelled sentences at *step 3* above. For each sentence, we computed its business adoption and social adoption

Categories lev. 1	Categories lev. 2	Categories lev. 3	# Words	Top 5 words (by term frequency)	Intra-cluster similarity
Business	Property	Property type	17	apartment, house, home, flat, private	0.48
		Interiors	43	clean, comfortable, room, bed, kitchen	0.42
		Facilities	17	water, hot, wifi, towels, tv	0.42
	Location	Location	109	quiet, area, walk, located, restaurants	0.40
	Professional Conduct	Communication	33	communication, questions, quick, communicative, responded	0.49
		Logistics	22	check, provided, arrival, late, keys	0.35
		Advice	11	information, recommendations, tips, advice, suggestions	0.61
Social	Social Interaction	Hospitality	35	helpful, welcoming, available, accommodating, responsive	0.37
		People	24	family, friend, husband, wife, daughter	0.44
		Personality	22	friendly, kind, warm, charming, sweet	0.47
		Sharing	6	share, sharing, experiences, stories, interests	0.60
		Talking	8	chat, conversation, talking, chatting, moments	0.55
		Meals	8	breakfast, delicious, fresh, dinner, meals	0.59

Table 3: Summary statistics of the final dictionary

(category level-1) values, using Eq. 2. We then compared these values to the manual classification of such sentences performed by crowd-workers. Table 4 shows the adoption of the business and social categories for the 1,868 manually annotated (ground truth) sentences. Let us consider the adoption of the business category first. As expected, the metric is much higher when computed over the business set than when computed over the social set (20% against 3% – a decrease of -85%). Conversely, the adoption of the social category is substantially higher when computed on the social set of reviews rather than the business set (at 10% compared to 2%, an increase of +400%). This result is preliminary evidence that our dictionary and metrics are able to correctly distinguish the two level 1 categories.

	Business set	Social set
Business adoption	20%	3%
Social adoption	2%	10%

Table 4: Business and social adoption in our corpus of 1,868 manually annotated sentences

Table 4 also shows that the highest adoption of the business category is twice as high as the highest adoption of the social category (when computed over the business and social sets respectively). One may question whether this simply derives from the fact that the business vocabulary used by Airbnb guests (287 specific words concerning the property, location of the property, and professional conduct of the host) is substantially wider than the social one (68 specific terms concerning the social interaction between guest-host). To investigate whether this is indeed the case, we restricted the business and social lexicon to have the same number of words (i.e., we kept in our lexicon only the top  $n$  words according to their term frequency for the business and social categories, with  $n = \{10, 20, 40\}$ ). We found the exact same trend for all  $n$ . We take this as indication that guests' reviews are genuinely more prone to contain more business terms than social ones.

**Validation 2 – Unlabeled reviews.** For the second validation test, we used unlabelled data and, specifically, all the 3.2 million guests' reviews. Airbnb guests can choose to rent 'whole apartments' as well as 'shared/private rooms'. We expect to have more social interactions between host-guest when guests rent 'shared/private rooms', compared to those occurring when guests rent 'entire home/apt'. Therefore, we also expect that reviews associated with

'shared/private rooms' contain more social terms than reviews associated with 'entire home/apt'. To verify whether our dictionary and metrics can capture this intuition, we grouped reviews according to the type of property listed (i.e., 'shared/private rooms' vs. 'entire home/apt') and applied Eq. 2 to each set of reviews.

Table 5 shows the change of adoption of the two level-1 categories in our dictionary, for 'shared/private rooms' relative to the reference class 'entire home/apt'. We observe a slight decrease of adoption for the business category (from -11% in Great Manchester, to -38% in San Francisco) and a boost of adoption for the social category (from +76% in San Francisco, to +209% in Melbourne) when shifting from 'entire home/apt' to 'shared/private rooms'. This finding meets the intuition that reviews written for shared/private rooms discuss less business-related topics and more social-related topics than reviews written for entire apartments. We take this as further confirmation of the reliability of our dictionary.

Relative change of ...	Great Manchester	London	Melbourne
... business adoption	-11%	-28%	-20%
... social adoption	+173%	+107%	+209%

Relative change of...	New York City	San Francisco	Sydney
... business adoption	-19%	-38%	-20%
... social adoption	+106%	+76%	+150%

Table 5: Change of the business and social adoption for the set of reviews associated with 'shared/private rooms', relative to the reference class 'entire home/apt'

We next proceed to illustrate four examples of market research questions that one can perform using our dictionary and metrics.

## 5 THE SOCIAL-BUSINESS DICHOTOMY

We conduct four different investigations that aim to shed light onto the big debate of whether Airbnb is a social interaction vs. business transaction platform. First, we operate at the level of *categories* defined in our dictionary to analyze at different granularities how Airbnb is evolving over a period of 10 years and for 6 different cities (Section 5.1). Second, we zoom in at the level of *lexicon* defined in our dictionary to detect micro-variations in trends, once again over time and space (Section 5.2). By segmenting reviews even further, we then investigate whether the business-social dichotomy varies for different groups of hosts (Section 5.3) and for properties located in different neighbourhoods within the same city (Section 5.4).

## 5.1 The dichotomy over the years

To begin with, we investigate whether Airbnb is evolving as a platform where guests are more concerned with business aspects of the service or with social ones. We perform this analysis by grouping reviews on a per year and per city basis, and by computing the adoption metric defined in Eq. 2. Figure 1 plots the adoption of the ‘business’ and ‘social’ level 1 categories across each year and for each analysed city; the different color shades of the plot show the adoption of the level 3 categories.

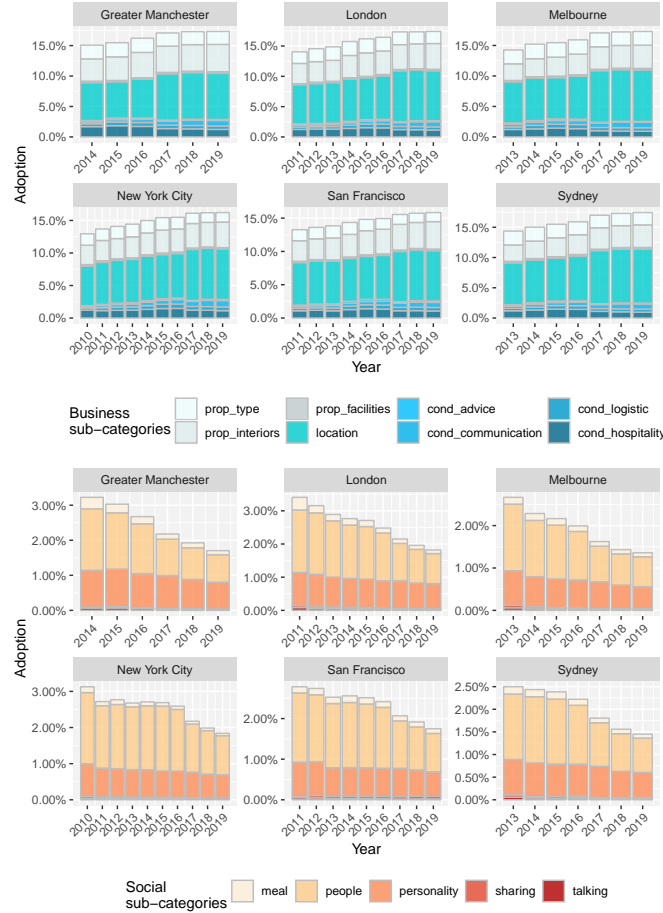


Figure 1: Adoption metric by year and city

Overall, we find that the adoption of the ‘business’ category is increasing over time, while the adoption of the social category is steadily decreasing. For example, in London, the adoption of the business category in 2011 is 14%, whereas it is 17.5% in 2019 – the relative increase is of 25% in a 9 year temporal window, with a growth of 2.8% per year. Conversely, the adoption of the ‘social’ category is 3.5% overall in 2011, whereas it decreases to 1.9% in 2019. This represents a reduction of 45% in a 9 year temporal window, with a relative decrease of 5% per year.

To gain more fine-grained insights, we inspect Figure 1 further, and observe trends within the ‘business’ category. We find that, at level 3 categories, ‘location’, ‘property type’ and ‘interiors’ are

the most frequently discussed ones; as an example, in London they collectively gather 14.5% of adoption in 2019, with a constant growth since 2011. ‘Hospitality’ is the level 3 category with highest adoption within the ‘business conduct’ level 2 (sub)category; as an example, in London it reached its highest adoption level (around 1.5%) in 2015–2016 with a consistent increase from 2011, but either stalled or even slightly decreased afterwards. If we now move our attention to the ‘social’ category, we find that ‘people’ and ‘personality’ are the most frequently discussed level 3 (subsub)categories, with about 1% and 0.5% adoptions in 2019, respectively. However, both of them exhibit a negative slope of adoption rate across all years. The other level 3 categories, namely ‘meal’, ‘sharing’ and ‘talking’, are rarely used in the whole observation period.

Figure 1 also shows that all the identified trends are confirmed for each city, despite the fact they are located in different countries/continents. This suggests that Airbnb language evolution is happening at a global scale, at least in Western countries.

*Sanity checks.* To gain confidence in the results presented above, we performed both a statistical validation of the proposed metrics and an analysis of potential confounding factors. In terms of statistical validation, we built a null (random) model by shuffling the years in our dataset and then repeated the whole analysis on it. Figure 2 shows the adoption metric applied to the (random) null model averaged across all cities, since trends were found to be similar. We observe flat trends throughout. Furthermore, we compared the distribution of adoption of both business and social categories in the years from 2010 to 2012 against their distribution in the years from 2017 to 2019, by running the *Wilcoxon rank sum test*. The obtained *p-value*  $< 2.2 e^{-16}$  confirms the difference in the mean value showed above is statistically significant.

To control for potential confounding factors, we considered both *review length* and *room type*. Intuitively, both of them could be a cause of the observed phenomenon: a recorded prevalence of short reviews in the late years could cause the reduction of the adoption of the ‘social’ category; this is true also if the number of reviews associated with ‘entire apartments’ as opposed to ‘shared/private room’ increases drastically in the last years considered in our investigation. To exclude these confounding factors, we proceeded by binning the reviews in our dataset according to their length and room type. We re-plotted the ‘business’/‘social’ adoption for each year from 2010 to 2019 and for each bin. As an example, Figure 3 illustrates the adoption metric for the ‘business’ and ‘social’ categories across each year and across each value of room type (‘entire home/apt’, ‘private room’, ‘shared room’). Results have been averaged by city since trends were found to be similar. The illustrated trends are consistent with those already shown in Figure 1, suggesting that the findings reported in this section are not a consequence of these confounding factors.

## 5.2 The dichotomy in words

The above analysis suggests that Airbnb is evolving as a platform where guests are more concerned with business aspects, rather than with social aspects, of the hospitality. We can examine more precisely what aspects of the service are behind this trend by operating at the level of the lexicon and computing the term frequency gain metric (Eq. 3). As an example, we binned reviews so to cover a



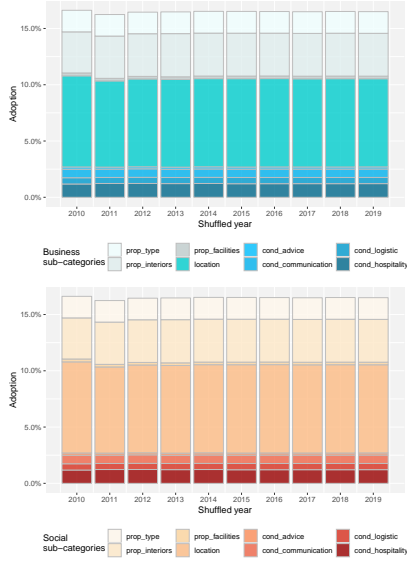


Figure 2: Adoption metric for the null model

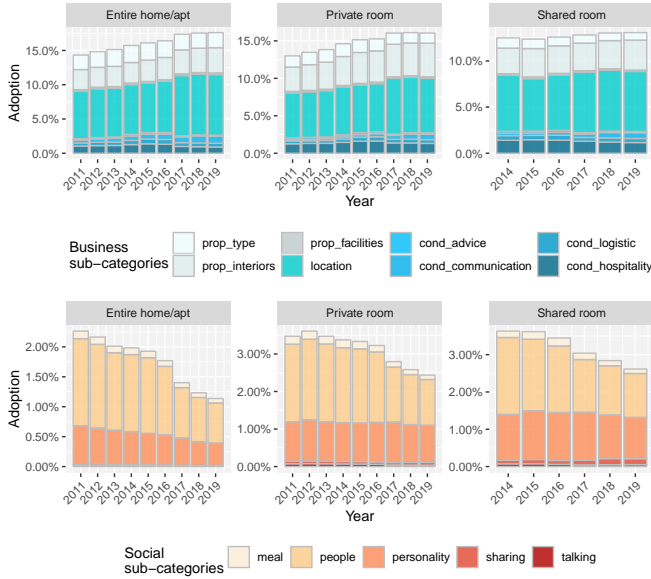
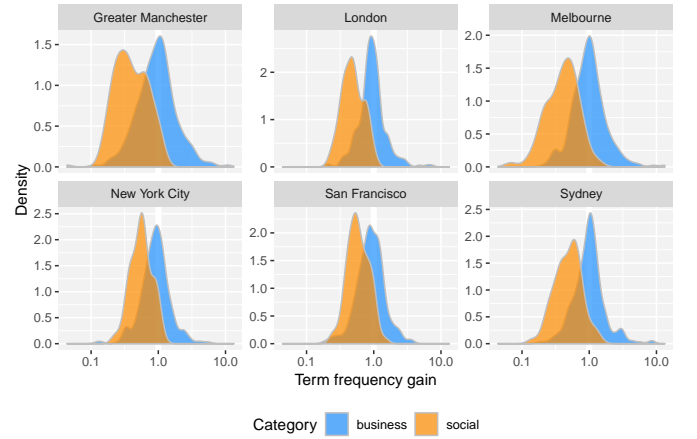


Figure 3: Adoption metric by year and room type

late period (set  $A$ , reviews written between 2017–2019) and an early one (set  $B$ , reviews written between 2010–2012). For each review belonging to the periods above, we consider each word (unigram)  $w$  and compute the *term frequency gain*  $tf_{gain}^{A/B}$  as defined in Eq. 3. This metric allows us to detect words that are over-used in the late period compared to the early one ( $tf_{gain}^{A/B} > 1$ ), and vice versa ( $tf_{gain}^{A/B} < 1$ ). To ease the analysis, we discarded very unpopular words, i.e., words having a total term frequency lower than  $10^{-5}$ .

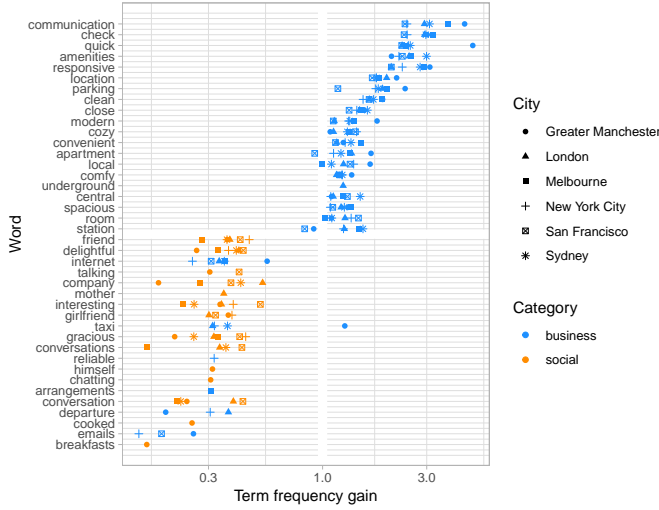
Figure 4: Density distribution of  $tf_{gain}^{A/B}$  of each category in each city

We start this investigation by plotting the density distribution of this metric for the two broadest categories of our dictionary ('business' and 'social'). Figure 4 shows the results for each analysed city and highlights two interesting trends. First, the 'business' category has words associated with both positive and negative term frequency gain. This means that some aspects of the 'business' category are indeed over-emphasised in the late period compared to the early years when Airbnb was a young service; however, there are also 'business' words which experience an opposite trend. Second, the great majority of words that are part of our 'social' lexicon are under-used in the late period compared to the early years when Airbnb was a young service. This is valid in all cities under study.

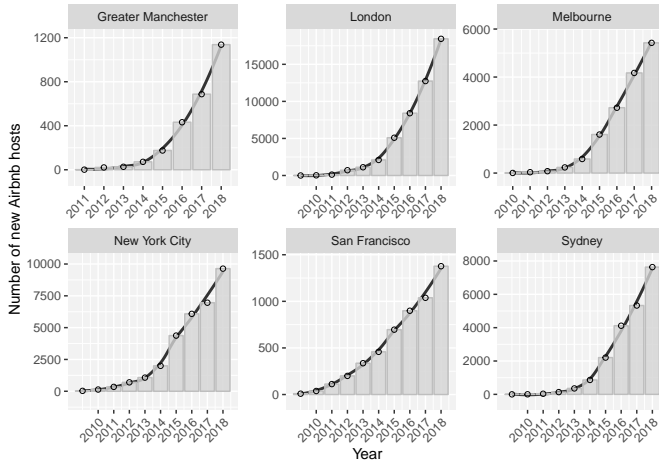
Figure 5 shows the top-20 words in our dictionary that exhibit the strongest decline/increase of term frequency gain (computed as average across all cities). As one would expect based on our findings so far, all the top-20 words belong to the 'business' category; interestingly, we observe that it is the 'location' (sub)category that mostly drives this trend, with words such as 'parking', 'local', and 'central'. Conversely, 70% of the bottom-20 words belong to the 'social' category, and these words span different social (sub)categories, from 'personality' (e.g., 'gracious'), to 'people' (e.g., 'friend'), to talking (e.g., 'delightful', 'company', 'conversations').

### 5.3 The dichotomy across market segments

By segmenting reviews not only by city and year but also by host (i.e., service provider) characteristics, our approach enables investigations across different market segments. As an example, in this section we segment hosts based on the concept of technology adoption [16, 36, 41, 43]. In the literature, users are classified as: *innovators* (first 2.5% of users) adopting a new technology, *early adopters* (subsequent 13.5% of users), *early majority* (the following 34% of users), *late majority* (34% of remaining users), and *laggards* (last 16% of users). To understand the current Airbnb adoption era, we computed the number of new users for each year and for each city involved in our analysis; we plot the corresponding results in Figure 6.



**Figure 5: Words contained in our dictionary having the top/bottom 20 term frequency gain**

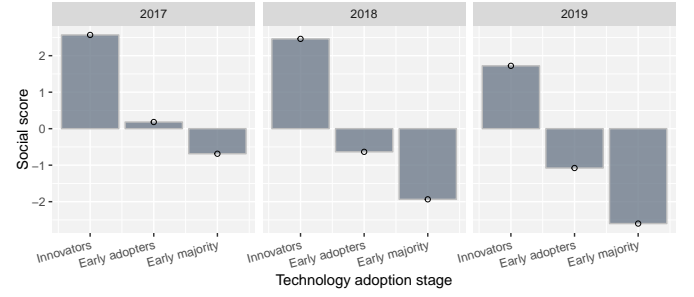


**Figure 6: Adoption of Airbnb in each city**

By comparing the shape of the obtained curves with the Gaussian-shaped one reported in the related literature [36, 41, 43], we conclude that Airbnb has not reached the late majority phase yet, and this result is homogeneous across all the cities in our investigation. Moreover, by observing the adoption rates, we hypothesise that Airbnb is in the middle of the early majority phase. Following this reasoning, we segment hosts in our dataset in: innovators (first 5%), early adopters (subsequent 45%), and early majority (remaining 50%). We then linguistically analyse the reviews that each such bin collected.

For ease of presentation, we focus this analysis on the two level-1 categories only, and present results in a concise way by means of what we call ‘social score’:

$$\text{socialscore} = z(\% \text{adp}(\text{social})) - z(\% \text{adp}(\text{business})), \quad (4)$$



**Figure 7: Social score against Airbnb technology adoption**

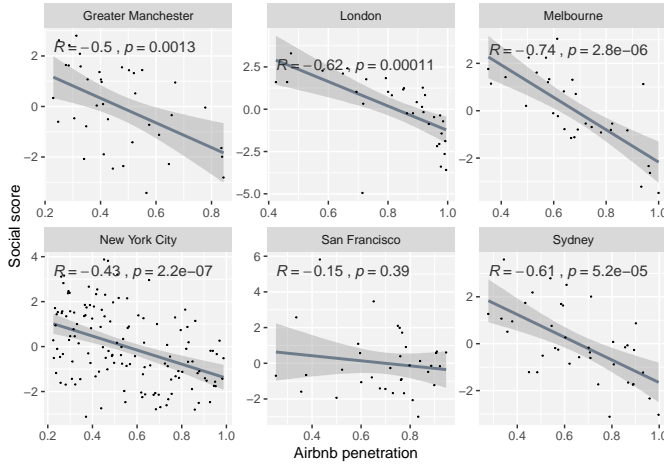
where  $z(\% \text{adp}(\text{social}))$  and  $z(\% \text{adp}(\text{business}))$  are, respectively, the z-scores of the ‘social’ and the ‘business’ adoption on a set of reviews. Note that, since both  $z(\% \text{adp}(\text{social}))$  and  $z(\% \text{adp}(\text{business}))$  are normalised and unitless numbers, Eq. 4 is able to directly compare the ‘social’ and ‘business’ adoption despite their difference of scale, thus telling us whether a particular set of reviews is biased towards the former or the latter category.

Figure 7 shows results averaged across all cities, since trends were found to be similar. We observe that, in each year, innovators receive reviews with a higher social score than the other categories of hosts do. We speculate that innovators may engage more in social interactions (as the original sharing economy manifesto wanted), and thus receive reviews with higher social scores. Figure 7 also indicates that the percentage of social reviews received by innovators is decreasing over time. This result is coherent with a platform adaptation phenomenon, and may suggest that, even though innovators remain overall more social than the other categories of users, they are undergoing an adaptation process following the evolution of Airbnb towards a more business-oriented model. After controlling for the same confounding factors discussed in the previous section, we find that neither review length nor room type affected our results.

#### 5.4 The dichotomy across neighbourhoods

As a final example of market research investigation one may perform, we illustrate how to segment reviews at a finer level of spatial granularity, so to investigate the possible presence of varying platform adoption dynamics within a single city. To illustrate how, we subdivided each city in its electoral districts, which are geographic areas of different size designed to have a similar number of residents. For each such area, we computed two scores: a social score, computed using Eq. 4 over the set of reviews left for Airbnb properties located in such area; and an *Airbnb penetration rate*, computed as the ratio of the number of active Airbnb listings (that is, the number of listings receiving at least one review), over of the maximum number of listings in any given district, so to normalise such a rate between  $[0, 1]$ . The latter has previously been found to be a good proxy for central / tourist areas [35].

Figure 8 shows the scatter plot (along with Pearson Correlation) between the Airbnb penetration rate and the social score for neighbourhoods in each city in our dataset. We observe that neighbourhoods with very high Airbnb adoption rates show lower social scores than those with lower penetration rates (Pearson correlation



**Figure 8: Social score against area Airbnb penetration rate (on a per city basis)**

up to -0.74). Results are valid across all cities considered, excluding San Francisco where the correlation is not statistically significant ( $p$ -value  $> 0.01$ ). These results would suggest that the Airbnb hospitality service is being valued more from a business point of view in central and tourist areas, whereas the social element is to be found once we move towards off-the-beaten track areas.

## 6 DISCUSSION AND CONCLUSION

According to the ‘crossing the chasm’ theory [33], there are five segments of technology adoption: the first three segments ‘innovators’, ‘early adopters’, and ‘early majority’ (which together make the chasm) are followed by ‘late majority’ and ‘laggards’. Based on our results, Airbnb has crossed the chasm by moving its adoption from the innovators and early adopters to the larger market segment of the early majority, and is still expected to grow in this segment.

With the arrival of the early majority, norms have changed though: early adopters engaged more in social interactions than the early majority are currently doing. Given this majority, Airbnb now revolves around business-focused experiences rather than socialisation. That is surprisingly consistent across the various countries our six Western cities are located in. Interestingly, the behaviour of early adopting hosts does not apply to early adopting areas: despite tourist areas being the initial ones to be offered on the platform [35], they are predominantly engaged in business-oriented experiences compared to properties offered in less central areas.

These findings were made possible thanks to a platform (Airbnb) specific dictionary that we constructed, starting from ready available guests’ reviews and using a combination of thematic analysis and machine learning techniques. Using this dictionary and a few metrics we defined, it is now possible to perform quantitative linguistic analysis of Airbnb experiences at scale, and use the findings to inform strategic business decisions that span different directions. For example, (i) *improved guest experience*: Airbnb developers could add functionalities to the platform, such as guest/guest recommender systems, to connect like-minded people based on the topics they discuss in their reviews. Furthermore, rather than going

for a ‘one size fits all’ business model, they may try to leverage the platform diversification (in peer composition and district offerings), and enrich the service offered with new features tailored to a guest’s willingness to socialise by, for example, preferentially ranking hosts among the early adopters or properties in less tourist areas. (ii) *Improved host experience*: Airbnb could offer new hosts information and online training on how to attract guests, for example by recommending they offer services/experiences that guests to that city care the most about; as changes in what guests care about are detected, the platform can offer up-to-date vetting and training to its hosts, so to maintain guests’ satisfaction high. (iii) *Tailored marketing*: by knowing what guests value, Airbnb can create hyper local advertising campaigns, for example highlighting more the efficiency of the service rather than its hospitality, to appeal to certain market segments that can vary by geographic location and over time. (iv) *Data-driven regulation*: by knowing the local market position of Airbnb, authorities and platform owners can co-create policies that differentiate business/leisure travels.

Our dictionary and metrics can also support social science researchers in their investigations. For example, despite analysing cities in different countries and continents, we acknowledge that this work has so far been restricted to the Western world. We thus cannot answer questions of globalisation and platform adaptation that expand beyond it. Future studies should be conducted both in Eastern countries and in developing ones, both of which have largely been neglected by the current sharing economy literature, partly because of a lack of scalable analytical tools [9]. As we do so, our dictionary may have to be amended, or new ones may have to be developed, so to properly analyse Airbnb reviews in a new geographic context and/or in a different language. One can automatically assess whether our dictionary is valid (e.g., in a new city, at a future point in time, in a different - translated - language) by quantifying the proportion of words modelled by our dictionary, with the expectation that the same dictionary can be used for a number of years within cities sharing similar cultural traits. When a new custom dictionary needs to be built, the very same inductive approach proposed in this paper can be reproduced. This process offers great scalability advantages compared to the traditional market analysis approaches based on interviews, with the expectation that the effort to employ (thousands of) crowd-workers to conduct a new thematic analysis to be significantly less than that to recruit and interview (tens of) Airbnb users (with domain experts needed in both cases).

In addition to replicating our analysis, new lines of investigation can be pursued, delving deeper into peer segmentation, for example to explore questions of gender-specific and age-specific values in the Airbnb hospitality service. Last but not least, using a similar approach to the one proposed in this paper, one could go beyond the business-social dichotomy and develop dictionaries that enable orthogonal explorations, for example on the theme of trust, not least because trust is one of the main currency in the sharing economy. Nowadays, given Airbnb’s focus on business at the price of socialisation, cultivating trust might not be a priority and, as such, trust deficits might stand in the way of growth – the very same growth that could help Airbnb decisively move well beyond the chasm.

## REFERENCES

- [1] Abdulkareem Alsudais and Timm Teubner. 2019. Large-Scale Sentiment Analysis on Airbnb Reviews from 15 Cities. (2019).
- [2] Douglas Altman, David Machin, Trevor Bryant, and Martin Gardner. 2013. *Statistics with confidence: confidence intervals and statistical guidelines*. John Wiley & Sons.
- [3] Amita Inah Marie Bancoro and Adarsh Batra. 2018. Tourists's Motivation in Using Sharing Accommodation in Bangkok. In *Global Conference on Business, Hospitality, and Tourism Research (GLOSEARCH 2018)*.
- [4] Lajos Boros, Gábor Dudás, Tamás Kovalcsik, Sándor Papp, György Vida, et al. 2018. Airbnb in Budapest: analysing spatial patterns and room rates of hotels and peer-to-peer accommodations. *GeoJournal of Tourism and Geosites* 21, 1 (2018), 26–38.
- [5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [6] Judith Bridges and Camilla Vásquez. 2018. If nearly all Airbnb reviews are positive, does that make them meaningless? *Current Issues in Tourism* 21, 18 (2018), 2057–2075.
- [7] Mingming Cheng and Xin Jin. 2019. What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management* 76 (2019), 58–70.
- [8] Ha-Nhat Cuong, Van-Dang Tran, Linh Ngo Van, and Khoat Than. 2019. Eliminating overfitting of probabilistic topic models on short and noisy text: The role of dropout. *International Journal of Approximate Reasoning* (2019).
- [9] Tawanna R. Dillahun, Xinyi Wang, Earnest Wheeler, Hao Fei Cheng, Brent Hecht, and Haiyi Zhu. 2017. The Sharing Economy in Computing: A Systematic Literature Review. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 38 (Dec. 2017), 26 pages.
- [10] Bin Fang, Qiang Ye, and Rob Law. 2016. Effect of sharing economy on tourism industry employment. *Annals of Tourism Research* 57 (2016), 264–267.
- [11] Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin* 76, 5 (1971), 378.
- [12] Andrey Fradkin, Elena Grewal, and David Holtz. 2017. *The determinants of online review informativeness: Evidence from field experiments on Airbnb*. Technical Report. Working Paper.
- [13] Andrey Fradkin, Elena Grewal, Dave Holtz, and Matthew Pearson. 2015. Bias and reciprocity in online reviews: Evidence from field experiments on airbnb. In *Proc. of EC '15*. ACM, 641–641.
- [14] Yoav Goldberg and Omer Levy. 2014. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722* (2014).
- [15] Daniel Guttentag, Stephen Smith, Luke Potwarka, and Mark Havitz. 2018. Why tourists choose Airbnb: A motivation-based segmentation study. *Journal of Travel Research* 57, 3 (2018), 342–359.
- [16] Jerald Hage. 1980. *Theories of organizations: Form, process, and transformation*. Wiley New York.
- [17] Jason S Haukoos and Roger J Lewis. 2005. Advanced statistics: bootstrapping confidence intervals for statistics with "difficult" distributions. *Academic emergency medicine* 12, 4 (2005), 360–365.
- [18] Cindy Yoonjung Heo, Inès Blal, and Miju Choi. 2019. What is happening in Paris? Airbnb, hotels, and the Parisian market: a case study. *Tourism Management* 70 (2019), 78–88.
- [19] Tapio Ikkala and Airi Lampinen. 2015. Monetizing network hospitality: Hospitality and sociability in the context of Airbnb. In *Proc. of CSCW '15*. ACM, 1033–1044.
- [20] Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. 2019. Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey. *Multimedia Tools and Applications* 78, 11 (2019), 15169–15211.
- [21] George Joseph and Vinu Varghese. 2019. Analyzing Airbnb Customer Experience Feedback Using Text Mining. In *Big Data and Innovation in Tourism, Travel, and Hospitality*. Springer, 147–162.
- [22] Jiwon Jung, Susik Yoon, SeungHyun Kim, Sangkeun Park, Kun-Pyo Lee, and Uichin Lee. 2016. Social or Financial Goals?: Comparative Analysis of User Behaviors in Couchsurfing and Airbnb. In *Proc. of CHI '16*. ACM, 2857–2863.
- [23] David J Ketchen and Christopher L Shook. 1996. The application of cluster analysis in strategic management research: an analysis and critique. *Strategic management journal* 17, 6 (1996), 441–458.
- [24] Maximilian Klein, Jinhao Zhao, Jiajun Ni, Isaac Johnson, Benjamin Mako Hill, and Haiyi Zhu. 2017. Quality Standards, Service Orientation, and Power in Airbnb and Couchsurfing. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 58.
- [25] Airi Lampinen and Coye Cheshire. 2016. Hosting via Airbnb: Motivations and financial assurances in monetized network hospitality. In *Proc. of CHI '16*. ACM, 1669–1680.
- [26] Abdelaziz Lawani, Michael Michael R Reed, Tyler Mark, and Yuqing Zheng. 2018. Reviews and price on online platforms: Evidence from sentiment analysis of Airbnb reviews in Boston. *Regional Science and Urban Economics* (2018).
- [27] Carmen Kar Hang Lee, Ying Kei Tse, Minhao Zhang, and Jie Ma. 2019. Analysing online reviews to investigate customer behaviour in the sharing economy. *Information Technology & People* (2019).
- [28] Pearl MC Lin, Daisy XF Fan, Hanqin Qiu Zhang, and Chloe Lau. 2019. Spend less and experience more: Understanding tourists's social contact in the Airbnb context. *International Journal of Hospitality Management* 83 (2019), 65–73.
- [29] Jianwei Liu, Karen Xie, Qiang Ye, and Dong Jing. 2018. The Impact of Sharing Economy on Local Employment: Evidence from Airbnb. (2018).
- [30] Yi Luo. 2018. What Airbnb Reviews can Tell us? An Advanced Latent Aspect Rating Analysis Approach. (2018).
- [31] Yi Luo and Rebecca Liang Tang. 2019. Understanding hidden dimensions in textual reviews on Airbnb: An application of modified latent aspect rating analysis (LARA). *International Journal of Hospitality Management* 80 (2019), 144–154.
- [32] Eva Martín-Fuentes, Cesar Fernandez, Carles Mateu, and Estela Marine-Roig. 2018. Modelling a grading scheme for peer-to-peer accommodation: Stars for Airbnb. *International Journal of Hospitality Management* 69 (2018), 75–83.
- [33] Geoffrey A Moore and Regis McKenna. 1999. Crossing the chasm. (1999).
- [34] James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates* 71, 2001 (2001), 2001.
- [35] Giovanni Quattrone, Davide Proserpio, Daniele Quercia, Licia Capra, and Mirco Musolesi. 2016. Who Benefits from the Sharing Economy of Airbnb?. In *Proc. of WWW '16*. 1385–1394.
- [36] Everett M Rogers. 2010. *Diffusion of innovations*. Simon and Schuster.
- [37] Paolo Roma, Umberto Panniello, and Giovanna Lo Nigro. 2019. Sharing economy and incumbents' pricing strategy: The impact of Airbnb on the hospitality industry. *International Journal of Production Economics* 214 (2019), 17–29.
- [38] Sampo Satama et al. 2014. Consumer adoption of access-based consumption services-Case AirBnB. (2014).
- [39] Zahratu Shabrina, Elsa Arcaute, and Michael Batty. 2019. Airbnb's disruption of the housing structure in London. *arXiv preprint arXiv:1903.11205* (2019).
- [40] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.
- [41] Andrew H Van de Ven, Harold L Angle, and Marshall Scott Poole. 2000. *Research on the management of innovation: The Minnesota studies*. Oxford University Press on Demand.
- [42] David Wachsmuth and Alexander Weisler. 2018. Airbnb and the rent gap: Gentrification through the sharing economy. *Environment and Planning A: Economy and Space* 50, 6 (2018), 1147–1170.
- [43] Gerald Zaltman, Robert Duncan, and Jonny Holbek. 1973. *Innovations and organizations*. John Wiley & Sons.
- [44] Georgios Zervas, Davide Proserpio, and John Byers. 2015. A first look at online reputation on Airbnb, where every stay is above average. (2015).
- [45] Georgios Zervas, Davide Proserpio, and John Byers. 2016. The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. *Boston U. School of Management Research Paper* 2013-16 (2016).
- [46] Georgios Zervas, Davide Proserpio, and John W Byers. 2015. The impact of the sharing economy on the hotel industry: Evidence from airbnb's entry into the texas market. In *Proc. of EC '15*. ACM, 637–637.