

Reliability of Reviews on the Internet: The Case of TripAdvisor

Alton Y. K. Chua and Snehasish Banerjee

Abstract—Given the rampant growth of travel-related user-generated content on the Internet, this paper seeks to investigate the reliability of reviews in TripAdvisor, a popular travel review site. To support the goal, two objectives are submitted. The first is to locate clusters of highly-interlocked hotels that have been evaluated by a common pool of reviewers. This enables review baselines to be established, which in turn facilitates detection of anomalies. The second is to determine the inter-reviewer and intra-reviewer reliability of reviews. Results suggest that reviews in TripAdvisor could be largely reliable. The findings gleaned from the results are discussed.

Index Terms—user-generated content, online reviews, reliability, social network analysis

I. INTRODUCTION

THE participatory nature of the Internet in recent years has led to an explosive growth of travel-related user-generated content. Consequently, the tourism industry has morphed into a social force [1] where users' travel plans such as where to stay and what to do are largely shaped by the collective experiences and opinions of others.

Among various travel-related sites that support user generated content, TripAdvisor stands out most prominently in terms of usage and content [2]. Since its launch in 2000, it has garnered more than some 75 million reviews [3]. With such dime a dozen opinions, it is no wonder that TripAdvisor is recognized as an important information source among users for travel planning [4].

However, since reviews on TripAdvisor are not authenticated, they may not always be reliable. Some reviews could be biased and misleading while others could be frivolous and incomprehensible [5]. Despite community-policing measures such as the provision to flag dubious reviews, reliability on TripAdvisor remains a nagging question.

Hence, the overall goal of this paper is to study the reliability of user-generated content on the Internet, and in particular, reviews in TripAdvisor. To support this goal, two objectives are submitted. The first is to locate clusters of highly-interlocked hotels that have been reviewed by a common pool of users (henceforth, reviewers). Relying on

the concept of 'wisdom of crowds', the identification of such clusters enables review baselines to be established, which in turn facilitates detection of anomalies. The second objective is to determine the reliability of reviews on two levels, namely inter-reviewer reliability and intra-reviewer reliability. The former refers to the extent to which star ratings given by reviewers in clusters of highly-interlocked hotels are comparable to those attracted by a given hotel in that cluster. The latter on the other hand refers to the extent to which the star rating given by a reviewer is consistent with the textual comments provided.

This paper is significant on two counts. One, TripAdvisor has been studied from various perspectives including the ways its website features and reviews engender personal and systems trust [6] and how its reviews have been used in different stages of travel planning [2]. Yet, few works explicitly examine whether reviews in TripAdvisor are reliable in the first place. This paper thus focuses on a pertinent issue that has not been widely explored. Two, scholarly interest in establishing reliability of reviews is burgeoning. The common thread of investigation used hitherto has been either review-centric (e.g. [7]) or reviewer-centric (e.g. [8]). This paper not only combines both these approaches, but also uses social network analysis (SNA) as part of its methodology. It can therefore serve as a call for the development of more comprehensive approaches to study reliability of user-generated content.

II. LITERATURE REVIEW

A. User-Generated Content in Tourism

As the participatory nature of the Internet permits easy contribution of user-generated content without any editorial control, sites such as TripAdvisor that allow users to share post-trip travel experiences have acquired immense popularity [9]. These sites serve as avenues where users can post reviews to delineate, relive and reconstruct their experiences in hotels to online communities beyond any spatial or temporal constraints. Since potential travelers can browse through experiences of others, it is no wonder that more users are tapping into reviews prior to travel planning [10].

Given the importance of reviews in the context of tourism, numerous studies have been conducted to examine their impact on users' travel plans. For instance, in a study conducted to analyze the impact of reviews on the choice of holiday accommodation, it was found that more than 80% of users purchased accommodation recommended by reviews [11]. In another similar study, users were found to possess

Manuscript received July 01, 2013; revised July 23, 2013.

Alton Y. K. Chua is with Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore 637718 (phone: +65-6790-5810; e-mail: altonchua@ntu.edu.sg).

Snehasish Banerjee is with Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore 637718 (e-mail: snehasis002@e.ntu.edu.sg).

proclivity in trusting reviews even though they were mostly contributed by anonymous reviewers [12]. It appears that users are generally willing to trust reviews and embrace vulnerabilities based on favorable expectations from their online peers [13]. However, little scholarly efforts have been trained on studying the reliability of reviews in the first place, a lacuna that this paper attempts to fill.

B. Reliability of Reviews

Reliability of reviews in sites such as TripAdvisor remains a nagging question due to three reasons. One, unlike sites such as eBay or Expedia that allow users who have completed actual transactions to write reviews, any user can post reviews for any hotels in TripAdvisor. While some reviews could be real accounts of users' post-trip experiences, others could be maliciously contributed with business interests to promote or hurt the reputation of hotels [14]. Two, different reviewers have varying motivations to contribute reviews. While some hold altruistic commitments to write high quality reviews, others post frivolous or misleading reviews out of mischief [15]. Three, different users could have varying levels of expectation from the same hotel. When expectations differ, a hotel appreciated by a user could turn out abysmal for another [16].

For the purpose of this paper, reliability of reviews is established through two approaches. First, clusters of highly-interlocked hotels that have been evaluated by a common pool of reviewers are located. Identification of such clusters enables establishing review baselines by spotting users who could be potentially like-minded with one another in their review patterns, which in turn facilitates detection of anomalies. After all, one of the robust approaches users could follow to assess reliability of reviews in TripAdvisor lies in spotting reviewers who have commented on hotels where the user had already stayed to form a conjecture if their expectations can likewise be met [16].

Second, the reliability of reviews is further assessed on two levels, namely inter-reviewer reliability and intra-reviewer reliability. Inter-reviewer reliability is a measure of the extent to which star ratings given by reviewers in clusters of highly-interlocked hotels are comparable to those attracted by a given hotel in that cluster. Consistency in such rating patterns could suggest that the reviews are largely homogeneous and hence, reliable. Intra-reviewer reliability on the other hand refers to the extent to which the star rating given by a reviewer is consistent with the textual comments provided. This is particularly important because instead of being satisfied with star ratings, most users tend to look into the textual comments before making travel plans [17], [18]. Conceivably, if the rating and the comment of a given review are not consistent with each other, users may deem that review unreliable.

III. METHODOLOGY

A. Data Collection

As indicated earlier, data for this paper were drawn from TripAdvisor, one of the most popular sources of reviews for

hotels. Since its inception in 2000, it has garnered more than some 75 million reviews on about a quarter million hotels spread across 85,000 destinations worldwide and attracts around 40 million unique visitors per month on an average [3], [7], [19].

For the purpose of this paper, all hotels in Singapore that were featured in TripAdvisor as of August, 2011 were chosen for analysis. A web scraper was used to collect reviews posted from January, 2004 to December, 2010 for the hotels. The data items retrieved for the hotels include name, average reviewer rating and star rating. For all reviewers who evaluated the hotels, the data items retrieved include reviewer ID, number of reviews, hotels that they have reviewed, rating for the current hotel, textual comment and date of review. Reviews that were not in English were eliminated. The final dataset admitted for analysis included a total of 249 hotels in Singapore that cumulatively attracted 19,691 reviews from 17,021 unique users.

B. Data Analysis

Data analysis for this paper not only combines review-centric and reviewer-centric approaches, but uses SNA as part of its methodology. It was a three-step process. In the first step, SNA was used to find highly-interlocked hotels that have been evaluated by a common pool of reviewers. For this purpose, the 249 hotels x 249 hotels adjacency matrix was fed into UCINET. When a reviewer of one hotel submits an entry for another hotel, an interlock is said to be formed between the two hotels under consideration [20]. On the other hand, a hotel remains isolated if it does not share reviews contributed by at least one reviewer who has also evaluated another hotel. For the purpose of this paper, a hotel is said to be highly-interlocked if it has five or more interlocks.

In the second step, a reviewer-centric approach will be used to determine inter-reviewer reliability. Reviews posted by the top 100 reviewers in terms of volume of entries contributed for the highly-interlocked hotels will be used for further analysis. Specifically, a quantitative approach will be used to examine the extent to which ratings given by reviewers in clusters of highly-interlocked hotels are comparable to those attracted by a given hotel in the cluster.

In the third step, a review-centric approach will be used to study intra-reviewer reliability. Specifically, extreme positive reviews (those with five-star reviewer rating) and extreme negative reviews (those with one-star reviewer rating) contributed by the top 100 reviewers for the highly-interlocked hotels would be analyzed qualitatively to disinter the consistency between ratings and textual comments.

IV. RESULTS

A. Descriptive Statistics

Among the 249 hotels in Singapore, two-star hotels were most abundant (113). They were followed by four-star hotels (67), three-star hotels (43), five-star hotels (23) and one-star hotels (3).

In terms of reviewer ratings, five-star hotels were rated

most favorably (4.26). They were followed by four-star hotels (3.69), one-star hotels (3.16), two-star hotels (2.92) and three-star hotels (2.65). Reviewers contributed 1.16 reviews on an average. The lowest number of reviews contributed by a reviewer was 1 while the highest was 16.

B. Results of SNA

The results of SNA indicated 58 isolated hotels that were not evaluated by any reviewer who had also commented on some other hotel in Singapore. These included 39 two-star hotels, 17 three-star hotels and two four-star hotels. The remaining 191 hotels (249 - 58) were interlocked cumulatively accounting for 4,192 interlocks with one another.

The highly-interlocked hotels with five or more interlocks were used for further analysis. There were 46 such hotels accounting for 310 interlocks with one another. The hotels included three three-star hotels, 23 four-star hotels and 20 five-star hotels. Fig. 1 shows the interlocking patterns of the 46 hotels. The thickness of an edge between two hotels denotes the pool of common reviewers for both hotels. The highest number of interlocks was found to be 24 between Pan Pacific Singapore and Crowne Plaza Hotel Changi Airport.

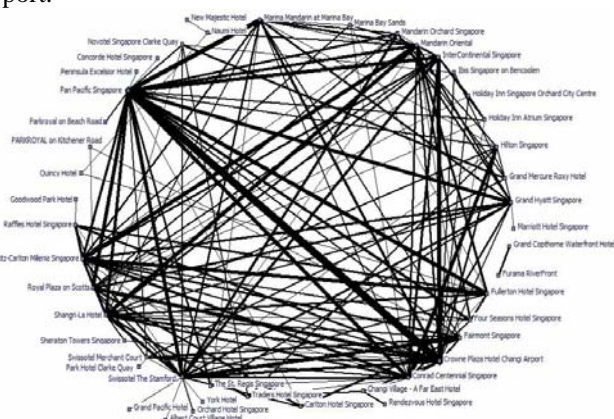


Fig. 1. Interlocking patterns of the 46 hotels with more than five connections.

C. Reliability of users

Reviews posted by the top 100 reviewers in terms of volume of entries for the selected 46 interlocked hotels were used for further investigation. The top 100 reviewers contributed 424 reviews with an average of 8.83 reviews per hotel.

The minimum number of reviews contributed by these reviewers for the 46 hotels was 1 while the maximum number was 14. Specifically, the three three-star hotels attracted 13 reviews with an average reviewer rating of 3.50, the 23 four-star hotels attracted 152 reviews with an average of 3.63, and the 20 five-star hotels attracted 259 reviews with an average of 4.05.

To study inter-reviewer reliability, the reviewer's rating for a given hotel was compared with the average reviewer rating for the hotel. For the purpose of this paper, rating difference (RD) of a review posted by a reviewer for a given hotel is defined as the difference between the reviewer's rating and the hotel's average reviewer rating. Put differently, RD would be positive for users who were overly impressed with a given hotel vis-à-vis other users, but it

would be negative for users who were mostly dissatisfied vis-à-vis others.

If the absolute value of RD for a given reviewer is bounded within 1, such reviews can be considered as fairly reliable. The dataset revealed that 84 of the 100 reviewers contributed reviews with a mean RD of less than one. This suggests that reviews in TripAdvisor largely exhibited inter-reviewer reliability. Of the remaining 16 users, two had a positive mean RD while 14 had a negative mean RD. It appears that though the 46 hotels were top hotels, they often failed to live up to users' expectations. This in turn raises concerns about the reliability of reviews posted by the two users with positive RD. On delving deeper, it was found that both the users reviewed only five-star hotels. Perhaps, they were familiar only with five-star hotels and generally provided lenient ratings. Hence, their ratings might not be objective insofar as appraising services in five-star hotels.

To study intra-reviewer reliability, a qualitative analysis was conducted on the extreme positive and negative reviews contributed by the top 100 reviewers to determine if ratings were consistent with textual comments. In particular, there were 126 extreme positive reviews and 12 extreme negative reviews.

Among the extreme positive reviews, 103 were consistent and discussed only about the merits of hotels. For example, a user commented about a five-star hotel, "...Cannot fault it from the moment we stepped out of our taxi...Stay there at least once in your life - you won't regret it." Another user posted about a four-star hotel, "Stayed in the executive floor, very good service, excellent breakfast,...excellent swimming pool area... no problem getting taxis either." Among the extreme negative review, all were consistently fraught with lambasting comments, often supplemented with sarcasm. For example, a user commented about a four-star hotel, "I hereby crown the newly elected 'worst hotel'... Not only will I NEVER return to this property, I am going to do mankind a favour by telling all my fellow travellers to stay away from this place...running this junk as a 'business' is a crime." Yet another user indicated, "...When we saw our room my heart sank. TINY! ... What was I paying for???... How does this HOTEL earn 4 stars...???" This suggests that reviews in TripAdvisor largely exhibit intra-reviewer reliability.

However, there were 23 extreme positive reviews that were inconsistent. Even though those reviews indicated a five-star reviewer rating, the textual comments contained noticeable complaints. For example, a user commented about a five-star hotel, "...Service at this hotel was great from start to finish...The only complaints I'd note were ..." Yet another user indicated, "I am a regular at this hotel almost every time I visit Singapore... It's a great hotel with pretty much everything....However, this last trip..." Such reviews could be ambiguous and may not be reliable.

Furthermore, it was found that all extreme negative reviews were lengthy and criticized hotels comprehensively. On the other hand, numerous positive extreme reviews were terse. For example, a user commented about a five-star hotel, "Grt place with luxury away from crowd". Another user mentioned, "From arrival to departure all was fantastic, all staff very helpful and friendly." Such succinct negative extreme reviews were however inconspicuous.

V. DISCUSSION

Three major findings could be gleaned from the results. One, reviews in TripAdvisor fared well in terms of both inter-reviewer and intra-reviewer reliability. This is in line with prior studies such as [7] which also suggested that the vast majority of reviews in TripAdvisor could be reliable. Of late however, there have been a few instances of hotels either asking workers to post fake reviews in TripAdvisor [21] or incentivizing travelers to write glowing comments in the site by offering discounts [22]. In this context, it is important to note that the dataset for this paper included reviews contributed from 2004 to 2010. At a time when social media was still growing, it seems that users have been contributing largely reliable reviews. To maintain the integrity of such reliable corpora of reviews in sites like TripAdvisor, it is important to thwart growing instances of posting misleading reviews.

Two, there were a few instances when the reliability of reviews was called into question. For example, there were 23 reviews with five-star ratings that highlighted demerits of hotels. Such reviews could be ambiguous and unreliable for users. Furthermore, there were six instances in the dataset where a reviewer was found to evaluate the same hotel twice. For example, a reviewer contributed two five-star reviews for a five-star hotel in 2009 and 2010. The first read *"This hotel is absolutely amazing...Highly recommended!!"* while the second indicated *"Every time I go...I find it even better than the time before..."* However, for a four-star hotel, the review posted by a reviewer in 2008 contradicted the one contributed in 2010. While the former criticized the hotel and indicated *"This hotel was a disaster..."*, the latter applauded stating *"The hotel has an amazing..."* There was no mention about his previous comment in the second review. Such dubious cases suggest that even though reviews in TripAdvisor could be largely reliable, they should be taken with a pinch of salt [21], [22].

Three, no one-star or two-star hotels in the dataset emerged as highly-interlocked. Greater interlocks among top hotels imply that such hotels tend to attract a common pool of users who are largely mobile, travel frequently and do not stick to the same hotel across multiple visits. On the other hand, low-tiered hotels attract users who either travel less or stay in the same hotel over multiple visits. Similar to the findings in [7], the volume of reviews attracted by top hotels outnumbered that by low-tiered hotels. Hence, it is no wonder that such top hotels would have more common reviewers among one another, and hence, greater interlocks. This also implies that prolific reviewers who contribute numerous reviews generally frequent top hotels. This justifies why most reviews in TripAdvisor are contributed for top hotels by reviewers who are highly educated with relatively higher levels of income [7], [23], [24]. It is to be acknowledged however that the lack of interlocks among low-tiered hotels did not facilitate a fair analysis of the reliability of reviews contributed for such hotels. This remains a limitation of the paper that future research should address.

VI. CONCLUSION

Given that user-generated on the Internet may not always be reliable, this paper investigated the reliability of reviews in TripAdvisor, a popular review site for sharing travel-related information. First, using SNA, clusters of highly-interlocked hotels that have been evaluated by a common pool of reviewers were located. On establishing review baselines, the inter-reviewer and intra-reviewer reliability of contributions was studied using reviewer-centric and review-centric approaches respectively. Results suggest that reviews in TripAdvisor could be largely reliable.

The contribution of this paper is two-fold. One, it offers insights into the pertinent issue on reliability of reviews, an area of research that has not been adequately explored. The results suggest that reviews in TripAdvisor could be fairly reliable. However, there are also cases where reliability of reviews could be questionable. This suggests that review sites such as TripAdvisor should employ more stringent gate-keeping procedures to render more reliability to the site. Users may also utilize the findings of this paper to conjecture which reviews are likely to be reliable for making travel plans. Two, to the best of our knowledge, this paper represents one of the earliest attempts to amalgamate SNA, reviewer-centric and review-centric approaches to analyze reliability of reviews. By serving as a call for the development of more comprehensive approaches to study reliability of user-generated content, this paper represents a modest attempt to expand the boundaries of extant research methodologies.

However, there are some limitations inherent in the paper that future research needs to address. For one, the dataset used for analysis comprised hotels only within Singapore. Future research can include a more comprehensive dataset to offer better generalizability of findings. Two, the results are constrained by the window of data collection period. Interlock analyses on reviews contributed from January, 2004 to December, 2010 failed to capture the richness and complexity of temporal dynamics. It was possible for two hotels to be interlocked even though a reviewer had contributed review for one hotel in 2004 and for the other in 2010. Future research can include a more granular analysis based on such temporal variations. Three, data were collected during the period when social media was still growing. Perhaps, it was a time when malpractices such as posting of irrelevant and misleading reviews were not too rampant, which in turn, painted a reliable picture of reviews in TripAdvisor. Similar scholarly inquiry using more recent dataset is needed to analyze longitudinally the reliability patterns of such reviews. Four, given that this paper was primarily based on scraped data, future research can also look to develop a system that will allow real users of the system to rate and comment on the reliability of reviews.

ACKNOWLEDGMENT

The authors would like to thank Cayco Nadia David, Saravanan Kiran Kumar, and Kanyakumari Ponnappan Laleetha for their help in data collection.

REFERENCES

- [1] F. Higgins-Desbiolles, "More than an "industry": The forgotten power of tourism as a social force," *Tourism Management*, vol. 27, no. 6, pp. 1192-1208, Dec. 2006.
- [2] U. Gretzel and K. Yoo, "Use and impact of online travel reviews," in *ICT in Tourism 2008*, P. O'Connor, W. Hopken and U. Gretzel (Eds.), pp. 35-46, Springer.
- [3] TripAdvisor. (2013). Fact Sheet [Online]. Available: http://www.tripadvisor.com/PressCenter-c4-Fact_Sheet.html
- [4] J. Y. Chung and D. Buhalis, "Web 2.0: A study of online travel community," in *ICT in Tourism 2008*, P. O'Connor, W. Hopken and U. Gretzel (Eds.), pp. 70-81, Springer.
- [5] P. M. O'Mahony and B. Smyth, "Learning to recommend helpful hotel reviews," in *Proc. ACM Conf. RecSys*, pp. 305-308.
- [6] I. Jeacle and C. Carter, "In TripAdvisor we trust: Rankings, calculative regimes and abstract systems," *Accounting, Organizations and Society*, vol. 36, no. 4-5, pp. 293-309, May-Jul., 2011.
- [7] P. O'Connor, "User-generated content and travel: A case study on TripAdvisor.Com," in *ICT in Tourism 2008*, P. O'Connor, W. Hopken and U. Gretzel (Eds.), pp. 47-58, Springer.
- [8] E-P. Lim, V-N. Nguyen, N. Jindal, B. Liu and H. W. Lauw, "Detecting product review spammers using rating behaviors," in *Proc. ACM CIKM 2010*, pp. 939-948.
- [9] I. S. Lo, B. McKercher, A. Lo, C. Cheung and R. Law, "Tourism and online photography," *Tourism Management*, vol. 32, no. 4, pp. 725-731, Aug. 2011.
- [10] B. Pan and D. R. Fesenmaier, "Online information search: Vacation planning process," *Annals of Tourism Research*, vol. 33, no. 3, pp. 809-832, Jul. 2006.
- [11] K. L. Sidali, H. Schulze and A. Spiller, "The impact of online reviews on the choice of holiday accommodations," in *ICT in Tourism 2009*, W. Höpken, U. Gretzel and R. Law (Eds.), pp. 87-98, Springer.
- [12] T. Hennig-Thurau, "Word-of-mouse: Why consumers listen to each other on the internet," *Yearbook of Marketing and Consumer Research*, vol. 3, no. 3, pp. 37-58, 2005.
- [13] D. M. Rousseau, S. B. Sitkin, R. S. Burt and C. Camerer, "Not so different after all: A cross-discipline view of trust," *Academy of Management Review*, vol. 23, no. 3, pp. 393-404, Jul. 1998.
- [14] N. Jindal and B. Liu, "Opinion spam and analysis," in *Proc. ACM WSDM 2008*, pp. 219-230.
- [15] T. Hennig-Thurau, K. P. Gwinner, G. Walsh and D. D. Gremler, "Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet?," *Journal of Interactive Marketing*, vol. 18, no. 1, pp. 38-52, 2004.
- [16] N. Keates. (2007, June 1). Deconstructing TripAdvisor [Online]. *The Wall Street Journal*. Available: <http://online.wsj.com/article/SB118065569116920710.html>
- [17] P. Chatterjee, "Online reviews: do consumers use them?," in *Proc. Advances in Consumer Research 2001*, pp. 129-133.
- [18] J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: online book reviews," *Journal of Marketing Research*, vol. 43, no. 3, pp. 345-354, Aug. 2006.
- [19] TripAdvisor. (2011). About TripAdvisor media group [Online]. Available: <http://www.tripadvisor.com>
- [20] U. Brandes, P. Kenis, J. Lerner and D. V. Raaij, "Network analysis of collaboration structure in Wikipedia," in *Proc. WWW 2009*, pp. 731-740.
- [21] E. Allen. (2012, October 9). 'Dear staff. We could do with some positive comments': Hotel boss is caught telling his workers to post fake reviews on TripAdvisor [Online]. *MailOnline*. Available: <http://www.dailymail.co.uk/news/article-2214974/Hotel-boss-caught-telling-workers-post-fake-reviews-TripAdvisor.html>
- [22] C. Fernandez. (2011, July 11). TripAdvisor bribes: Hotel owners offer free rooms in return for glowing reviews [Online]. *MailOnline*. Available: <http://www.dailymail.co.uk/travel/article-2013391/Trip-advisor-Hotel-owners-bribe-guests-return-good-reviews.html>
- [23] H. Ogut and B. Tas, "The influence of internet customer reviews on the online sales and prices in hotel industry," *The Service Industries Journal*, vol. 32, no. 2, pp. 1979-214, Jan. 2011.
- [24] U. Gretzel, K. Yoo and M. Purifoy, "Online travel review study: The role and impact of online travel reviews," *Laboratory for Intelligent Systems in Tourism 2007*.