# THE INTERNATIONAL JOURNAL OF BUSINESS & MANAGEMENT

# Finding Hotel and Tourism Expert through Social Network Analysis

**Sri Widiyanesti**
Lecturer, Faculty of Economic and Business, Telkom University, Indonesia
**Ihsan Maulana Riyadi**
Student, Faculty of Economic and Business, Telkom University, Indonesia

*Abstract:*
*In world of tourism, mainly in this age of technology, the entrepreneurs should make a digital marketing, thereby could reach a larger market, gained a new customer easier, and maintained an existing customer loyalty. One of the examples of the digital marketing media is TripAdvisor. TripAdvisor is an online medium for entrepreneurs in tourism industry to market his/ her product or service. Even though TripAdvisor is one of the largest place for travellers to get some information regarding hotel, restaurant, and travel destination, but in the website someone was found making a fake review & commenting its competitors to attack each other. On the other hand, not just the business owner, the comment that isn't corresponding and considered overboard with TripAdvisor policy, will be deleted. The aim and result from this research is to find a key player and hotel & tourism community through Twitter, therefore found some referenced and trusted account for the travellers to gain an information surrounding hotel & tourism, without completely rely on Trip Advisor.*

*The data gathering method in this research was using a technique called data crawling (social network analysis), through Twitter. Then the data was cleaned, so the left was just actors and target that are relevant with the research goal, after that process, the data accounts modelled in a graphic, thus formed a tourism community pattern and found some accounts that can be trusted & have an information surrounding hotel and tourism.*

*Keywords: Social network analysis, key player, community pattern, tripasvisor, fake review, hotel & toursim*

## 1. Introduction

Lately, whereas checking, reviewing, and comparing one hotel with another, did it through an online platform. Lots of online forum media provide a place for people to check, compare, and review between hotels, venues, and restaurants. One of the platforms is TripAdvisor, known as the biggest travel review website in the world (TripAdvsior, 2019). TripAdvisor covering lots of hospitality venue, such as hotel, restaurant, landmarks, flight, etc. In tourism industry, online review and rating has a significant effect in performance and successfulness of a hotel. Gaining 1% of hotel reputation in an online media will increased the Average Daily Rate (ADR) by 0.89%, followed by occupancy rate by 0.54%, and revenue per available room by 1.42% (Anderson, 2012).

2007, in one of the review in TripAdvisor, a restaurant called 'The Shed at Dulwich', became a restaurant with the best review in London, but in reality, the restaurant was nowhere to be found. This kind of phenomenon was rarely found, but this kind of incident keeps increasing by the day. Fake review mostly found in a hotel and restaurant. TripAdvisor became a media that used by business owners to make a negative review against their competitors, even exaggerated positive review for themselves. Another report says negative review was deleted without reason, especially if TripAdvisor became an advertiser or used as a marketing campaign pay-per-click with a business partner. These things will lead to unrealistic review or opinion (traveladdicts, 2019).

An independent research conducted in 2019, with 250.000 reviews as the samples. 15% of overall samples are fake. These fake reviews lead to TripAdvisor. They acknowledged that 93% did use a fake review, in other word, business that had use fake review still make an appearance in TripAdvisor (theguardian.com, 2019).

A research conducted by Lappas et al. (2016), fake review affects significantly against hotel credibility, majority of the hotels in 17 cities that become the samples in the research were affected by it, even professional review-authoring companies dare to post a fake review for money. This study aims were to help users/ travelers to find the utmost reliable expert tourism account in Twitter. The tourism expert that was found will be a new source of information for tourists and travelers.

In this research we are focusing not on the TripAdvisor, but the Twitter users. Twitter users in this case are the experts in hotel and tourism. We crawled the data surrounding hotel and tourism through social media Twitter, then we applied community detection using modularity class, because of its efficiency and easy to understand compared to other method, and then we find the key player from it. From the result we can know how's the community will formed, how big and their size is, and who's the user that leads certain community, and who's the key players are.

## 2. Literature Review

### 2.1. Data Mining

Gathering information from data warehouse and changed it into something that can easily understand is the meaning of data mining. Data analytics definition started use as reference for data mining (Lake & Crowther, 2013). According to Scionti & Elliot (2003), to understand data mining in marketing is the gathering process that is unknown, comprehensive, and information that can be executed from big data storage, as well used as crucial business decision, help implement it, including marketing initiative strategy with successes measure.

### 2.2. Machine Learning

Machine learning needed by the computer as a tool to learn (or to increase performance) based by data. The centre research area for computer program was to learn a complex pattern automatically and make a smart decision according data has been read (Han et al, 2012). Machine learning is a scientific study about algorithm and statistic model that used by computer for doing certain tasks effectively without explicit instruction, with just relying on pattern and inference as a substitute. Machine learning algorithm builds mathematic model based on data sample, known as data training, to make prediction or decision without programmed explicitly for doing certain tasks (Koza et al, 1996). Machine learning algorithm used in many certain applications, such as sorting email, and computer vision. Machine learning highly related to computer statistic, which focused on making prediction using computer. Study about mathematic optimization gives method, theory, and application domain to machine learning discipline. Data mining is a part of machine learning study, which focused on data exploration analysis through unsupervised learning (Friedman, 1998). In business practices, machine learning also called as predictive analysis.

### 2.3. Social Network Analysis

Definition of social network analysis is a branch discipline from human behavior and graphs (Tsvetovat & Kouznetsov, 2011). According to Otte & Rousseau (2014), social network analysis can be used as a way to gather information, that contain interaction between one user and another, the interaction itself pictured by a graph. Social network analysis known as, in another definition, a social structure, which formed by a group of dots or nodes that represent individual or actor and connect between each other through ties, friendship, family, favourite things, unfavourite things, and social status (Wasserman, 2008).

For detecting community, we used modularity metrics, that based on research by Blonde et al. (2008). This method measures ratio in other communities. Connection between actors in certain community is bigger than outside of it. The formula for the modularity metric shown in (1). We have Q as modularity value, which is between [0,1], the bigger the value means the community has a clear structure, vice versa. $m$ represents the connection or edges number. $A_{ij}$ represents the adjacency matrix of network. The node itself represented by $i$ & $j$ and the degree of it represented by $k_i$ & $k_j$. Last, $\delta(c_i c_j)$ is a kronecker of community $i$ & $j$.

$$Q = \frac{1}{2m} \sum_{i,j} [A_{ij} - \frac{k_i k_j}{2m}] \delta(c_i c_j) \quad (1)$$

To analyze the most important person in certain community we use the key player method. There are several characteristic that we can use to find the key player, such as degree centrality, betweenness centrality, and closeness centrality. Degree centrality defined as the total relationship or ties someone have. Actor with lots of connection will have higher value of degree and will be the center of network structure (Umadevi, 2013). For degree centrality formula shown in (2). We have $CD(ni)$ as the degree value and $d(ni)$ as the degree from the actor.

$$CD(ni) = d(ni) \quad (2)$$

Betweenness centrality defined as the shortest path that goes through the actor. Actor with the highest betweenness centrality will become the important key to connect from certain community to the other, vice versa. Other than that, the actor can control and manipulate the information that goes within his/ her network. If this actor was to be removed from certain network, then the path to get information through certain communities will take much longer (Umadevi, 2013). For betweenness centrality formula shown in (3), with $C_b(n_1)$ as the value. 𝑔𝑗𝑖𝑘 as the total shortest path between actor j and k through actor i. $gjk$ represents the total shortest path from actor j to k.

$$C_b(n_1) = \sum_{j,k \neq i} \frac{gjik}{gjk} \quad (3)$$

Closeness centrality defined as the distance between actors with focusing on the shortest path. Closeness can be used as a measure for how long does it take to carry information to the actor and then to another (Umadevi, 2013). The formula for closeness centrality shown in (4). $C_c(C_c)$ is the closeness centrality value. (𝑑1, ) represents the distance between two actors.

$$C_c(C_c) = \sum_{i=Ni} \frac{1}{d(n1,nj)} \quad (4)$$

### 2.4. User Generated Content

Lately, the mix development of social network, gadget computation, and cloud computing is getting faster. This trend encouraged users to interact actively and work through their own network. Through social network, user can comment trend issues, asks, answers, reviews, and purchases product through an online store. Through their gadget, they can check-in in tourist destination, and share photo and video easily.

Social media content that made by the user itself, known as user generated content (UGC) (Moens et al, 2014). UGC meaning came from various resource, including social network site, like Facebook & LinkedIn, microblog sites, such as Twitter, or a place to share video, like YouTube and Vimeo. UGC can be called CGM (Consumer Generated Media). OECD (Organisation for Economic Cooperation and Development) says that UGC was a content that was published into a site and can be accessed by public or social media website that can open by users.

*2.5. Twitter*

Twitter is a microblog service or media to communicate with friend, family, and co-worker with messaging system called tweet with a maximum 140 characters. In Twitter user can upload tweet that contain photos, GIF, video, or links. Even Twitter itself provides a direct messaging system called DM with a requirement to follow each party (Twitter, 2019). According to Economic and Social Research Council (2019), Twitter is very popular among student, politician, government, and public figure. Twitter can be also used as:

- Promoting research, sharing through links to a blog or journal website
- Reach audience faster with tweet and retweet
- Follow a researcher with the same topic
- Making a contact with the experts
- Reach new audience or user
- Getting a feedback

## 3. Experiment and Analysis

We collect the data from a certain period of time and related to review object that appear in TripAdvisor. Thus we used certain hashtag, such as #flight, #hotel, #rental, #cruise, #restaurant, #landmark, and #destination. The period of time for crawling the data are 8 January 2020 – 9 February 2020. In total, we have 166.949 raw data. The software we used to crawl is R-Studio.

*3.1. Network Graph*

Firstly, we transform the raw data into clean one. Leaving only the source and the target. The user's screen name represents the source and the mentioned represented by the target. Then we connect each source and target one by one. Because some source has multiple targets, we need to make a new separate row, so each source has one target. After making new rows, we input it into a software called Gephi, which will making the graph.

*3.2. Community Pattern*

After the graph is drawn, we calculate it using the modularity class statistic to detect the community in the social network. There are more or less 7 group of communities in the network. Each community defined by a certain group of color, that applies as well to the actors/ nodes. Actors that has a same characteristic will grouped in the same community. From the Fig. 1 visualization, there are some major account that leads the community. Other than using modularity class, we use other method of calculation for measuring centrality, such as degree, betweenness centrality, and closeness centrality. From the figure below, the dominant actors are Rashada Writes and travelbag tours, with each has 317 and 302 degree respectively, leading the blue and orange community.
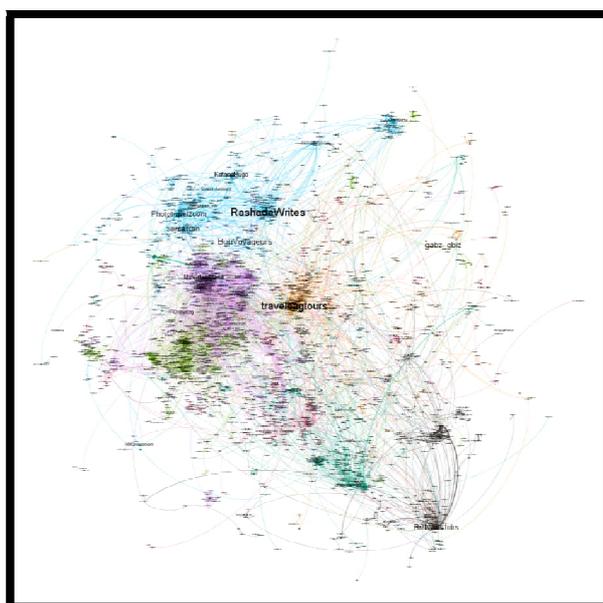


*Figure 1: Social Network with Modularity Class*

| No. | Actors | Degree | Betweenness Centrality | Closeness Centrality | Modularity Class |
|-----|--------|--------|------------------------|----------------------|------------------|
| 1 | RashadaWrites | 317 | 2485981.246 | 0.222 | 1888 |
| 2 | travelbagtours | 302 | 4861718.325 | 0.237 | 1528 |
| 3 | BonVoyageurs | 249 | 192656.056 | 0.187 | 1888 |
| 4 | gabz_gbiz | 229 | 23572.896 | 0.146 | 1828 |
| 5 | Phototravelzcom | 217 | 1401463.327 | 0.220 | 1888 |
| 6 | BizNewsJobs | 207 | 3642695.313 | 0.221 | 788 |
| 7 | KatanaHugo | 181 | 814519.340 | 0.210 | 1888 |
| 8 | SaveATrain | 172 | 1080392.295 | 0.219 | 1888 |
| 9 | MyVirtualVaca | 151 | 1733204.161 | 0.225 | 34 |
| 10 | HALcruises | 127 | 625680.562 | 0.211 | 34 |
| 11 | Youtube | 126 | 970102.448 | 0.212 | 2061 |
| 12 | Cruiselog | 119 | 534394.161 | 0.207 | 34 |

*Table 1: Most Influential Actors in the Network Ranked by Degree*

*3.3 Result Analysis*

Based on the Table 1 we have 12 dominant actors in the hotel & tourism community and in different group. But for travelbagtours, phototravelzcom, gabz_gbiz, BizNewsJobs, KatanaHugo, SaveATrain, and HALcruises, Youtube are not account that will give information surrounding hotel & tourism, instead they provide hotel booking, travel picture, a cinematographer, a service that connects supplier and buyer in hospitality industry, travel & nature picture, train ticket, world cruise, and content provider respectively, so we need to discard them. Thus, the source of information that the travelers that can be relied on regarding hotel & tourism are RashadaWrites, BonVoyageurs, MyVirtualVaca, and Cruiselog. Respectively they are a freelance writer & travel blogger, travel experts & influencer in social media, travel blogger, and a reporter & founder of USA Today mainly in cruises topic.

Thus, RashadaWrites, BonVoyageurs, MyVirtualVaca, and Cruiselog can be made as a new source of information regarding hotel & information. Travellers that are doubt with TripAdvisor review, they can lean on to these accounts. Unfortunately, some of the account are not covering each/ all review object in TripAdvisor. So, the travelers need to contact the person itself through a direct way, such as face to face consultation.

From the four of them, RashadaWrites has the biggest value of Betweenness Centrality (2485981.246) followed by BonVoyageurs (192656.056), MyVirtualVaca (1733204.161), and Cruiselog (534394.161). For the biggest value for Closeness Centrality lead by RashadaWrites (0.222), followed by My Virtual Vaca (0.225), Cruiselog (0.207), and Bon Voyageurs (0.187). For the biggest value in Modularity Class is RashadaWrites and BonVoyageurs with a same amount of value (1888), followed by MyVirtualVaca and Cruiselog with a same amount of value as well (34). From these explanations we can conclude that RashadaWrites is the account that counted as the key player followed by BonVoyageurs, MyVirtual Vaca, and Cruiselog. RashadaWrites and BonVoyageurs leads the blue and purple community with MyVirtualVaca and Cruise log acting as the connecting node between those two community. Thus as a whole the hotel & tourism in this research mainly focus on blue and purple community, while the other communities are irrelevant to the research.

## 4. Conclusion

This research shows that finding the community and the important people that experts in hotel & tourism can be done through social network analysis. From the result, majority of the actors are indirectly having a connection to hotel & tourism but don't have the knowledge or has the common idea about hotel & tourism, while only minority has it. Even the minority has a certain particular hotel & tourism area that were covered. In the end the minority always connect to each other through certain actor, leading to high, vast, and fast flow of information within the network.

## 5. Recommendation

This research focusing only in particular area of hotel & tourism and bound by object in TripAdvisor. While it answered all of the question in the research. To increase the accuracy and complexity of the research, we hope that the next research adds more object and sample to the research while combining with other methods.

## 6. References

i. Anderson, C. (2012). The Impact of Social Media on Lodging. *Cornell Center for Hospitality Research Publications*, 5-11.

ii. Chua, T.-S., Li, J., & Moens, M.-F. (2014). *Mining User Generated Content.* New York: CRC Press.

iii. Crowther, P., & Lake, P. (2013). *Concise Guide to Databases.* London: Springer.

iv. Economic and Social Research Council. (2019, Desember 15). *What is Twitter and why should you use it?* Retrieved from esrc.ukri.org: https://esrc.ukri.org/research/impact-toolkit/social-media/twitter/what-is-twitter/

v. Friedman, J. H. (1997). Data Mining and Statistic: What's the connection? *Proceedings of the 29th Symposium on the Interface Between Computer Science and Statistics*, (pp. 5-9).

vi. Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques.* Waltham: Elsevier Inc.

vii. Kouznetsov, A., & Tsvetovat, M. (2011). *Social Network Analysis for Startups.* Sebastopol: O'Reilly Media.

viii. Koza, J. R., Bennett III, F. H., Andre, D., & Keane, M. A. (1996). Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. *Artificial Intelligence in Design '96*, 151-170.

ix. Longwell, L. (2019, Desember 8). *What's Wrong with TripAdvisor, and What to Do About It*. Retrieved from TravelAddicts: https://traveladdicts.net/tripadvisor-fake-reviews/

x. Otte, E., & Rousseau, R. (2002). Social Network Analysis: A Powerful Strategy, Also for the Information Sciences. *Journal of Information Science*, 441-453.

xi. Scionti, R., & Elliott, K. (2003). *The Confluence of Data Mining and Market Research for Smarter CRM.* Two Rivers.

xii. TheGuardian. (2019, September 6). *TripAdvisor is failing to stop fake hotel reviews, says Which?* Retrieved from The Guardian: https://www.theguardian.com/travel/2019/sep/06/tripadvisor-failing-to-stop-fake-hotel-reviews-which

xiii. TripAdvisor. (2019). *About TripAdvisor*. Retrieved from TripAdvisor Media Center: https://tripadvisor.mediaroom.com/US-about-us

xiv. Twitter. (2019, Desember 16). *Using Twitter*. Retrieved from Twitter Help Center: https://help.twitter.com/en

xv. Umadevi, V. (2013). Case Study – Centrality Measure Analysis On Co-Authorship Network. *Journal of Global Research in Computer Science*, 67-70.

Valkanas, G., Sabnis, G., & Lappas, T. (2016). The Impact of Fake Reviews on Online Visibility: A Vulnerability Assessment of the Hotel Industry. *Information System Research*, 940-961.Anderson, C. (2012). The Impact of Social Media on Lodging. *Cornell Center for Hospitality Research Publications* , 5-11.

Chua, T.-S., Li, J., & Moens, M.-F. (2014). *Mining User Generated Content.* New York: CRC Press.

Crowther, P., & Lake, P. (2013). *Concise Guide to Databases.* London: Springer.

Economic and Social Research Council. (2019, Desember 15). *What is Twitter and why should you use it?* Retrieved from esrc.ukri.org: https://esrc.ukri.org/research/impact-toolkit/social-media/twitter/what-is-twitter/

Friedman, J. H. (1997). Data Mining and Statistic: What's the connection? *Proceedings of the 29th Symposium on the Interface Between Computer Science and Statistics*, (pp. 5-9).

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques.* Waltham: Elsevier Inc.

Kouznetsov, A., & Tsvetovat, M. (2011). *Social Network Analysis for Startups.* Sebastopol: O'Reilly Media.

Koza, J. R., Bennett III, F. H., Andre, D., & Keane, M. A. (1996). Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. *Artificial Intelligence in Design '96* , 151-170.

Longwell, L. (2019, Desember 8). *What's Wrong with TripAdvisor, and What to Do About It*. Retrieved from TravelAddicts: https://traveladdicts.net/tripadvisor-fake-reviews/

Otte, E., & Rousseau, R. (2002). Social Network Analysis: A Powerful Strategy, Also for the Information Sciences. *Journal of Information Science* , 441-453.

Scionti, R., & Elliott, K. (2003). *The Confluence of Data Mining and Market Research for Smarter CRM.* Two Rivers.

TheGuardian. (2019, September 6). *TripAdvisor is failing to stop fake hotel reviews, says Which?* Retrieved from The Guardian: https://www.theguardian.com/travel/2019/sep/06/tripadvisor-failing-to-stop-fake-hotel-reviews-which

TripAdvisor. (2019). *About TripAdvisor*. Retrieved from TripAdvisor Media Center: https://tripadvisor.mediaroom.com/US-about-us

Twitter. (2019, Desember 16). *Using Twitter*. Retrieved from Twitter Help Center: https://help.twitter.com/en

Umadevi, V. (2013). Case Study – Centrality Measure Analysis On Co-Authorship Network. *Journal of Global Research in Computer Science* , 67-70.

Valkanas, G., Sabnis, G., & Lappas, T. (2016). The Impact of Fake Reviews on Online Visibility: A Vulnerability Assessment of the Hotel Industry. *Information System Research* , 940-961.

Wasserman, S. (1994). *Social Network Analysis: Methods and Applications.* Cambridge University Press.