Resources, Conservation & Recycling 164 (2021) 105146

Contents lists available at ScienceDirect

Resources, Conservation & Recycling

journal homepage: www.elsevier.com/locate/resconrec

Full length article

Data analytics of social media publicity to enhance household waste management

Peng Jiang[a,*], Yee Van Fan[b], Jiří Jaromír Klemeš[b]

[a] Department of Systems Science, Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), Singapore 138632, Singapore
[b] Sustainable Process Integration Laboratory – SPIL, NETME Centre, Faculty of Mechanical Engineering, Brno University of Technology - VUT Brno, Technická 2896/2, 616 69 Brno, Czechia

ARTICLE INFO

Keywords:
Digital waste management
User engagement
Internet of Things
Publicity improvement
Text data mining

ABSTRACT

Household waste segregation and recycling is ranked at a high priority of the waste management hierarchy. Its management remains a great challenge due to the high dependency on social behaviours. The integration of Internet of Things (IoT) and subscription accounts on social media platforms related to household waste management could be an effective and environmentally friendly publicity approach than traditional publicity via posters and newspapers. However, there is a paucity of literature on measuring social media publicity in household waste management, which brings challenges for practitioners to characterise and improve this publicity pathway. In this study, under an integrated framework, data mining approaches are employed or extended for multidimensional publicity analytics using the data of online footprints of propagandist and users. A real-world case study based on a subscription account on the WeChat platform, Shanghai Green Account, is analysed to reveal useful insights for personalised improvements of household waste management. This study suggests that the current publicity related to household waste management leans towards propagandist-centred in both timing and topic dimensions. The identified timing, which has high user engagement, is 12:00–13:00 and 21:00–22:00 on Thursday. The overall relative publicity quality of historical posts is calculated as 0.95. Average user engagement under the macro policy in Shanghai was elevated by 138.5% from 2018 to 2019, during which the collections of biodegradable food waste and recyclable waste were elevated by 88.8% and 431.8%. Intelligent decision support by publicity analytics could enhance household waste management through effective communication.

1. Introduction

Effective household waste management is a crucial challenge due to the involvement of multiple stakeholders, including managers, sanitation companies, workers and vast residents. The publicity of household waste management acts as a communication channel for the stakeholders. The manner of publicity evolves rapidly during the past three decades. The primary publicity pathway before the year 2000 was traditional one-way media, such as newsletters, magazines and posters. In the internet era after around 2003, more and more publicity of household water management was spread through web pages. In recent years, the use of social media has been widespread. According to the Global Digital 2019 reports, 3.48×10^9 population worldwide (i.e. 45%) were active social media users with a 9% increasing rate compared to 2018 (Simon, 2019). Social media information has been demonstrated to create operational value (Cui et al., 2018), promote online value co-creation (Frempong et al., 2020) and improve policy regulations (Sun et al., 2020). With the development of Internet of Things (IoT), subscription accounts on social media platforms are being applied to household waste management innovatively, such as food waste management using Facebook and Instagram (Martin-Rios et al., 2018) and waste electronic equipment collection with the help of WeChat, Weibo and Sina blogs (Zuo et al., 2020). In a survey with 850 samples about the information pathways of waste recycling, the proportions of newspaper and television, internet and social media were 29%, 26% and 45% (Ramzan et al., 2019). Social media (45%) appears as a compelling pathway in conveying messages. However, compared to other industrial sectors, the digitalisation in waste management was still in its infancy (Sarc et al., 2019). This paper focuses on social media publicity related to household waste management.

In the field of household waste management, social media provides a sustainable and environmentally friendly publicity pathway

* Corresponding author at: 1 Fusionopolis Way, #15-16, Connexis North Tower, Singapore 138632.
E-mail addresses: jiang_peng@ihpc.a-star.edu.sg, peng.dr.jiang@gmail.com (P. Jiang).

https://doi.org/10.1016/j.resconrec.2020.105146
Received 16 June 2020; Received in revised form 24 August 2020; Accepted 28 August 2020
Available online 02 September 2020
0921-3449/

compared to traditional publicity via posters and newspapers. It brings incomparable convenience for residents in at least three aspects:

i) Information on social media presents responsive access to global society. For example, a subscription account on Facebook, @ WasteManagement, had 181,464 active users by 22 August 2020 (Facebook, 2020). Users are quickly informed of the latest regulations, news and opportunities by social media. Considering the recency of regulations has an impact on behaviours and the related implementation (Mintz et al., 2019), social media can promote waste recycling (DEEP, 2020) and increase the participation of stakeholders by communicating with residents persuasively (Knickmeyer, 2019).
ii) Social media can provide instant interaction and feedback regarding their concerns and suggestions (Zamri et al., 2019). The computer-based feedback systems can be tailored to track recycling activity and send feedback in the whole social networks, which help to significantly increase segregation accuracy and nurture recycling behaviour (Knickmeyer, 2019).
iii) Social media serves as a facilitator for strengthening the willingness to adopt more environmentally friendly attitudes. The public reaction in Australia was quite positive when managers advised the general public to engage in recycling and resource recovery via social media (WMR, 2015). Wamuyu (2018) proposed Web 2.0 and social media technologies and platforms to foster environmentally friendly behaviours for residents, even those residents in low-income urban communities.

The ultimate goals of social media publicity are to raise environmental awareness of residents (Mallick and Bajpai, 2019) and influence/change public behaviours, such as waste reduction behaviour (Young et al., 2017) and recycling behaviour (Sujata et al., 2019). The indirect influence mechanism of information publicity on behaviour and intention has been investigated by Wang et al. (2018). The social media publicity and characteristics of a subscription account should be figured out adequately to promote awareness of residents and guide public behaviours on waste reduction, segregation and recycling. There is a paucity of literature on measuring social media publicity systematically in household waste management. Generally, the popularity of a subscription account on social media could be reflected by user engagement (Aldous et al., 2019), such as views (Jiang et al., 2019), likes (Ksiazek et al., 2016), user-to-user interaction (Hu et al., 2017) and user comments (Gaenssle and Budzinski, 2020). Previous observations indicated that the popularity of a social media post regarding household waste management was related to its publishing time (Ma and Chen, 2017), published title and topics (Jing and Wei, 2016) in attracting the user attention. However, the unsystematic measures and limited statistical analytics bring significant challenges for practitioners and managers to characterise and improve social media publicity related to household waste management.

This study aims to offer systematic measures for social media publicity related to waste management by analysing the online footprints of propagandist (i.e. the person who manages and oversees a subscription account) and users (i.e. residents who subscribe the account). The analysed results are useful in further improving social media publicity, thereby targeting to raise environmental awareness of residents and change public behaviours of waste segregation and recycling. The major contributions of this study are summarised as follows:

a) The online footprints of propagandist and users of the subscription account, including both time series data and textual data, are innovatively applied to social media data mining in household waste management to measure and improve social media publicity in terms of publishing time, publicity topics, user popularity, status and trend, view-like relation, relative publicity quality and user feedback.
b) An integrated framework for related publicity analytics is constructed, which is a closed-loop system on data collection, measure design, data analytics, visualisation and decision-support insights. For ease of applications, visualisation representations of social media publicity are offered for managers to gain the intuitive impressions of publicity analytic results of a specific subscription account.
c) The observations yielded through a real-world case study in Shanghai assist in improving the effects of social media publicity and enhance household waste management. The extracted information in user comments helps to promote household waste segregation and recycling.

This paper is structured by sections. Section 2 presents the methodology. Section 3 shows a real-world application. Section 4 concludes this paper.

2. Methodology

This section introduces the integrated framework, data collection method, and data mining approaches for measuring and analysing social media publicity.

2.1. Framework

The proposed integrated framework of measures and analytics for social media publicity is presented in Fig. 1. The framework is a closed-loop system including the subscription account management, data collection, publicity analytics, visualisation, insight generation and waste management enhancement. A specific subscription account that connects user clusters directly or indirectly publishes posts mainly about waste management related contents. The collected data, online footprints of propagandist and users, are modelled by statistical and machine learning techniques to answer the following questions:

i) What are the time distribution and peak time of published posts under uncertainties?
ii) What are the ranking order and the contribution levels of the major publicity topics?
iii) What is the user popularity in terms of the number of views or likes of these posts?
iv) What are the life cycle status and trend of publicity in a specific subscription account?
v) What is the potential relationship between views and likes in a subscription account?
vi) What is the relative publicity quality by those posts with a few likes or without likes?
vii) What is the user feedback, and what are the topics behind active user comments?

Based on the seven questions, the design of inter-related measures is presented in Fig. 1, where seven modelling tasks are matched with the related measures. After measuring, analysing and visualising social media publicity, the insights and possible decision support of practical applications are provided for managers to improve social media publicity and enhance household waste management further.

2.2. Data collection

For a subscription account on a social media platform, the data to be collected are divided into two categories—namely, online propagandist footprints and online user footprints. Table S1 in Supporting Information shows the ranking order and some online footprints of the top ten social media platforms with the subscription account functions. Overall, except Snapchat, the other nine platforms have similarity in online footprints. The propagandist footprints include publishing time,

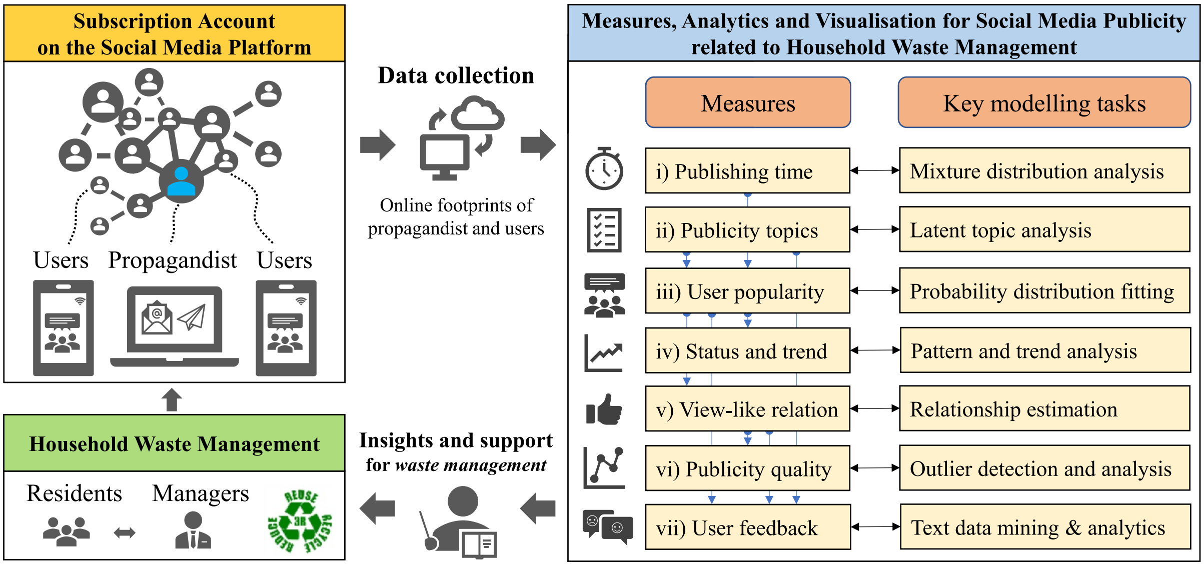

Fig. 1. The integrated framework of measures and analytics for social media publicity related to household waste management.

title information, published contents and replies for user comments. The user footprints include information of views, likes (e.g. thumb, love and voting) and user comments. In this study, the analysis approaches of published topics are developed for the title information of posts (i.e. text data rather than voice or video data) so that the approaches have broader applications for different social media platforms.

Different cities develop respective subscription accounts that oversee their own publicities of interest. Since the subscription accounts are relatively independent in terms of published posts, this study focuses on measuring social media publicity of an individual account. The corresponding online footprints of anonymous users are collected and matched with the information of historical posts in a subscription account. As posts and online footprints on most platforms are public data, they can be easily accessed and collected. What researchers need to be careful is the privacy protection which can be guaranteed via using anonymous information for data analytics. Regarding how long the time duration of data collection should be, we suggest collecting at least a whole year data since the trend and topic analyse need enough data to be insightful.

2.3. Publicity measures and analytics

2.3.1. Publishing time

The publishing time or date may not be fixed due to numerous uncertain factors involved in the publishing process. It is foreseeable that the peak time of different subscription accounts varies significantly. To measure the publishing date, the conventional way, frequency distribution histogram (Gu et al., 2015), can count the week pattern of the number of posts. However, for the publishing time within a day, larger uncertainties and the resultant multiple peak values in the frequency distribution histogram hinder the estimation of the real publishing behaviour. A continuous probability distribution offers a robust way to estimate the distribution of publishing time and the potential peaks in the distribution. A Gaussian mixture model can fit complicated distributions using several Gaussian components, which has been widely used for error distribution fitting (Zhang et al., 2019) and conditional probability distribution fitting (Jiang and Liu, 2016). The Gaussian mixture distribution $p(t)$ is designed to fit the distribution of publishing time:

$$p(t) = \sum_{k=1}^{K} c_k p(t|k) = \sum_{k=1}^{K} c_k \Phi[t|\mu_k, \Sigma_k], \tag{1}$$

$$\sum_{k=1}^{K} c_k = 1, \tag{2}$$

$$t^* = arg \max_{\forall t} p(t), \tag{3}$$

where t denotes the publishing time; $\Phi[t|\mu_k, \Sigma_k]$ is Gaussian density; c_k, μ_k and Σ_k represent the weight, mean vector and covariance matrix for the mixture component k. The sum of weights equals 1, as shown in Eq. (2). The publishing peak time t^* is given by Eq. (3). Although better fitting results can be yielded by adding more components, more parameters in the mixture distribution with multiple components imply a higher risk of overfitting. The Gaussian mixture distribution with the minimal Akaike's information criterion (AIC) value is regarded as the best-fitted distribution (Akaike, 2011). After modelling and analyses, the relationship between the historical publishing time/date and the resultant user engagement can be analysed to improve the future publishing time/date.

2.3.2. Publicity topics

The publicity topics on social media are generally diverse to keep the followers interested. Cole (2016) suggested that a large percentage of diverse and high-quality contents in a subscription account should be curated by informative social media posts or blog posts from other waste management accounts. It is beneficial for managers to identify the underlying topics in social media publicity related to household waste management. The latent Dirichlet allocation (LDA) (Blei et al., 2003) is a commonly used topic analysis model which infers the keyword probabilities in topics and discovers major underlying topics in the text data. The prior parameters of an LDA model can be tuned by heuristic algorithms, such as the differential evolution (Agrawal et al., 2018) and the adaptive particle swarm optimisation (Jiang et al., 2017). More details about the topic modelling and its applications are referred to Jelodar et al. (2019). This study builds an LDA model for the title information analytics regarding social media publicity to discover its underlying topics. As the data volume is significantly reduced by just modelling the title information, the inference algorithm of collapsed Gibbs sampling (Porteous et al., 2008) can be faster and more accurate than various variational Bayes algorithms (Jelodar et al., 2019). The optimal number of underlying topics of the LDA model is determined by the minimal perplexity of a held-out set of posts (Mathworks, 2020a). The trained LDA model generates the optimal J underlying topics and the probabilities ζ_{ij} of topic mixtures ($j \in \{1, 2, ..., J\}$) for each post i, $\forall i \in \{1, 2, ..., I\}$. The probabilities ζ_{ij} satisfy Eq. (4). The sum of probabilities ψ_j in Eq. (5) is then defined as the usage-degree-of-topic in historical posts for each topic j.

$$\sum_{j=1}^{J} \zeta_{ij} = 1, \quad \forall\, i \tag{4}$$

$$\psi_j = \sum_{i=1}^{I} \zeta_{ij}, \quad \forall\, j \tag{5}$$

The topic analysis is further extended to understand the publicity better. The probabilities ζ_{ij} generated by the LDA model and the number of views (or likes) are innovatively linked via a regression model to investigate the contribution-level-of-topic for the number of views (or likes). As the topics yielded by an LDA model are independent, we build a multivariate linear regression model for the number of views (or likes) and the probabilities of topics, as expressed as Eq. (6) and Eq. (7). The contribution-level-of-topic is represented by the regression coefficients $\mathbf{w}^*$:

$$\hat{y}_i = \sum_{j=1}^{J} w_j \zeta_{ij}, \quad \forall\, i, \tag{6}$$

$$\mathbf{w}^* = \arg\min_{\forall \mathbf{w}} \frac{1}{T} \sum_{i=1}^{I} (y_i - \hat{y}_i)^2 = \arg\min_{\forall \mathbf{w}} \frac{1}{T} \sum_{i=1}^{I} \left(y_i - \sum_{j=1}^{J} w_j \zeta_{ij} \right)^2, \tag{7}$$

where y_i is the number of views or likes of the post i; $\hat{y}_i$ denotes the estimated value of y_i. The contribution-level-of-topic results could improve social media publicity by promoting more publications with great public attention.

2.3.3. User popularity

Views, likes, comments and replies were suggested to indicate the popularity of social media by Jiang et al. (2019). In this section, the numbers of views and likes are considered preferentially to measure the user popularity of published posts. The comments and replies are discussed later and regarded as two-way feedback in Section 2.3.7. Since the number of views or likes has different scales for different posts, a probability distribution of the number of views or likes is a competent and robust way to estimate the user popularity of historical posts. Judged by the AIC (Akaike, 2011) as shown in Eq. (8), the most appropriate distribution d^* with the smallest AIC value in Eq. (9) is identified for the popularity distribution amongst the 19 commonly used probability distributions (Jiang et al., 2020):

$$\mathrm{AIC}_d = -2\ln(L_d) + 2n_d, \quad \forall\, d, \tag{8}$$

$$d^* = arg \min_{\forall d} \mathrm{AIC}_d, \tag{9}$$

where L_d is the likelihood value of the distribution fitting under the probability distribution d, $d \in \{1, 2, ..., 19\}$; n_d is the number of parameters of the candidate distribution d.

There are scarce symmetric distributions for the number of views or likes in real-world cases. The sample skewness S (Doane and Seward, 2011), as shown in Eq. (10), is a natural statistical measure to estimate the skewness of a distribution:

$$S = \frac{\sqrt{I(I-1)}}{I-2} \frac{\frac{1}{I}\sum_{i=1}^{I} (y_i - \bar{y})^3}{\left(\sqrt{\frac{1}{I}\sum_{i=1}^{I} (y_i - \bar{y})^2}\right)^3}, \tag{10}$$

where I is the number of published posts; y_i is the number of views or likes, $i \in \{1, 2, ..., I\}$; $\bar{y}$ denotes the sample mean.

2.3.4. Status and trend

For different social media platforms, views and likes are fundamental factors to measure user popularity and the success of a post on social media. The number pattern of views or likes can be further analysed to measure the status and trend of social media publicity. Amongst statistical tools, the boxplot (Frigge et al., 1989) is a competent candidate for such status and trend analytics as it can assist in discovering time series patterns from statistical data (Jiang et al., 2020). The division of time duration for trend analysis depends on the availability of historical data. If the numbers of views and likes are collected during several years, the boxplot generates not only annual mean values but also their five percentiles, including a top whisker (a 95th percentile), a top of the box (a 75th percentile), a median value (a 50th percentile), a bottom of the box (a 25th percentile) and a low whisker (i.e. the minimum value) (Eriksen et al., 2018). The status of social media publicity is revealed by comparing the boxes. The time-series trend of the mean values and the five percentiles are further used to analyse the publicity trend of a subscription account. The status and trend can also help to judge the life-cycle stages of a subscription account by referring to the product life cycle (Franses, 2015). Since the contents of published posts in a subscription account would be different for different years, another interesting point regarding the annual trend analysis is the ratio of annual waste management related posts in total annual posts. The trend of annual ratios could assist in explaining the changes in user popularity.

2.3.5. View-like relation

For a published post in social media, from the perspective of a propagandist, a greater number of views that may trigger more likes from users implies better publicity. From the perspective of users, users may recommend their relatives and friends viewing the posts of interest, which, in turn, increases the number of post views. The number of views and the number of likes might have a correlation. Since the Spearman correlation test has no indicated limitation on the variable distribution (Myers et al., 2010), it has been used to analyse the potential relationship between the number of views and the number of likes. The Spearman correlation coefficient $\rho(v, l)$ is given by:

$$\rho(v, l) = \mathrm{cov}(\mathrm{rg}v, \mathrm{rg}l)/(\sigma_{\mathrm{rg}v}\sigma_{\mathrm{rg}l}), \tag{11}$$

where v denotes the variable of the number of views; l denotes the variable of the number of likes; $\mathrm{cov}(\mathrm{rg}v, \mathrm{rg}l)$ is the covariance between the rank variables $\mathrm{rg}v$ of v and $\mathrm{rg}l$ of l; $\sigma_{\mathrm{rg}v}$ and $\sigma_{\mathrm{rg}l}$ are standard deviations of $\mathrm{rg}v$ and $\mathrm{rg}l$.

Under the condition of a strong and statistically significant correlation, the view-like relation can be expressed as a view/like ratio, if a bivariate graph (e.g. scatterplot) suggests a relatively linear relationship. Otherwise, the view-like relation should be fitted by an appropriate nonlinear function, such as the polynomial, exponential and radial basis function. For the view/like ratio estimation, the strong assumption of normal distribution in traditional linear regression restricts its practical application. The parameter estimation of a linear regression under the Bayesian paradigm (Desgagné and Gagnon, 2019) has the superiority to deal with a heavy-tailed distribution and the ratio outliers. The view-like relation can be expressed as:

$$v = \begin{cases} a_{\mathrm{Bayes}} l + b, & if linear \\ f(l) + b', & if nonlinear \end{cases}, \tag{12}$$

where a_{Bayes} denotes the estimated ratio; b and b' are intercepts; $f(l)$ denotes a fitted nonlinear function. In a traditional linear regression, the estimated ratio can be solved in a closed form by a least-squares method. However, the a_{Bayes} under the Bayesian paradigm can only be solved by an approximation algorithm, e.g. a Markov chain Monte Carlo.

2.3.6. Publicity quality

Since user scales are different for subscription accounts, and publicity quality fluctuates as the time changes, it is challenging to estimate the absolute publicity quality of a subscription account. This study proposes a concept of relative publicity quality (RPQ) for social media publicity. By this concept, all historical posts are compared under the same subscription account. The publicity quality via social media depends on the user behaviour of likes. Considering the final

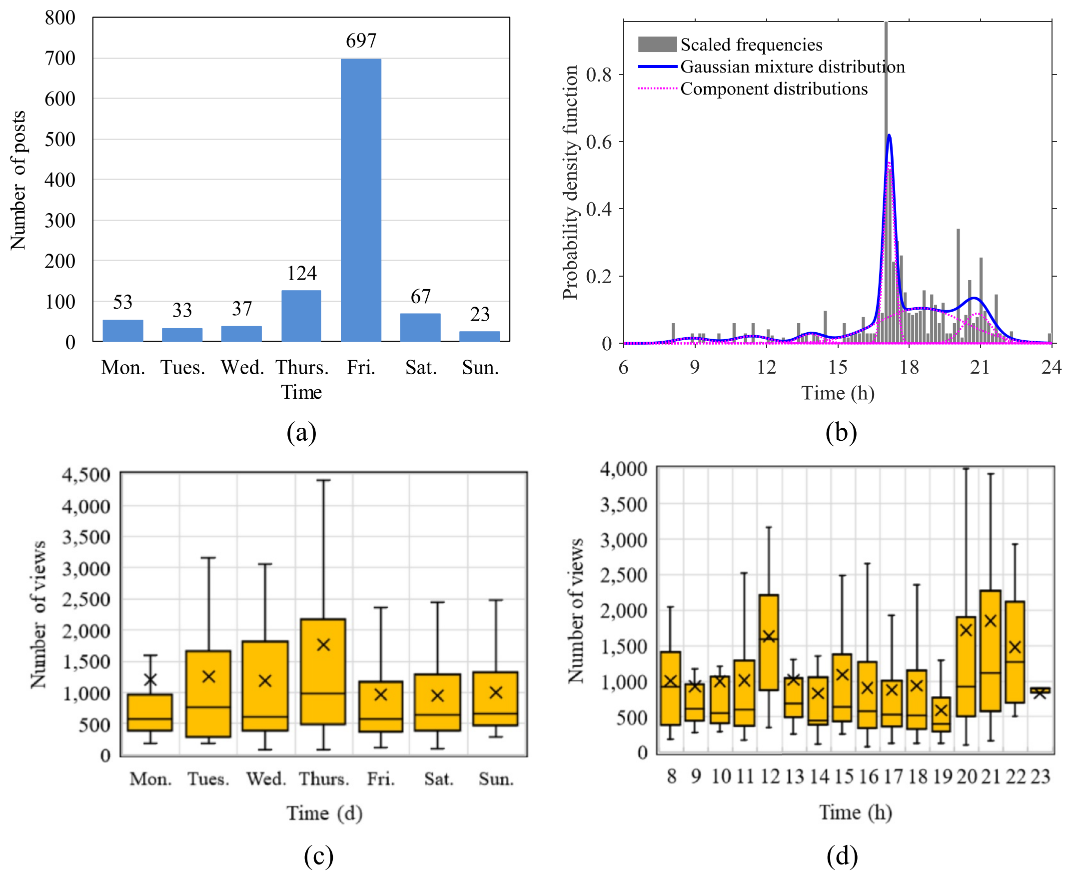

Fig. 2. The probability distribution of publishing time on a week scale (a) and a day scale (b) and the number distribution of user views on a week scale (c) and a day scale (d).

goal—namely, behaviour changes (Sujata et al., 2019) and environmental awareness promotion (Mallick and Bajpai, 2019)—of social media publicity related to household waste management, the number of likes is a natural measure that reflects potential willingness and awareness of users. RPQ is determined by two categories of posts with poor quality, i.e. the posts without likes and the posts related to outliers of the time-serial view/like ratio. The expression of RPQ and the identification of outliers are given by:

$$\mathrm{RPQ} = 1 - \frac{W}{I} - \frac{O}{I}, \tag{13}$$

$$F_{\mathrm{lower}} = Q_1 - \tau(Q_3 - Q_1), \tag{14}$$

$$F_{\mathrm{upper}} = Q_3 + \tau(Q_3 - Q_1), \tag{15}$$

where RPQ in Eq. (13) is scoped in the range of [0, 1] with 1 being a preferable value; I denotes the total number of historical posts; W denotes the number of the posts without likes; O denotes the number of outliers. In Eqs. (14) and Eq. (15), the ratio beyond its fence range $[F_{\mathrm{lower}}, F_{\mathrm{upper}}]$ is regarded as an outlier; Q_1 and Q_3 are conditional quartiles; $Q_3 - Q_1$ denotes the interquartile range for each ratio, where the range is estimated by the non-parametric quantile-based method (Mathworks, 2019) based on the quantile regression forests (Meinshausen, 2006); τ is a control coefficient for the fence range which is set as 1.5 (Krounbi et al., 2019).

2.3.7. User feedback

User comment is the primary feedback pathway on a subscription account. More likes and views possibly bring more user comments for a post. The Spearman correlation test (Myers et al., 2010) is used to analyse the relationship between the number of user comments and the number of views (or likes). Managers may also want to know the topics behind the posts with active user comments. The keyword frequency statistics can answer such a question for managers. The Bag-of-n-grams model (Mathworks, 2020b) is employed to record the number of times that each keyword appears in historical posts. Besides the correlation measure and the topic analyses, the reply rate for user comments, i.e. a concept of feedback to user feedback, is regarded as a measure to analyse the enthusiasm of the propagandist. The user feedback and interactions are informative since different questions, errata identifications, useful suggestions and critical comments from users have been recorded. It is potential in improving social media publicity of a subscription account and enhancing household waste management if user comments are extracted and interpreted properly.

3. Case study

3.1. Case statement

A subscription account—namely, Shanghai Green Account—on the WeChat platform is taken as a testing ground. The collected data are from 20 May 2015 to 20 December 2019. This subscription account with over 40,000 active users is used in the great Shanghai area. Shanghai is not only an economic centre but also an innovation centre of China. In 2014, Shanghai launched a Green Account programme to help residents to participate in a new incentive plan on waste segregation and recycling. The Shanghai Green Account published posts on the WeChat platform starting from 20 May 2015. Residents were suggested to register their green accounts and conducted waste segregation and recycling with incentives. Residents can exchange daily necessities and vouchers using green points in their green accounts. In July 2019, the Green Account programme regarding waste segregation and recycling was implemented mandatorily in the great Shanghai area. More details are referred to the Green Account programme (Xiao et al., 2020) and the incentives based on green points (Jiang et al., 2020). In the case study, Excel 2016, R 3.6.3 and MATLAB 2019b are used for the data storage, pre-processing, data mining and result visualisation.

3.2. Results and discussion

This subsection measures social media publicity of the Shanghai Green Account and delivers some personalised improvements for its publicity.

3.2.1. Publishing time and its improvement

The publishing date in a week is shown in Fig. 2a. The ranking order of the number of posts is 697 (Friday), 124 (Thursday), 67 (Saturday), 53 (Monday), 37 (Wednesday), 33 (Tuesday) and 23 (Sunday). More than 65% posts (i.e. 697/1,034) were published on Friday. The purpose of posting on Friday is difficult to understand from both the waste dumping behaviour and user engagement perspectives. The big data analysis results indicated that the most frequent waste dumping occurred on Sunday rather than Friday in a pilot community in Shanghai (Jiang et al., 2020). Although the optimal post time of social media varied across markets and industries, posting on Monday to Thursday (earlier in a week) seemed to get a higher user engagement than posting on Friday (Lin et al., 2017). The combination of the Friday (Fig. 2a) and the peak time duration 17:00–18:00 (Fig. 2b) seems to be more likely the end of working time during a week. Regarding the distribution of publishing time in Fig. 2b, six Gaussian components are identified as the best combination in the Gaussian mixture model according to the minimal AIC (Table S2 in Supporting Information). Amongst the estimated six peaks, the top two peaks occur at 17:00–18:00 and 20:00–21:00. The relatively loose distribution in Fig. 2b indicates that the timing under the Shanghai Green Account does not obey the suggestion of publishing posts on a particular time (Cole, 2016). The online surfing time of users depends on their daily schedule. The publishing time should be specified as a fixed time period of the day (Ma and Chen, 2017). The Gaussian mixture distribution has a potential advantage in a further application. The mixture distribution can be matched with the optimal time distribution on the social media platform to measure the quantitative matching degree.

Regarding the improvement of publishing time, Fig. 2c indicates that the posts published on Thursday have the greatest number of views amongst a week. Fig. 2d suggests that the posts published in 12:00–13:00 and 20:00–23:00 (especially 21:00–22:00) have greater numbers of views than other time durations. These observations imply that most users view the posts when they have relatively spare time, apart from workloads and other dominated activities (e.g. entertainment on weekends). The identified golden time durations by Fig. 2c and 2d do not match with the publishing time in Fig. 2a and 2b. The above observations imply that the current publicity is propagandist-centred in the time dimension. Other evidence of the propagandist-centred publicity is discussed further in Section 3.2.2. The identified golden time durations, 12:00–13:00 and 21:00–22:00 on Thursday based on Fig. 2c and 2d, can guide to improve potential publicity effects related to household waste management.

3.2.2. Underlying topics and topic improvement

Historical posts are divided into two sets. An in-sample set of 80% posts is taken to train an LDA model. The remaining held-out set (i.e. the validation set) of 20% posts is employed for model validation. Two quantitative measures are used to show the reliability of the results. The first one—namely, perplexity—is a general quality measure for the entire model (Blei et al., 2003). Table 1 shows the validation perplexities under the number of topics from 1 to 15. The model with ten topics achieves the minimal validation perplexity of 214.9, which indicates that the identified model is more reliable than the other models with different numbers of topics. The perplexity measure was also adopted

Table 1
The validation perplexities under the different numbers of topics.

Number of topics	Validation perplexity	Number of topics	Validation perplexity	Number of topics	Validation perplexity
1	257.8	6	221.8	11	218.0
2	260.0	7	223.5	12	215.8
3	242.2	8	220.2	13	226.2
4	223.3	9	219.6	14	216.8
5	220.7	10	214.9	15	227.3

Note: The computational time and the trend of validation perplexity are presented in Figure S1 in Supporting Information.

by Sun et al. (2020) to find the most effective topic analysis model for the negative sentiments of residents. The second one—namely, topic similarity—is a quality measure to show the individual representativity of an identified topic. The cosine similarity of topic vectors was adopted to measure the topic similarity (Ramage et al., 2009). The keyword probabilities per topic generated by the LDA model are used to calculate the cosine similarity of an identified topic to the other topics. The mean cosine similarity of the ten topics is calculated as 0.076 (by averaging 0.151, 0.093, 0.090, 0.003, 0.029, 0.139, 0.111, 0.049, 0.003 and 0.093). Although sporadic keywords exit in more than one topic, the low mean cosine similarity indicates that these topics are representative and relatively independent for the posts of the Shanghai Green Account.

Fig. 3a shows the ten topics and a descending ranking order of ten topics (by corpus topic probabilities observed in published posts, i.e. 0.131, 0.123, 0.117, 0.097, 0.096, 0.095, 0.094, 0.087, 0.081 and 0.079). The LDA model has generated 698 keywords. The different font sizes of keywords in each word cloud in Fig. 3a are associated with the keyword probabilities in each topic. For the ten topics, the maximal keyword probability per topic is 0.168, 0.187, 0.093, 0.195, 0.089, 0.150, 0.103, 0.202, 0.142 and 0.093. The ten topics have particular focuses. Judged by the first four keywords per topic (Gurcan and Cagiltay, 2019), the focuses of the ten topics are household waste segregation (Topic 1), green account/point and district waste management (Topic 2), incentives related to green accounts (Topic 3), environmental protection and education (Topic 4), new stories and knowledge (Topic 5), waste segregation and industry highlights (Topic 6), people related to waste management (Topic 7), Shanghai city and its fun places/activities (Topic 8), good modes in waste management (Topic 9) and welfares and new changes of Shanghai (Topic 10).

Fig. 3b shows the topic mixtures of the ten most popular posts (1 to 10) and the ten least popular posts (11 to 20). Each post in Fig. 3b is a combination of the above ten topics. Overall, Topic 5 has larger weights in the ten most popular posts than weights in the ten least popular posts. The posts 11, 14 and 20 that focus on Topics 1 and 4 are all from the ten least popular posts. According to Section 2.3.2, the usage-degree-of-topic in Fig. 3c indicates that Topics 1 to 4 are used most frequently by propagandist in historical posts. The multivariate linear regression function between the number of views and the ten topics is calculated as Eq. (16). The regression coefficients in Eq. (16) i.e. the contribution-level-of-topic in Fig. 3c shows that Topics 2, 3 and 5 contribute more for the number of views, but Topics 4 and 8 contribute less than other topics. The mismatch between the usage-degree-of-topic and the contribution-level-of-topic, especially for Topics 4 and 5, indicates that the current publicity related to household waste management is propagandist-centred in the topic dimension. The analytics of topic contribution and underling topics offer a possibility to improve the publicity by incorporating more high-contribution topics in newly published posts. The pure preaching on environmental protection and the introduction of shanghai city and its fun places seem to be difficult to stimulate the interests of residents, compared to incentive activities and new stories and knowledge related to waste management. Such quantitative analysis results are in line with the statement and

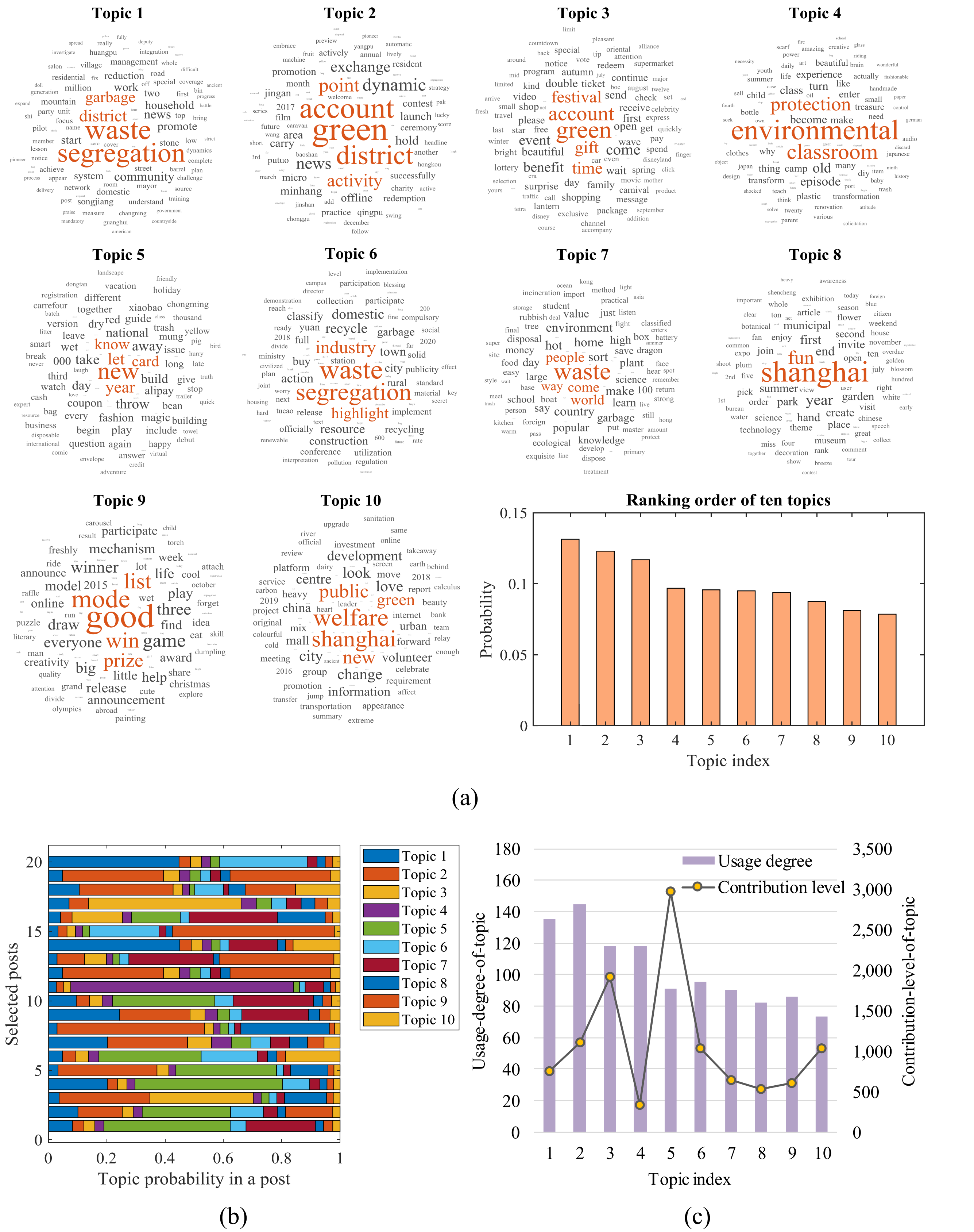

Fig. 3. Major topics, topic mixtures and topic usage/contribution. (a) the ten topics of the Shanghai Green Account and the ranking order of topics. (b) topic mixtures of the ten most popular posts (1 to 10 in the y-axis) and the ten least popular posts (11 to 20 in the y-axis) and (c) the comparison between the usage-degree-of-topic and the contribution-level-of-topic.

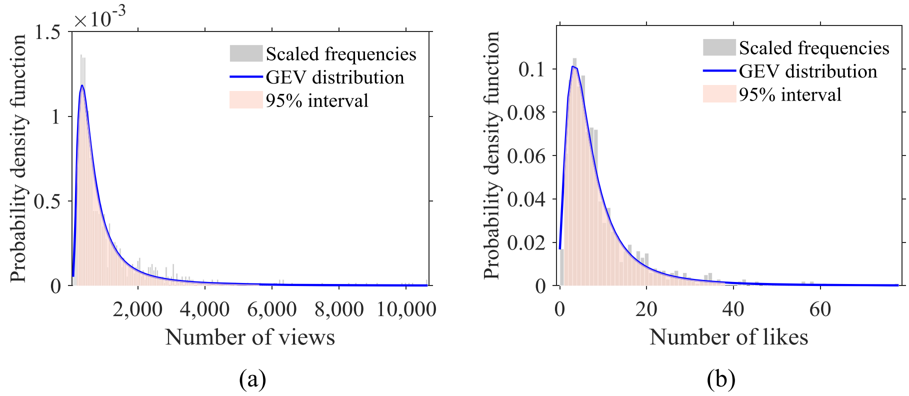

Fig. 4. The probability distributions of (a) the number of views and (b) the number of likes.

suggestion for waste segregation in Beijing (Zhao, 2017).

$$v = 749.76\zeta_1 + 1,110.27\zeta_2 + 1,919.13\zeta_3 + 336.71\zeta_4 + 2,977.32\zeta_5 + 1,035.59\zeta_6 + 637.30\zeta_7 + 526.00\zeta_8 + 601.65\zeta_9 + 1,033.42\zeta_{10} \quad (16)$$

3.2.3. Distribution of user popularity

According to the minimal AIC (i.e. 16,203.02 and 6,574.57 from the goodness-of-fit results in Table S3 and Table S4 in Supporting Information), the generalised extreme value (GEV) distribution is identified as the most suitable one for the number of views or likes amongst the 19 potential probability distributions. Both of the two fitted distributions have apparent long-tail characteristics with the sample skewness of 3.71 (Fig. 4a) and 2.52 (Fig. 4b). The propagandist of a subscription account should focus more on those posts in the long-tail parts which offer better guidance to attract user engagement. The shape, scale and location parameters of the GEV distribution for the number of views (Fig. 4a) are 0.62, 366.10 and 492.50. The average number of views is estimated at 1,282.70. The GEV parameters for the number of likes (Fig. 4b) are 0.42, 3.90 and 4.58. The average number of likes is estimated at 9.57. These distribution parameters and mean values can be used to analyse user engagement (Aldous et al., 2019). For the Shanghai Green Account, there are average 1,283 views, and ten likes for a post. Such quantitative results collectively with the skewness values and the estimated distribution parameters can be set as benchmarks for future publicity improvement on waste segregation and recycling in Shanghai.

3.2.4. Annual status and trend

The annual status and trend based on views and likes are shown in Fig. 5a and 5b. The annual number of views had a stable increase from 2015 to 2017, but it was reduced in 2018. This is basically in line with the conclusion that social media popularity peaks four years later since entry (Franses, 2015). According to the product life cycle concept, the subscription account was in its decline stage in 2018. It is interesting that the annual number of views in 2019 was elevated significantly. After the trial operation in broad pilot communities in 2018 (Jiang et al., 2020), the policy on mandatory waste segregation was implemented on 1 July 2019 in Shanghai (Xiao et al., 2020). According to the data released by the Shanghai Landscaping & City Appearance Administrative Bureau (SLCAAB, 2020), the average daily waste collections in 2019 included dry waste 17,731 t/d, biodegradable food waste 7,453 t/d (also called as wet waste in Shanghai), recyclable waste 4,049 t/d and hazardous waste 0.6 t/d. Compared to those at the end of 2018, dry waste was reduced by 17.5%; biodegradable food waste, recycling waste, and hazardous waste were elevated by 88.8%, 431.8% and 504.1%. The elevation of the annual number of views, a growth rate of 138.5% by the annual mean from 2018 to 2019, can be well explained since many residents rely on the post information on waste segregation and recycling in 2019. But the managers should be aware that this is only a passive success to refresh the product life cycle of this subscription account, which heavily benefits from the external macro policy. Under a concept of dynamic product life cycle management, the publicity innovation with interesting and useful posts should be introduced for the subscription account, once the trend transfers to a maturity or decline stage from a growing stage.

The results suggest similar overall trends on views and likes in Fig. 5a and 5b, which implies that either views or likes can be employed for the status and trend analysis. This is a useful observation since some social media platforms, e.g. Reddit, only count the number of voting (i.e. a type of likes) rather than views. Although the 'Live Story' function of Facebook can provide information on both views and likes, easily accessible public information in the subscription accounts on Facebook, e.g. @WasteManagement on Facebook (Facebook, 2020), is the number of likes.

The Shanghai Green Account started publishing posts from 20 May 2015. The numbers of total annual posts in Fig. 5c are stable from 2016 to 2019, which implies a disciplined publicity behaviour of this subscription account. Fig. 5c also shows the annual trend of the ratio of annual waste management related posts in total annual posts. The average ratio from 2015 to 2019 is 87.9%. The relatively high ratio of waste management related posts can also be observed in the subscription account @WasteManagement (Facebook, 2020). The ratio of 84.3% in 2018 has an apparent reduction compared to the other four-year average ratio of 88.8%. Through tracking the posts in 2018, amongst 36 posts without relevance to waste management, 28 posts focused on a new topic—namely, Getting Around Shanghai—which introduces tourist attractions in Shanghai. The average numbers of views and likes belonging to the 28 posts are 578 and 5.5, which are significantly lower than their mean values in 2018, as shown in Figs. 5a and 5b. The new topic on tourist attractions partially leads to the unfavourable user popularity in 2018. That is why this new topic was cancelled for publicity in 2019.

3.2.5. Relationship between views and likes

By the Spearman correlation test, the correlation coefficient between the number of views and the number of likes is 0.73 ($p < 0.001$), which suggests a strong and statistically significant correlation. By a distribution fitting, the view/like ratio obeys a GEV distribution (Fig. 6a) with the shape, scale and location parameters of 0.30, 38.63 and 75.55, indicating a heavy-tailed distribution. The scatterplot in Fig. 6b suggests a relatively linear relationship between the two variables. The ratio estimation result, 96.64, under the Bayesian linear regression in Fig. 6b has an apparent difference with the estimation, 108.39, under the least-squares linear regression. This observation demonstrates the importance of the distribution verification and model selection when the models are applied to real cases in the waste management field. For the studied case, the estimation function indicates that a basic number of views (i.e. about 105 views) activates the

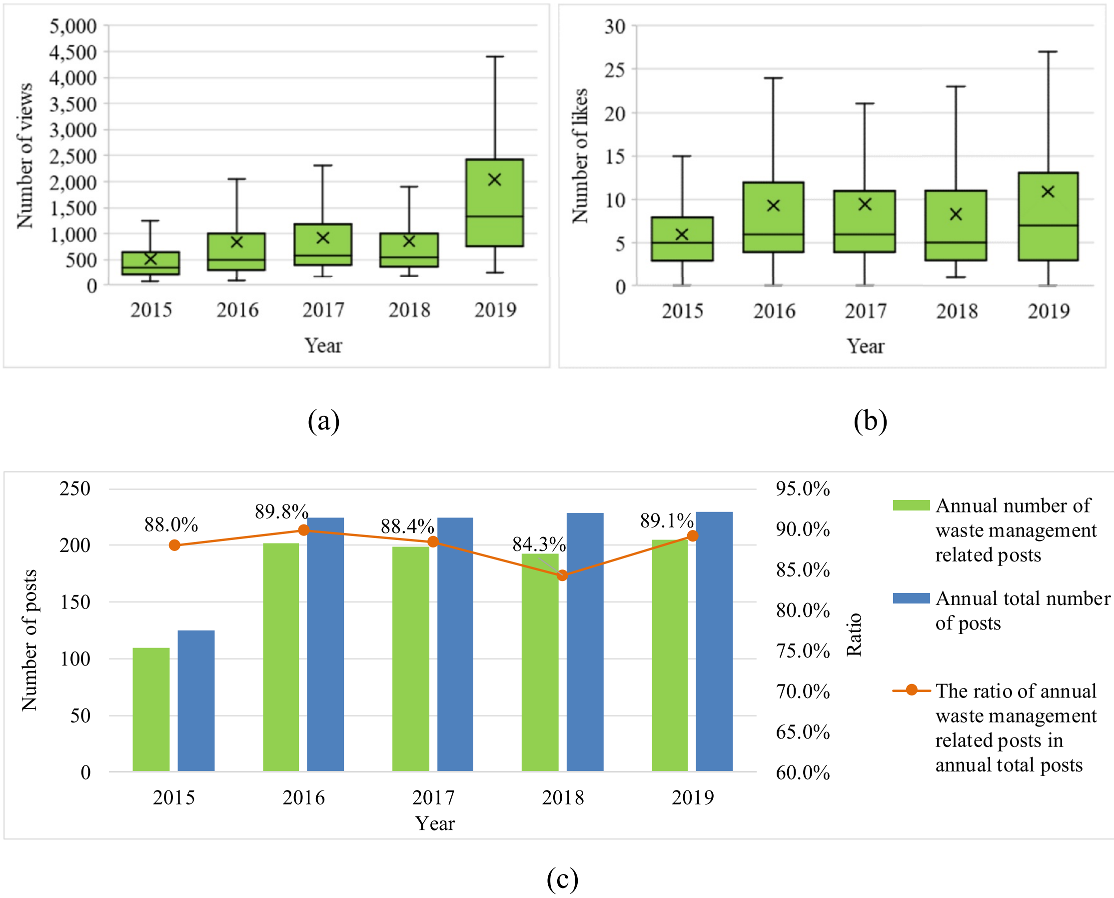

Fig. 5. Trends of (a) the number of views, (b) the number of likes and (c) the ratio of annual waste management related posts in total annual posts from 2015 to 2019. The five lines on a boxplot denote five percentiles, and the notation ' × ' denotes an annual mean.

behaviour of likes statistically, and one like per about 97 views occurs after the activation averagely. There are two kinds of applications based on the above observations. First, the ratio estimation model under the Bayesian paradigm could guarantee more robust applications on different social media platforms. Second, the generated view/like ratio could be set as a benchmark for further improvement in social media publicity and household waste segregation and recycling. More likes imply more satisfaction regarding the publicity of household waste management. Satisfaction can impact enthusiasm and participation in waste segregation positively (Wang et al., 2020).

3.2.6. *Relative publicity quality of historical posts*

According to the definition in Section 2.3.6, the proportion of posts without likes takes 1.6% (i.e. 17 amongst 1,034 posts). There are 30 outliers of the time-serial view/like ratio, as shown in Fig. 7a, which take 2.9% amongst the 1,034 posts. Collectively, the poor-quality posts take 4.5% being equivalent to a relative publicity quality (RPQ) of 0.95. The RPQ measure could have two applications. First, managers can compare the RPQ value of the subscription account of interest with RPQ values of other accounts to obtain the motivation for further quality improvements. Second, the text analytic statistics in Section 2.3.7 can

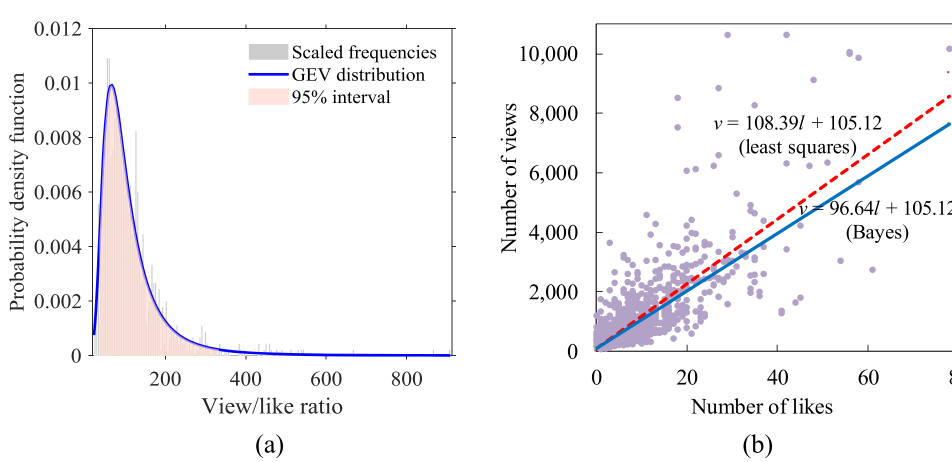

Fig. 6. The relationship between views and likes. (a) the distribution of the view/like ratio and (b) the ratio estimation based on the Bayesian linear regression (solid blue line). Note: the estimation based on the least-squares linear regression (red dashed line) is used as a comparison.

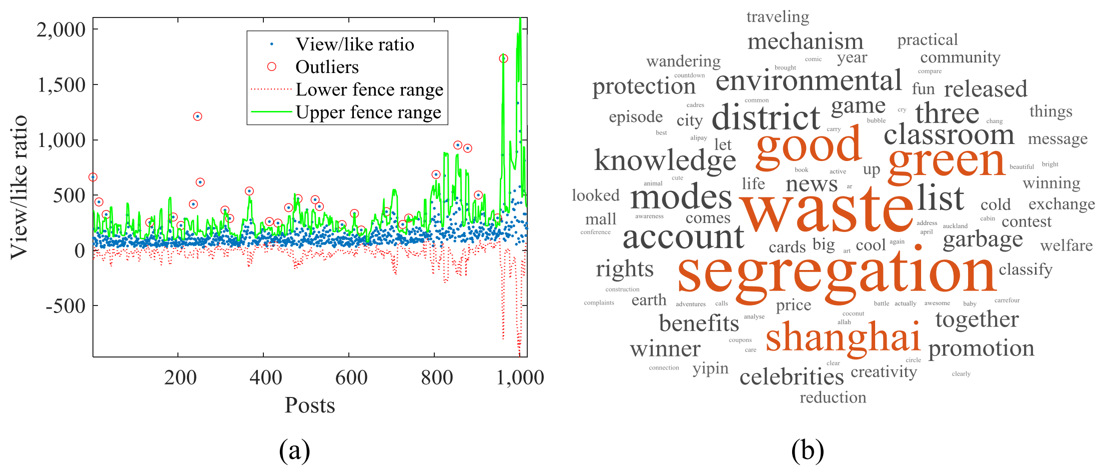

Fig. 7. (a) the detection results of outliers of the time-serial view/like ratio and (b) the keyword frequency results of these poor-quality posts.

help to figure out the keyword frequency of these poor-quality posts. The results are shown in Fig. 7b. By excluding the common keywords, i.e. 'waste segregation', 'green account' and 'Shanghai', the apparent keywords are 'good modes', 'environmental classroom', 'district' and 'knowledge'. For one thing, social media users may not have common interests in straightforward exhortation and annoying knowledge. For another, the Shanghai Green Account was designed for users in the whole city. It is better to publish posts targeting issues or news for the city-wide residents rather than the district-wide residents, let alone there are already district-level subscription accounts, such as Yangpu Green Account, Jinshan Green Account, Hongkou Green Account and Chongming Green Account. Overall, the analytics of poor-quality posts could improve the strategy adjustment for social media publicity on environmental protection and education. For example, the story presentation needs to be more attractive to avoid boring preaching. Otherwise, the media (including social) related to waste management would lose the engagement or attention of users due to the lack of attractiveness (Zorpas, 2020).

3.2.7. Feedback and interactions on the platform

By the Spearman correlation test, the correlation coefficient between the number of comments and the number of views is 0.66 ($p < 0.001$). The correlation coefficient between the number of comments and the number of likes is 0.56 ($p < 0.001$). Although the correlation test results are statistically significant, the correlation coefficients are both less than 0.70. The resulting R^2 values of curve fitting are both less than 0.40 (Figure S2 in Supporting Information). The results of the correlation analysis and curve-fitting suggest relatively loose relationships between the number of comments and the number of views (or likes). In other words, a hot topic with high user popularity does not necessarily guarantee active user comments. The keyword frequency analysis results associating with the number of comments are shown in Figure S3 in Supporting Information to discover the topics behind posts with active user comments. Overall, the results imply that the Shanghai Green Account has published quite relevant contents in the scope of its publicity tasks. The posts that receive over 15 comments (Figures S3a and S3b in Supporting Information) more likely link to the incentive/reward/promotion programmes after excluding those common keywords, i.e. 'waste segregation', 'green account' and 'Shanghai'. The results in Figures S3a and S3b imply that incentives promote interactions between users and household waste managers on the social media platform, and the welfare/gifts exchange has more promotion effects than virtual points in accounts of residents. These observations could guide better interactions between users and managers, thereby elevating the user participation in household waste management.

Regarding the feedback to user feedback. From 20 May 2015 to 20 December 2019, a total of 209 amongst 1,207 user comments have been replied by the propagandist. The reply rate is 17.3%. Although a low replying rate is common for a subscription account since some comments are repeated questions or not questions at all, some unique questions and suggestions are not replied. For example, the question about how to register and log into green accounts and the question regarding what kind of waste (hazardous or recyclable) is the fluorescent lamp? Since satisfaction would impact participation and enthusiasm in waste segregation positively (Wang et al., 2020), the unsuccessful interactions might reduce the enthusiasm and engagement of residents in community waste management.

Regarding the contents of user comments. Massive group wisdom in user engagement can promote household waste segregation and recycling. Some examples from user comments are listed as follows.

(i) Questions: 'Considering that the materials of Tetra Pak include plastic, aluminium and paper, which category should the Tetra Pak be thrown for recycling?'. The answer from the subscription account is 'the paper category because of the high paper content in Tetra Pak'. After implementing the mandatory waste segregation in Shanghai on 1 July 2019, more plastic bags were used for packing different categories of waste. For the post 'Restriction or ban for plastic?' on 06 September 2019, the question from a user is 'When will the society develop environmentally friendly alternative materials to forbid and replace plastics?'. This might be an open question without conclusive answers in the current phase. Klemeš et al. (2020b) suggested plastics are not natural foes and what we need to do currently is to promote the '10Rs programme' for a circular economy. Although plastics have been providing convenience for goods packing and assisting in fighting COVID-19 pandemic, more attention should be paid to 'Plastic Waste Footprint' worldwide (Klemeš et al., 2020c).

(ii) Errata in posts: In a post on 4 February 2019, the colour patterns of the waste segregation bins were found not to meet the latest colour standard. On 17 October 2019, some users found an error hyperlink in the post. Some active users even pointed out typos in posts. Fortunately, the propagandist replied to the comments and modified the errata promptly, which demonstrates the unique advantage in terms of interactions by social media publicity compared to traditional one-way media publicity, such as newsletters, posters and magazines.

(iii) Suggestions: Some users suggested replacing the offline exchange of green points by online exchange manners to make the incentives more environmentally friendly. Some users claimed that the members of mobile video applications attracted them the most than daily necessities. The users' feedback was implemented several months later, and the exchanges of video members were successful online on the Green Account website.

(iv) Critical comments: After implementing the new policy, only one dumping facility is left in partial communities with more than 500 households. The waste dumping time windows are not rational for some residents. Although hazardous waste bins are full in some communities, residents still have to wait until the designated day for hazardous waste collection and transportation. User comments have raised many useful suggestions and practical issues in daily life, which help to enhance household waste segregation and recycling management significantly.

4. Conclusions

This study has proposed an integrated framework with statistical and machine learning techniques to measure and improve social media publicity related to household waste management. A total of seven inter-related measures and analytics approaches have been applied to understand the behaviours or patterns of the propagandist and users of a subscription account in social media. A specific case study of the Shanghai Green Account, on the WeChat platform, has demonstrated the potency of the framework and approaches in revealing the useful information for personalised improvements of household waste management. The framework and approaches could be easily replicated to publicity analytics for other subscription accounts on different social media platforms, such as Facebook, YouTube, WeChat, Instagram and Twitter.

The main observation of this study is that, if decision-makers are not supported by data analytics on online footprints of propagandist and users, social media publicity related to household waste management would likely be propagandist-centred in both timing and topic dimensions. From the perspective of publishing time, posts were published when the time is more comfortable for the propagandist, rather than when most users can read the posts. From the perspective of published topics, posts stressed more on the usefulness and knowledge of published topics but lacked the balance regarding how to attract users simultaneously. Although social media publicity should not be totally user-centred to keep the diversity of published contents, the propagandist could improve the story presentation as suggested, especially for environmental education/knowledge related topics.

The main quantitative results are:

i) With the mandatory implementation of waste segregation and recycling in Shanghai, compared to those in 2018, the average daily dry waste in 2019 was reduced by 17.5%, biodegradable food waste and recyclable waste were elevated by 88.8% and 431.8%, and the average number of views in the Shanghai Green Account was elevated by 138.5%.
ii) In 2015, 2016, 2017 and 2019, the average ratio of waste management related posts in total annual posts was 88.8%. The smallest annual ratio, 84.3%, occurred in 2018 when the user popularity on the subscription account was also observed to be weak.
iii) The historical publishing time has a loose distribution with six Gaussian components. The identified golden time durations with most views are 12:00–13:00 and 21:00–22:00 on Thursday.
iv) Ten underlying topics are identified for the publicity of waste segregation and recycling in the Shanghai Green Account. The new stories/knowledge and the incentives related to green accounts contribute the most for the post view. In contrast, the environmental protection/education contributes the least for it.
v) The number of views and the number of likes show a statistically significant linear relationship (Pearson coefficient > 0.7 and $p < 0.001$). The number of likes, which can reflect the satisfaction (awareness) and engagement (willingness), is still having a large room for improvement where the current status is one like per average 97 views after a basic view activation.
vi) Amongst the historical posts, there are 17 posts without likes and 30 posts that are outliers of time-serial view/like ratio. They result in the relative publicity quality of 0.95 collectively.
vii) Based on the keyword frequency analysis, the engaging posts (e.g. with more than 15 comments) are identified as having a close relationship with promotion or reward programmes (e.g. keywords of 'points', 'exchange' and 'welfare').

Several directions on the social media publicity analytics related to household waste management are worthy of being investigated further considering this is a relatively new area under the background of IoT and digitalisation in waste management.

a) When social media posts are not proposed and designed in a qualified manner, they could be even contra-productive. The adverse effects of some posts on household waste segregation and recycling are worthy of being assessed systematically.
b) The integrated framework is also applicable to the other platforms such as Facebook, Twitter and Instagram, which have different user bases in worldwide waste management systems. Future work can be conducted to compare the differences in results from the spatial and demographic aspects.
c) The indirect influence of publicity on behaviour was identified by Wang et al. (2018). It is beneficial to rethink the awareness promotion mechanism based on the improved outputs of social media publicity related to household waste management.
d) Social media may not replicate 'face to face' of social influence (Young et al., 2017) on the publicity related to household waste management. The information in user comments on social media platforms could provide enough insights into different types of media (Park et al., 2019), including the 'face to face' communication.
e) The publicity analytics deserves investigation in waste management before, during and after the COVID-19 pandemic, considering that environmental footprints (Klemeš et al., 2020a) and waste generation behaviours (Fan et al., 2020) have changed significantly during this crisis period.

CRediT authorship contribution statement

Peng Jiang: Conceptualization, Data curation, Methodology, Formal analysis, Software, Visualization, Writing - original draft, Writing - review & editing. **Yee Van Fan:** Conceptualization, Formal analysis, Writing - original draft, Writing - review & editing. **Jiří Jaromír Klemeš:** Conceptualization, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is partially supported by the National Natural Science Foundation of China under the programme: AI + Waste to Resources Management toward Sustainable City. The EU supported project Sustainable Process Integration Laboratory –SPIL funded as project No. CZ.02.1.01/0.0/0.0/15_003/0000456, by Czech Republic Operational Programme Research and Development, Education, Priority 1: Strengthening capacity for quality research is also gratefully acknowledged.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.resconrec.2020.105146.

References

Agrawal, A., Fu, W., Menzies, T., 2018. What is wrong with topic modeling? and how to fix it using search-based software engineering. Inf. Softw. Technol. 98, 74–88.
Akaike, H., 2011. Akaike's information criterion. Int. Encycl. Stat. Sci 25-25.
Aldous, K.K., An, J., Jansen, B.J., 2019. View, like, comment, post: analyzing user engagement by topic at 4 levels across 5 social media platforms for 53 news organizations. In: Proceedings of the International AAAI Conference on Web and Social Media, pp. 47–57.
Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent Dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022.
Cole, K., 2016. Communications: social media and the waste management sector: eight strategies to use social media more effectively. Waste Adv. Mag. https://wasteadvantagemag.com/communications-social-media-and-the-waste-management-sector-eight-strategies-to-use-social-media-more-effectively Accessed 12.06.2020.
Cui, R., Gallino, S., Moreno, A., Zhang, D.J., 2018. The operational value of social media information. Prod. Oper. Manag. 27 (10), 1749–1769.
DEEP, 2020. Using Social Media to Promote Recycling. Department of Energy and Environmental Protection (DEEP). https://portal.ct.gov/DEEP/Reduce-Reuse-Recycle/Recycling/Using-Social-Media-to-Promote-Recycling Accessed 12.06.2020.
Desgagné, A., Gagnon, P., 2019. Bayesian robustness to outliers in linear regression and ratio estimation. Braz. J. Probabi. Stat. 33 (2), 205–221.
Doane, D.P., Seward, L.E., 2011. Measuring skewness: a forgotten statistic? J. Stat. Educ. 19 (2), 1–18.
Eriksen, M.K., Pivnenko, K., Olsson, M.E., Astrup, T.F., 2018. Contamination in plastic recycling: influence of metals on the quality of reprocessed plastic. Waste Manag. 79, 595–606.
Facebook, 2020. Waste Management @WasteManagement. https://www.facebook.com/WasteManagement Accessed 22.08.2020.
Fan, Y.V., Jiang, P., Hemzal, M., Klemeš, J.J., 2020. An update of COVID-19 influence on waste management. Sci. Total Environ. 142014. https://doi.org/10.1016/j.scitotenv.2020.142014.
Franses, P.H., 2015. The life cycle of social media. Appl. Econ. Lett. 22 (10), 796–800.
Frempong, J., Chai, J., Ampaw, E.M., Amofah, D.O., Ansong, K.W., 2020. The relationship among customer operant resources, online value co-creation and electronic-word-of-mouth in solid waste management marketing. J. Clean. Prod. 248, 119228.
Frigge, M., Hoaglin, D.C., Iglewicz, B., 1989. Some implementations of the boxplot. Am. Stat. 43 (1), 50–54.
Gaenssle, S., Budzinski, O., 2020. Stars in social media: new light through old windows? J. Media Bus. Stud. https://doi.org/10.1080/16522354.2020.1738694.
Gu, B., Wang, H., Chen, Z., Jiang, S., Zhu, W., Liu, M., Chen, Y., Wu, Y., He, S., Cheng, R., 2015. Characterization, quantification and management of household solid waste: a case study in China. Resour. Conserv. Recycl. 98, 67–75.
Gurcan, F., Cagiltay, N.E., 2019. Big data software engineering: analysis of knowledge domains and skill sets using LDA-based topic modeling. IEEE Access 7, 82541–82552.
Hu, K.-C., Lu, M., Huang, F.-Y., Jen, W., 2017. Click "like" on Facebook: the effect of customer-to-customer interaction on customer voluntary performance for social networking sites. Int. J. Hum.–Comput. Interact. 33 (2), 135–142.
Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., Zhao, L., 2019. Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey. Multimed. Tools Appl. 78 (11), 15169–15211.
Jiang, P., Fan, Y.V., Zhou, J., Zheng, M., Liu, X., Klemeš, J.J., 2020. Data-driven analytical framework for waste-dumping behaviour analysis to facilitate policy regulations. Waste Manag. 103, 285–295.
Jiang, P., Liu, X., 2016. Hidden Markov model for municipal waste generation forecasting under uncertainties. Eur. J. Oper. Res. 250 (2), 639–651.
Jiang, P., Liu, X., Shoemaker, C.A., 2017. An adaptive particle swarm algorithm for unconstrained global optimization of multimodal functions. In: Proceedings of the 9th ACM International Conference on Machine Learning and Computing (ICMLC). Singapore. ACM, pp. 221–226.
Jiang, T., Guo, Q., Chen, S., Yang, J., 2019. What prompts users to click on news headlines? Evidence from unobtrusive data analysis. Aslib J. Inf. Manag. 72 (1), 49–66.
Jing, F., Wei, L., 2016. A study on influential factors of WeChat public accounts information transmission hotness (In Chinese). J. Intell. 35 (2), 157–162.
Klemeš, J.J., Fan, Y.V., Jiang, P., 2020a. The energy and environmental footprints of COVID-19 fighting measures – PPE, disinfection, supply chains. Energy. https://doi.org/10.1016/j.energy.2020.118701.
Klemeš, J.J., Fan, Y.V., Jiang, P., 2020b. Plastics: friends or foes? The circularity and plastic waste footprint. Energy Sources Part A. https://doi.org/10.1080/15567036.2020.1801906.
Klemeš, J.J., Fan, Y.V., Tan, R.R., Jiang, P., 2020c. Minimising the present and future plastic waste, energy and environmental footprints related to COVID-19. Renew. Sustain. Energy Rev. 127, 109883.
Knickmeyer, D., 2019. Social factors influencing household waste separation: a literature review on good practices to improve the recycling performance of urban areas. J. Clean. Prod., 118605.
Krounbi, L., Enders, A., van Es, H., Woolf, D., van Herzen, B., Lehmann, J., 2019. Biological and thermochemical conversion of human solid waste to soil amendments. Waste Manag. 89, 366–378.
Ksiazek, T.B., Peer, L., Lessard, K., 2016. User engagement with online news: conceptualizing interactivity and exploring the relationship between online news videos and user comments. New Media Soc. 18 (3), 502–520.
Lin, H.-C., Swarna, H., Bruning, P.F., 2017. Taking a global view on brand post popularity: six social media brand post practices for global markets. Bus. Horiz. 60 (5), 621–633.
Ma, Y., Chen, J., 2017. Effects of Platform and Content Attributes on Information Dissemination on We Media: a Case Study on WeChat Platform. In: Wuhan International Conference on e-Business (WHICEB), pp. 64.
Mallick, R., Bajpai, S.P., 2019. Impact of social media on environmental awareness, environmental awareness and the role of social media. IGI Global 140–149.
Martin-Rios, C., Demen-Meier, C., Gössling, S., Cornuz, C., 2018. Food waste management innovations in the foodservice industry. Waste Manag. 79, 196–206.
Mathworks, 2019. Detect Outliers Using Quantile Regression. MathWorks, Inc., USA. www.mathworks.com/help/stats/outlier-detection-using-quantile-regression.html#responsive_offcanvas Accessed 12.06.2020.
Mathworks, 2020. Choose Number of Topics for LDA Model. MathWorks, Inc., USA. https://www.mathworks.com/help/textanalytics/ug/choose-number-of-topics-for-LDA-model.html Accessed 12.06.2020.
Mathworks, 2020. Bag-of-n-grams Model. MathWorks, Inc., USA. https://www.mathworks.com/help/textanalytics/ref/bagofngrams.html Accessed 12.06.2020.
Meinshausen, N., 2006. Quantile regression forests. J. Mach. Learn. Res. 7, 983–999.
Mintz, K.K., Henn, L., Park, J., Kurman, J., 2019. What predicts household waste management behaviors? Culture and type of behavior as moderators. Resources. Conserv. Recycl. 145 (145), 11–18.
Myers, J.L., Well, A., Lorch, R.F., 2010. Research Design and Statistical Analysis. HarperCollins, New York.
Park, J., Ahn, C., Lee, K., Choi, W., Song, H.T., Choi, S.O., Han, S.W., 2019. Analysis on public perception, user-satisfaction, and publicity for WEEE collecting system in South Korea: a case study for Door-to-Door Service. Resour. Conserv. Recycl. 144, 90–99.
Porteous, I., Newman, D., Ihler, A., Asuncion, A., Smyth, P., Welling, M., 2008. Fast collapsed Gibbs sampling for latent Dirichlet allocation. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 569–577.
Ramage, D., Hall, D., Nallapati, R., Manning, C.D., 2009. Labeled LDA: a supervised topic model for credit attribution in multi-labeled corpora. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pp. 248–256.
Ramzan, S., Liu, C., Munir, H., Xu, Y., 2019. Assessing young consumers' awareness and participation in sustainable e-waste management practices: a survey study in Northwest China. Environ. Sci. Pollut. Res. 26 (19), 20003–20013.
Sarc, R., Curtis, A., Kandlbauer, L., Khodier, K., Lorber, K., Pomberger, R., 2019. Digitalisation and intelligent robotics in value chain of circular economy oriented waste management–a review. Waste Manag. 95, 476–492.
Simon, K., 2019. Global Digital 2019 Reports. https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates Accessed 12.06.2020.
SLCAAB, 2020. Shanghai Landscaping & City Appearance Administrative Bureau (SLCAAB). http://lhsr.sh.gov.cn/sites/ShanghaiGreen/dyn/ViewIndex.ashx Accessed 12.06.2020.
Sujata, M., Khor, K.-S., Ramayah, T., Teoh, A.P., 2019. The role of social media on recycling behaviour. Sustain. Prod. Consum. 20, 365–374.
Sun, Y., Wang, Z., Zhang, B., Zhao, W., Xu, F., Liu, J., Wang, B., 2020. Residents' sentiments towards electricity price policy: evidence from text mining in social media. Resour. Conserv. Recycl. 160, 104903.
Wamuyu, P.K., 2018. Leveraging Web 2.0 technologies to foster collective civic environmental initiatives among low-income urban communities. Comput. Human. Behav. 85, 1–14.
Wang, Q., Long, X., Li, L., Kong, L., Zhu, X., Liang, H., 2020. Engagement factors for waste sorting in China: the mediating effect of satisfaction. J. Clean. Prod. 267, 122046.
Wang, Z., Guo, D., Wang, X., Zhang, B., Wang, B., 2018. How does information publicity influence residents' behaviour intentions around e-waste recycling? Resour. Conserv. Recycl. 133, 1–9.
WMR, 2015. The Last Word: does social media have a place in improving bin behaviour? Waste Manag. Review (WMR). https://wastemanagementreview.com.au/the-last-word-does-social-media-have-a-place-in-improving-bin-behaviour Accessed 12.06.2020.
Xiao, S., Dong, H., Geng, Y., Francisco, M.-J., Pan, H., Wu, F., 2020. An overview of the municipal solid waste management modes and innovations in Shanghai, China. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-020-09398-5.
Young, W., Russell, S.V., Robinson, C.A., Barkemeyer, R., 2017. Can social media be a tool for reducing consumers' food waste? A behaviour change experiment by a UK retailer. Resour. Conserv. Recycl. 117, 195–203.
Zamri, G.B., Azizal, N.K.A., Nakamura, S., Okada, K., Nordin, N.H., Othman, N., Akhir, F.N.M., Sobian, A., Kaida, N., Hara, H., 2019. Delivery, impact and approach of household food waste reduction campaigns. J. Clean. Prod., 118969.
Zhang, J., Yan, J., Infield, D., Liu, Y., Lien, F.-S., 2019. Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model. Appl. Energy 241, 229–244.
Zhao, O., 2017. Guiding Waste Segregation By Incentives. Beijing Morning PostAccessed 30.07.2020. http://opinion.people.com.cn/n1/2017/0209/c1003-29068217.html.
Zorpas, A.A., 2020. Strategy development in the framework of waste management. Sci. Total Environ. 716, 137088.
Zuo, L., Wang, C., Sun, Q., 2020. Sustaining WEEE collection business in China: the case of online to offline (O2O) development strategies. Waste Manag. 101, 222–230.