

Article

# Emojis as Contextual Indicants in Location-Based Social Media Posts

Eva Hauthal , Alexander Dunkel  and Dirk Burghardt 

Institute of Cartography, Technische Universität Dresden, 01069 Dresden, Germany;  
alexander.dunkel@tu-dresden.de (A.D.); dirk.burghardt@tu-dresden.de (D.B.)

\* Correspondence: eva.hauthal@tu-dresden.de

**Abstract:** The presented study aims to investigate the relationship between the use of emojis in location-based social media and the location of the corresponding post in terms of perceived objects and conducted activities connected to this place. The basis for this is not a purely frequency-based assessment, but a specifically introduced measure called typicality. To evaluate the typicality measure and examine the assumption that emojis are contextual indicants, a dataset of worldwide geotagged posts from Instagram relating to sunset and sunrise events is used, converted to a privacy-aware version based on a Hyperloglog approach. Results suggest that emojis can often provide more nuanced information about user activities and the surrounding environment than is possible with hashtags. Thus, emojis may be suitable for identifying less obvious characteristics and the sense of a place. Emojis are already explored in research, but mainly for sentiment analysis, for semantic studies or as part of emoji prediction. In contrast, this work provides novel insights into the user's spatial or activity context by applying the typicality measure and therefore considers emojis contextual indicants.



**Citation:** Hauthal, E.; Dunkel, A.; Burghardt, D. Emojis as Contextual Indicants in Location-Based Social Media Posts. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 407. <https://doi.org/10.3390/ijgi10060407>

Academic Editors: Wolfgang Kainz, Arianna D'Ulizia, Patrizia Grifoni and Fernando Ferri

Received: 14 April 2021  
Accepted: 10 June 2021  
Published: 12 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** emojis; location-based social media; context

## 1. Introduction

In a way, social media is an ambivalent source of data. It comes with unique strengths, but also entails many challenges. It is a relatively new and extensive data source that can be seen as a supplement to official data. The data are available at a high temporal and partly also spatial resolution, often refer to current events, and are semantically very rich, especially as they contain subjective information. However, this data source is methodologically challenging as it involves very large amounts of data, so that extracting information from it requires a great effort in processing, especially since the quality, consistency and trustworthiness are neither consistent nor equivalent to those of official data. Likewise, the motivation of users to provide these data varies greatly. For this and other reasons, information obtained from social media can by no means be considered representative [1,2], not least because social media is not used by all demographic groups [3]. Another critical issue is data protection and privacy. Nevertheless, these data have the unbeatable advantage of containing subjective and user-related information that would be impossible to collect in this magnitude by more conventional survey methods.

We are interested in social media data that are geo-referenced and contain emojis, i.e., these seemingly playful ideograms and pictograms that are used especially in electronic messages. We aim to investigate the relationship between the use of emojis and the location of the corresponding social media post in terms of conducted activities and perceived objects connected to this place. The topic of sunrise and sunset was chosen as the thematic framework for this investigation, since these are natural events that recur daily all over the world, are commonly perceived and appreciated and thus are to be found in social media. The information content and meaningfulness of certain emojis will be examined on a global level as well as for Italy at a national level in order to work out differences in this respect.

We see a lot of potential in emojis. On the one hand, because the number of available emojis [4] and their use is increasing and communication with emojis is changing [5], and on the other hand, because we see emojis as language-independent [6,7] indicators of a user's context that can help circumvent error-prone language processing. To investigate this, the spatial component of location-based social media is included, as this has been a gap in existing emoji research so far. In addition, a brief comparison with the information content of hashtags is given to underpin the use of emojis.

## 2. State of the Art

Computer-mediated communication is increasingly permeating most areas of life. In the case of textual communication, however, this entails the absence of non-verbal and paralinguistic cues, such as countenances, body gestures, intonations and speaking speed, which are essential when communicating face-to-face [8,9]. A remedy in written communication can be, for example capital letters as an indicator for shouting, several exclamation marks as a sign for excitement, or symbols replacing facial expressions [10]. The latter ones were initially called emoticons, a sequence of ASCII characters drawing pictures such as “:-)” or “:-(“ . However, emoticons are increasingly being replaced by emojis [11], which are pictorial symbols included in the Unicode characters list since 2010 and that represent not only emotional states, but also people, animals, objects, activities and places, and therefore enable us to use them beyond expressing emotions.

Roughly summarized, emojis fulfil two functions: an emotional and a semantic function [5]. Accordingly, they are utilized for research in various fields. Due to their function as emotional signals [6,7] adding a sentimental information to a written text, sentiment analysis or emotion identification is one application field within computer science [7,9,12–17], which, among other things, led to the result that the utilization of emojis helps to improve sentiment scores [18,19]. Other research questions related to emojis are of semantic origin and explore their meaning and use [14,20–22]. Studies with a focus on communication examine how emojis are interpreted, what they convey and what influencing factors need to be taken into account here [11,23–27]. A technology intended to support the use of emojis is emoji recommendation [28–31] which intends to suggest suitable emojis to the user when typing messages. Apart from the already mentioned fields, emojis are also utilized for research in marketing, behavioral science, psychology, medicine and education [5].

Since emojis occur in large numbers in the textual part of social media posts, they might come along with a spatial reference. This has only been used very sporadically in research so far, particularly for investigating differences in meaning and usage between languages or cultures [32–34]. Moreover, it turns out that emoji use gives a fairly realistic picture of living conditions in different parts of the world [35]. Another study connects emojis and hashtags in order to detect the attitude towards a political event in consideration of the location, compares this with actual election results for the regions concerned, and can identify correlations [36].

What has not been studied up to now is how far emojis provide information about the location-based context of users. This can be characterized by activities that are associated with the current location or by objects that are perceived in the physical environment. It can be assumed that emojis also reflect this because neither words nor emojis are chosen randomly [37]. When being used in combination with images (like on Instagram), emojis provide additional contextual information about what is shown in an image [30]. Additionally, detached from an image, emojis do not only express emotional states or alter the tone of a text; they also serve as decoration, as a replacement for lexical units, and deliver information not explicitly stated in the text [37–40]. Thus, already a single emoji can increase the expressiveness of a text [18]. When mentioning merely a city name in a social media post, this gives no further information. However, an emoji can enrich the post with contextual information, such as about the feelings, activities or perceptions of the

respective person related to the current location, without requiring language processing. Therefore, we argue that emojis can be considered contextual indicants.

A possible pitfall might be the fact that the semantic interpretation and thus the usage of emojis varies between cultures and languages, but also between viewing platforms, as emojis are rendered differently [25,27]. However, the connotative meaning of an emoji is ambiguous, but the denotative one hardly is [38], which is the meaning of interest for the presented study.

### 3. Materials and Methods

#### 3.1. Dataset

Instagram is a very popular social network, and is a mixture of a microblog and an audio–visual platform. A geotag function is available. In June 2018, Instagram reached the 1 billion active users mark and the number has been growing ever since; Refs. [41,42] state that, in comparison to other social media platforms, Instagram digs deeper into emotions and has become a medium to share life experiences in a creative way.

Instagram was chosen for this study because emojis are widely used here. For the period August to December 2017 and partially January 2018, all geo-referenced posts worldwide were selected that contain terms for sunrise and sunset in English, German, French and Dutch. Therefore, we have considered different synonyms for both sunset and sunrise within each language as well as various variant forms of spelling. This gave us a total of 11 million posts. That total dataset was divided into two sub-datasets relating to sunset and sunrise. The distinction between sunset and sunrise was only made on the basis of terminology. If terms for both sunset and sunrise appear in one post, this post was assigned to both sub-datasets. Although there is a strong correspondence between the post's time of day and the sun event addressed, the timestamp was not taken into account for the formation of sub-datasets, since it indicates the time of the post publication and not the time the photo was taken. When extracting emojis from all these posts, differentiations by gender and skin tone were disregarded, as they have no effect on the intended study.

In order to ensure a privacy-aware use of this dataset, an integrated and component-based approach was used [43], based on the data abstraction format HyperLogLog (HLL), first described by [44]. In summary, during data retrieval, emojis are extracted from posts and quantitative measurements (post counts, user counts) are stored as approximate HLL sets for each distinct emoji. This results in an estimation of results rather than in exact measurements, preventing identification of individual IDs from the resulting dataset. The resulting dataset is therefore considered statistical data, in contrast to more sensitive personal or pseudonymized data. A consequence is that measurements feature a consistent error rate between 3 and 5%. In combination with existing approaches, HLL can improve privacy for quantitative studies, such as the one presented herein, while still providing some degree of flexibility for the analysis process.

#### 3.2. Calculating Typicality

In the presented study, the occurrence of emojis is of central importance. Since emojis are only available in limited numbers and therefore cannot have nearly the same diversity as words, neither the absolute nor the relative frequency of their occurrence provides meaningful results. The web app emojiTracker [45] shows the real-time use of emojis on Twitter. In our experience, the emojis that this app shows as most used in general are also among the most used in particular. Therefore, a formula was developed that combines two relative frequencies and can determine typical emojis.

For the application of the formula, at least one sub-dataset out of a total dataset is required, and any occurrence can be subject of the formula (i.e., not only emojis, but also terms, hashtags, etc.). The calculated value is termed *typicality* in the following, which is a noun derived from the adjective *typical* and describes “the state of being that is typical” [46].

The two relative frequencies on which the calculation is based are the relative frequencies of an emoji or similar in both the sub-dataset and the total dataset. The relative

frequency ( $f$ ) puts the absolute number of, for example, a specific emoji ( $n$ ) in relation to the absolute number of all emojis ( $N$ ).

$$f = \frac{n}{N} \quad (1)$$

For the calculation of typicality, the relative frequency in the total dataset ( $f_t$ ) is subtracted from the one in the sub-dataset ( $f_s$ ), and this interim result is divided by the relative frequency in the total dataset ( $f_t$ ):

$$t = \frac{\frac{n_s}{N_s} - \frac{n_t}{N_t}}{\frac{n_t}{N_t}} = \frac{f_s - f_t}{f_t} \quad (2)$$

The resulting typicality ( $t$ ) is a proportional number that indicates whether an occurrence is typical (positive value) or atypical (negative value) for a sub-dataset, compared to the total dataset. The more distinct the value, the more typical or atypical the occurrence. The division by the relative frequency in the total dataset is carried out for reasons of normalization. Without this division, only the difference between the two relative frequencies is formed, which provides less significant results regarding typicality than the normalized calculation. This will be briefly addressed again in a later section. There are also other approaches to measure normalization and relative differences, such as tf-idf or chi-squared test, but their calculation is comparatively complex and partly bound to preconditions. In this respect, typicality is an uncomplicated and easily comprehensible measure, which will be evaluated in the following in the form of a case study.

## 4. Global Analysis

### 4.1. Comparing Hashtags and Emojis

The formula for calculating typicality introduced in the previous chapter has been applied to both hashtags and emojis in the two sub-datasets described beforehand, one containing georeferenced Instagram posts related to sunset, the other one related to sunrise. The total dataset is the aggregate of both sub-datasets. Before calculating typicality, a preselection has been made: the 100 most frequently occurring hashtags and emojis were selected from both sub-datasets. After calculating the typicality for each of these preselected hashtags and emojis, only those with the 20 highest positive values in each sub-dataset are considered for the following comparison within this sub-chapter, as they are the most typical ones.

First, a comparison of the hashtags follows (see Figure 1). The purely frequency-based top 20 of both sub-datasets have a correspondence of 80% and are not particularly specific, i.e., both frequency-based top 20s merely reflect what the main topic of the post is and that it is a natural phenomenon, which was photographed. The two top 20s determined as typical with the help of the typicality formula already provide much more information about the context. That is, that the observation of sunrise takes place in the early morning, for example, after waking up or during breakfast, whereas the observation of sunset apparently often happens in the evening or night during social activities, which is indicated by the hashtags *#food* and *#funny*.

The same applies to the frequency-based top 20 emojis from both sub-datasets (see Figure 2). These correspond to 90% and likewise reflect the phenomenon itself, and the photographing, but also (and this is different from the hashtags) the joy about it. The top 20 emojis calculated as typical are even more specific and meaningful than those of the hashtags. Nature plays a greater role for sunrise than for sunset, as mountain hikes are often started early for weather reasons and thus the sunrise can be observed during that. Likewise, joggers seem to witness sunrise or early risers while drinking coffee. Sunset, on the other hand, seems to be perceived more in urban settings, at the beach (perhaps during a holiday, which means the sunset is perceived as more special than in everyday life) or at





In order to come back to the development of the typicality formula, it should be mentioned at this point that the mere subtraction of the relative frequency in the total dataset from the one in the sub-dataset would have emphasized the respective sun emojis the most for both sunrise and sunset. Only through division by the relative frequency in the sub-dataset could the presented results be achieved, which not only refer to the event itself, but also to the associated context, which is of interest for this study and is supposed to be extracted.

The previous explanations show, on the one hand, that with the help of the introduced typicality formula, typical emojis or terms for a sub-dataset can be identified. On the other hand, even after calculating this typicality, hashtags provide less detailed information than emojis about the context of an event users give hints about in their social media posts. This already promises a confirmation of the pre-assumption that emojis can be considered contextual indicants, which is why hashtags are disregarded in the following.

#### 4.2. Location-Specific Emojis

All emojis calculated as typical for the two sub-datasets were perused and some were attributed as location-specific, based on the consideration of whether an emoji shows an object from the physical environment where the sunrise or sunset is perceived, and whether it would be chosen for a social media post depending on the location. For example, the rainbow or sun emoji would be chosen regardless of whether the user is in the city, in the mountains or at the beach, but the cityscape, water wave, palm tree or mountain emoji would most likely not. Figure 3 shows the selection made and also indicates the sub-dataset for which each emoji was calculated as typical, whereas only one emoji is typical for both.



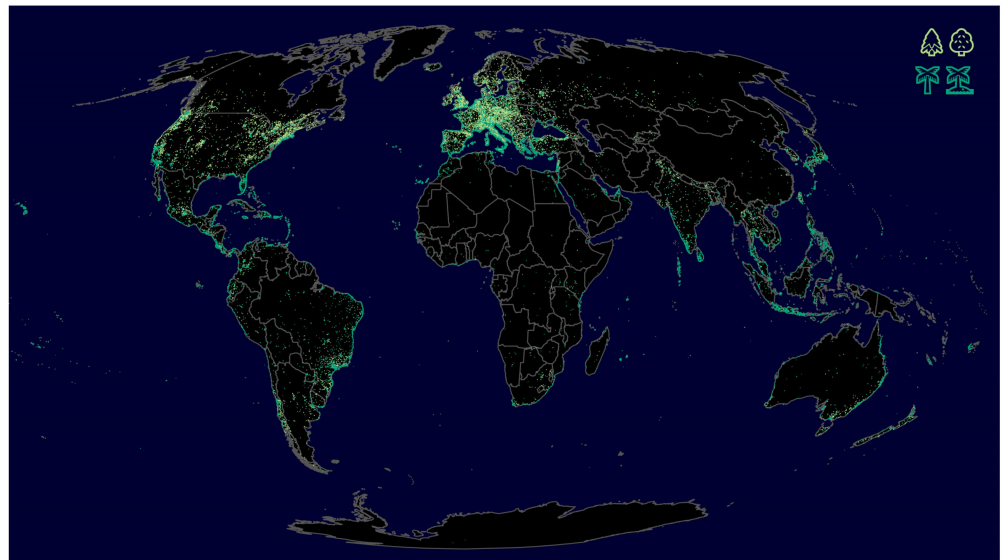
**Figure 3.** Emojis attributed as location-specific, selected from global sunrise- (purple) and sunset-typical (orange) emojis.

##### 4.2.1. An Initial Subdivision into Environmental Groups

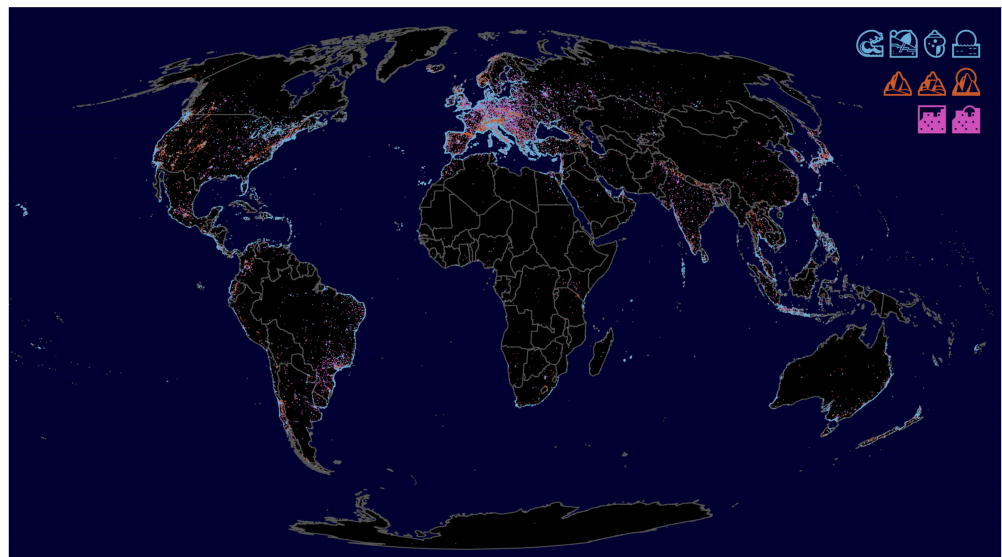
Figure 3 already shows that the emojis attributed as location-specific reflect different environments and can be grouped in this respect; however, this is not done based on their official Unicode name (because that is not shown to the user), but visually. Three groups refer to the geographical environment: urban, montane or maritime, although these three are not mutually exclusive. Two other groups refer to the predominant tree vegetation: palm trees on the one hand, and deciduous and coniferous trees on the other.

There is a very clear subdivision resulting from the calculation of typicality, which environment is typical for sunrise or sunset: emojis representing a montane environment are only typical for sunrise-related tweets. Urban and maritime environments, as well as palm trees, are only depicted by emojis typical for sunset. Only the deciduous tree emoji is typical for both sunrise and sunset, and coniferous tree only for sunrise.

The two maps in Figures 4 and 5 show the spatial distribution of the groupings of location-specific emojis just described (disregarding the distinction of sunrise and sunset), which represent the actual geographical conditions very well. Figure 4 shows that the palm tree emojis are mainly located along the coastlines, especially in tropical and subtropical areas, whereas those emojis indicating coniferous and deciduous trees are more likely to be found inland and outside the tropics. Additionally, along the coastlines are those emojis that represent a maritime environment (see Figure 5), but larger inland waters are recognizable as well. Mountain ranges and urban agglomerations stand out too.



**Figure 4.** Global spatial distribution of location-specific emojis representing the predominant tree vegetation.

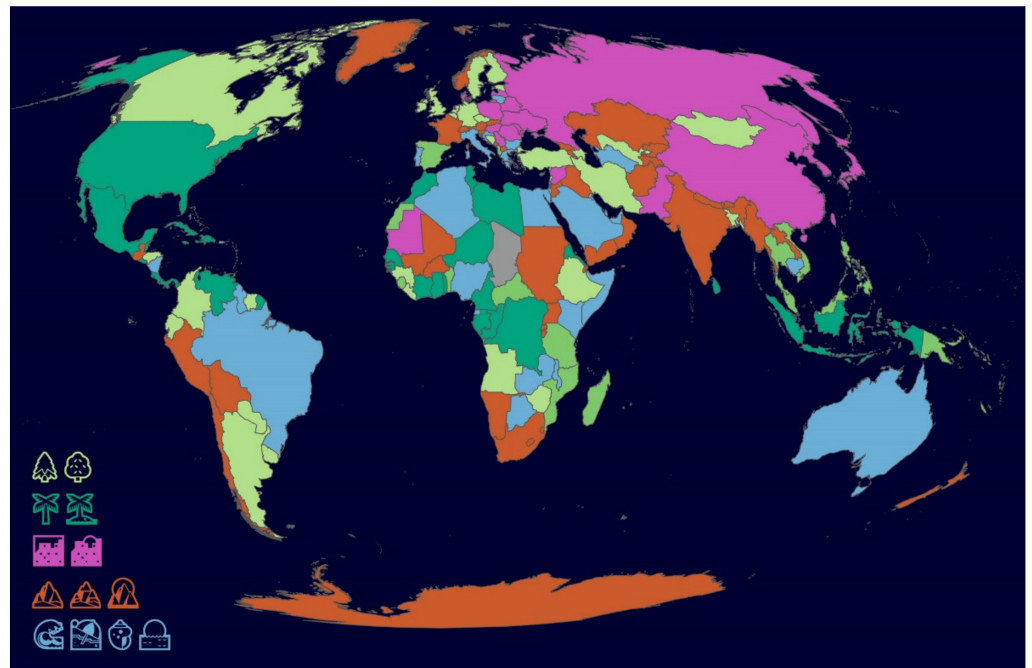


**Figure 5.** Global spatial distribution of location-specific emojis representing montane, urban and maritime environment.

These maps obviously do not provide any unexpected insights, but they do demonstrate, on the one hand, that emojis can provide very specific information about the user's perceived environment, and thus, on the other hand, that they are also likely to be suitable for answering less obvious questions.

#### 4.2.2. Country-Specific Differences

The defined location-specific emojis were furthermore examined on a country-specific level. For all these emojis, the typicality formula was applied again for each country. The total dataset is here the total number of location-specific emojis within a country, while the sub-dataset is the number of one of those emojis in the same country. In this way, the most typical location-specific emoji was calculated for each country (see Figure 6). For the values underlying Figure 6, see Supplementary Materials.



**Figure 6.** Countries by location-specific emojis calculated as typical, divided into environmental groups.








In order to examine how location-specific these emojis actually are and how meaningful the grouping made in advance is, various geographical data were added for each country, which in a sense represent the environmental groups previously formed:

- elevation span [47]—difference in altitude between the highest and lowest point of a country;
- coast ratio [47]—percentage of coast in the total country border;
- urban population [47]—percentage of population living in urban areas;
- forest coverage [48]—percentage of the country's area covered by forest.

Figure 4 implies that palm tree emojis are used in both a maritime and a tropical (i.e., characterized by a lot of vegetation) context, which corresponds to reality, as both are not mutually exclusive. Therefore, the environmental grouping of location-specific emojis has been further refined, i.e., the two palm tree emojis are classified differently, as they are difficult to properly assign to one group. This results in seven environmental groups, which can be seen in the first column of Table 1. The groups can be chosen according to the purpose of analysis. For each country, the typicality of all seven emoji groups was calculated and the correlation coefficient was determined with every geographical attribute described above. These values can be found in Table 1.

Hypothetically, it can be assumed that there is a positive correlation between an environmental emoji group and the geographical attribute that is representative for this group (e.g., between elevation span and emojis indicating a montane environment). For the groups montane, maritime I, maritime II, urban, palm and forest, this is entirely true. Thus, it seems that the location-specific emojis are apparently indeed chosen depending on the characteristics of the place where sunrise or sunset is observed. However, there are partially positive correlations for several geographical attributes per environmental emoji group, and the maximum correlation coefficient is not always evident for the representative geographical attribute.

**Table 1.** Correlation between the typicality of environmental groups of location-specific emojis within each country and the four geographical attributes (bold: respective representative geographical attribute; green: correlation coefficient for representative geographical attribute is positive and the maximum; yellow: correlation coefficient for representative geographical attribute is positive, but not the maximum; red: correlation coefficient for representative geographical attribute is negative).

		Elevation Span	Coast Ratio	Urban Population	Forest Coverage
montane		<b>0.40</b>	−0.43	−0.04	−0.09
maritime I		−0.19	<b>0.22</b>	0.04	−0.10
maritime II		−0.41	<b>0.50</b>	−0.07	0.11
urban		0.20	−0.29	<b>0.19</b>	−0.07
tree		0.20	−0.30	0.04	<b>−0.05</b>
palm		−0.30	<b>0.38</b>	−0.12	<b>0.22</b>
forest		−0.25	0.31	−0.12	<b>0.23</b>

Similarly, a calculation of the typicality of a location-specific emoji within the two sub-datasets relating to sunset and sunrise can be performed. Here, the total number of all these emojis in a country in the sunrise- or sunset-related sub-dataset is considered the total dataset, and within such a dataset the number of usages of a specific emoji in a country is considered the sub-dataset. Comparing the location-specific emojis calculated as typical for each country for sunrise and sunset, they match at 42%. Looking at the groups montane, maritime I, urban, tree and palm from Table 1, there is 59% correspondence. Unfortunately, these sunrise- and sunset-related typicalities do not reflect the typical activities and perceived surroundings described at the beginning. All location-specific emojis were taken into account, regardless of which of the two events they are typical for, which leads to the fact that the typicality formula for some reason determines the respective opposite as typical. However, a calculation for only the location-specific emojis typical for the respective event is also not meaningful, as there is virtually no intersection (only one tree emoji), and a comparison of sunrise and sunset therefore makes no sense. In this respect, sunrise and sunset are rather to be seen as events that cause social media users to perceive their surroundings and reflect them with the help of emojis.

Apart from the latter, the results obtained so far are quite conclusive, but it needs to be pointed out that a single emoji can hardly represent the geographical conditions of an entire country. Therefore, the scale of the study will become more detailed in the following. Furthermore, the emoji calculated as typical for one country is strongly influenced by the state boundaries, which are often not useful region separators, as both the shape and the scale influence the resulting values, thus introducing a modifiable area unit problem (MAUP).

## 5. Analysis of Italy

All the analyses that have already been carried out on a global scale are conducted again for Italy in the following. This results in different total and sub-datasets, as well as more detailed insights into the location-specific context of social media users. Apart from data availability, Italy was chosen because it is geographically very diverse: low and high mountain ranges, seacoast and inland waters, and thus different climatic and vegetation zones. In addition, there are numerous culturally and historically significant cities throughout the country.

### 5.1. Comparing Hashtags and Emojis

The procedure described in Section 4.1 for calculating the typicality, including a pre-selection of the 100 most frequent hashtags and emojis per sub-dataset, was also carried





For both the hashtags and the emojis, it is evident, as in the global context, that the frequency-based top 20s tend to reflect the phenomenon itself, and the top 20s calculated as typical provide information about activities and perceived objects in the environment, which are mainly the same both globally and in Italy. However, it is noticeable that for some reasons, the most typical emojis for Italy include a number of geometric shapes (circle, triangle, square). For the values underlying Figures 7 and 8, see Supplementary Materials.

Table 2 summarizes the percentage correspondences of sub-dataset top 20s, both within and across the global and at the Italian level. Table 2 (a) compares sunrise and sunset on the global and Italian level (e.g., at the global level, the frequency-based top 20 hashtags related to sunrise and sunset match by 80%). Table 2 (b), on the other hand, compares across levels (e.g., the global typicality-based emojis relating to sunrise match the Italian ones by 35%). It can be concluded that in both tables, the frequency-based sub-datasets show a higher correspondence than the typicality-based ones, i.e., they have more overlaps. Within the overall Italian dataset, these similarities are lower than globally (see Table 2 (a)), i.e., they are more differentiated. This makes sense in that the global view is a more general one and the country-based one more specific. A comparison across the global and Italian levels shows greater similarities than within each, whereby it is also clear here that the emojis calculated as typical in particular enable a more precise representation of the two events of sunrise and sunset.

**Table 2.** (a) Percentage correspondence between the top 20 of sunrise- and sunset-related sub-datasets at the global and Italian levels. (b) Percentage correspondence between top 20 s of sub-datasets at the global and Italian levels.

		Hashtags	Emojis
global	<i>frequency</i>	80%	90%
	<i>typicality</i>	5%	10%
Italy	<i>frequency</i>	55%	85%
	<i>typicality</i>	0%	0%

(a)

		Hashtags		Emojis	
		<i>sunrise</i>	<i>sunset</i>	<i>sunrise</i>	<i>sunset</i>
global vs. Italy	<i>frequency</i>	80%	85%	70%	80%
	<i>typicality</i>	40%	30%	50%	35%

(b)

### 5.2. Location-Specific Emojis

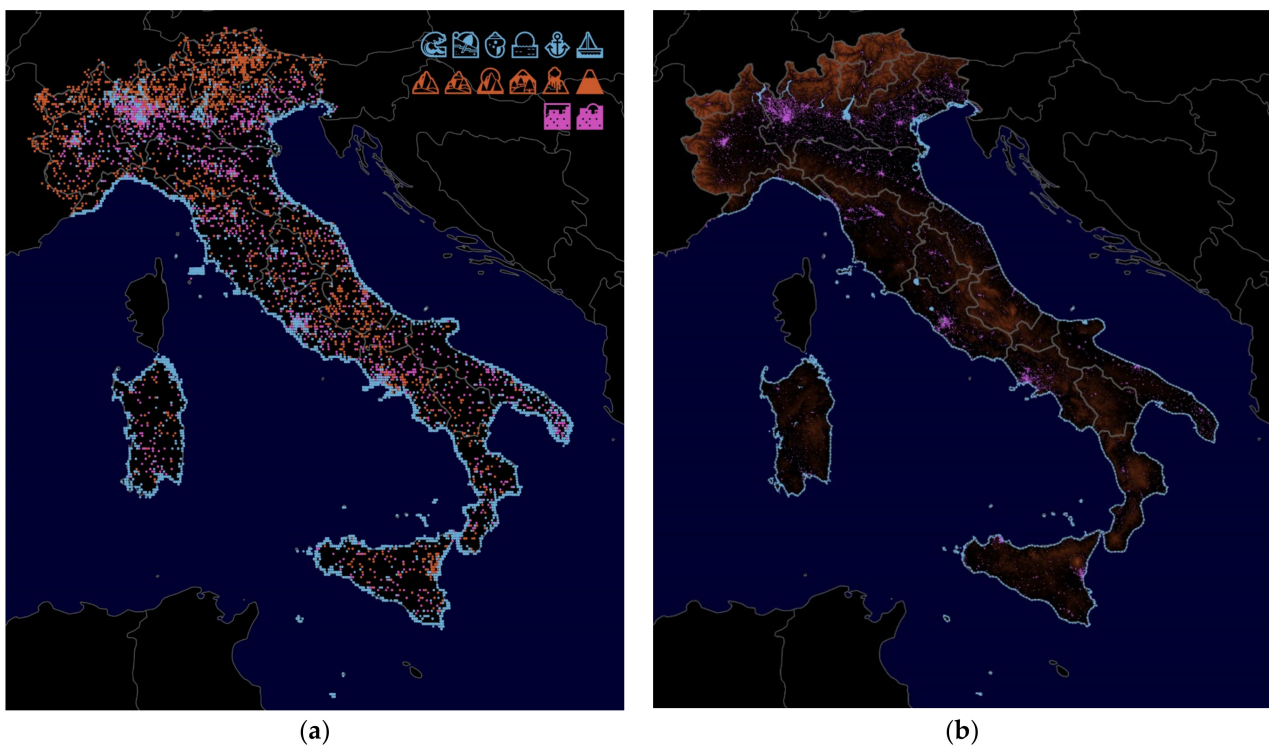
Since for Italy other emojis were calculated as typical due to different total and sub-datasets, these must be checked again for location-specific emojis according to the same criteria as described in Section 4.2. All emojis that were already attributed as location-specific in the global context are found again for Italy, as well as five further emojis (see Figure 9). The environments described by them remain the same: montane, urban, maritime, and in terms of tree vegetation. Two of the five additional Italy-specific emojis refer to maritime environment (anchor and sailboat), and three indicate a montane landscape. Among the latter is an emoji showing an erupting volcano and one that, according to the official Unicode name, represents Japan's Mount Fuji, which is also a volcano. However, as these official names are not directly obvious to the user, it can be noted that in Italy two volcano emojis are typically used, which is certainly due to the fact that there are five active volcanoes in Italy, with Mount Vesuvius and Mount Etna as the most famous. For the Italian location-specific emojis, there is no overlap within the sub-datasets, but there is also no clear separation of the environmental groups between sunrise and sunset, as is the case at the global level.



**Figure 9.** Emojis attributed as location-specific, selected from sunrise- (purple) and sunset-typical (orange) emojis of Italy (left part of each box: typical only for Italy; right part: typical on global and Italian level).

### 5.2.1. Spatial Occurrence of Environmental Groups

Since a map-based comparison with real conditions is less problematic at the country level than globally, the environmental emoji groups already spatially investigated globally are examined below in relation to the spatial patterns of the corresponding geographical conditions in Italy. When comparing Figures 10a and 10b, the correspondence can be clearly identified between the spatial distribution of these emojis representing montane, urban, as well as maritime environments on the one hand, and elevation, urban zones, large lakes and the Italian coastline on the other hand.

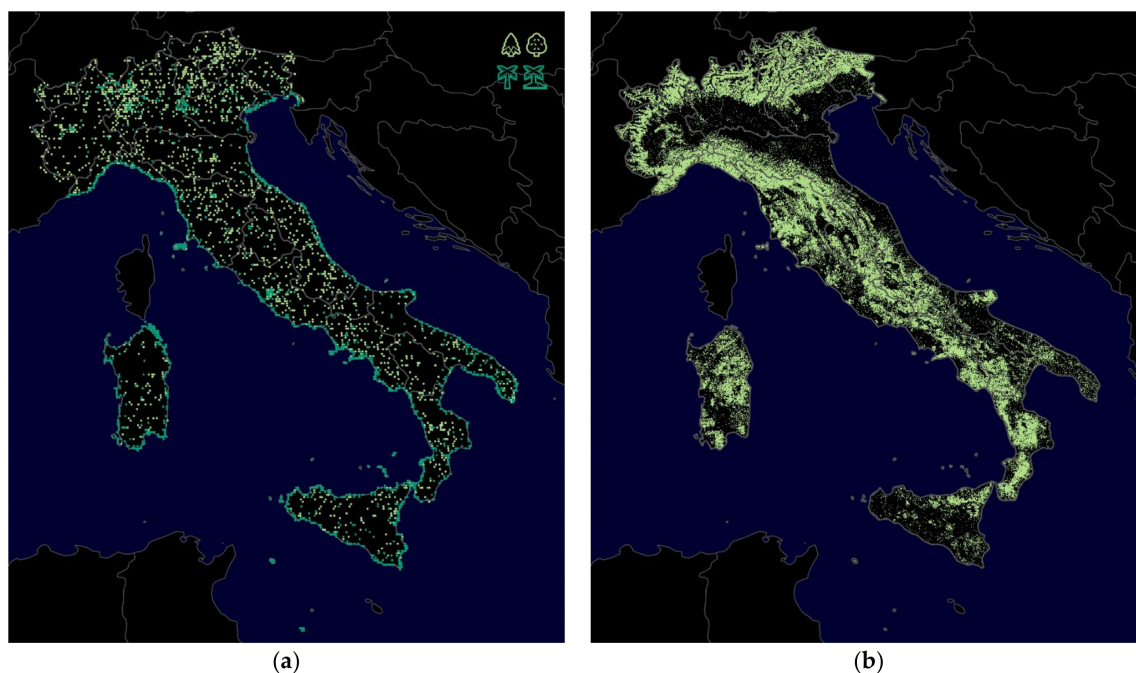


**Figure 10.** (a) Spatial distribution of location-specific emojis in Italy representing montane, urban and maritime environments; (b) elevation [49], urban zones, large lakes and coastline [50] of Italy in corresponding colors.

The situation is different for the location-specific emojis, which indicate tree vegetation. Similar to the global level, Italian palm tree emojis are used mainly along the seacoast, whereas the emojis for deciduous and evergreen trees are used in the inland (see Figure 11a). Figure 11b contains a map showing forest-covered areas in Italy. It is only partial, and still not particularly obvious: some less forested areas can be vaguely identified in the spatial distribution of the corresponding emojis (the Po Valley on the southern edge of the Alps, the south-eastern part of the mainland as well as the south-western half of large southernmost island Sicily), but they are not present at particularly large numbers. However, very striking is the stringent use of palm tree emojis along the coastline. Either these two emojis are used as a symbolic, simplified representation of a beach, or there are indeed palm trees on the Italian beaches, but they have not been identified as forest cover.

### 5.2.2. Region-Specific Differences

The calculations carried out in Section 4.2.2 at a global level to determine a typical location-specific emoji for each country were carried out in Italy for the 21 regions, which are the top level of administrative units (see Figure 12). For the values underlying Figure 12, see Supplementary Materials.



**Figure 11.** (a) Spatial distribution of location-specific emojis in Italy representing the predominant tree vegetation; (b) forest area in Italy [50].



**Figure 12.** Italian regions with location-specific emojis calculated as typical and color-coded by environmental group.

Likewise, for the refined environmental groups introduced in Table 1, the correlation was calculated between the typicality of these groups within each Italian region and the geographical attributes elevation span [49], coast ratio [50], urban population and forest coverage [51]. The results are presented in Table 3. The correlation coefficient is in every case positive for the geographical attribute representing an emoji group and compared to the correlation coefficients determined on a global level (see Table 1); the ones at the Italian level are higher. This suggests that a more specific representation of actual conditions at the country level is possible. Additionally, the effect of the different grouping of the two palm tree emojis is evident again.

**Table 3.** Correlation between the typicality of environmental groups of location-specific emojis within each Italian region and the four geographical attributes (bold: respective representative geographical attribute; green: correlation coefficient for representative geographical attribute is positive and the maximum; yellow: correlation coefficient for representative geographical attribute is positive, but not the maximum).






















		Elevation Span	Coast Ratio	Urban Population	Forest Coverage
montane		<b>0.69</b>	−0.64	−0.43	0.25
maritime I		−0.75	<b>0.74</b>	0.37	−0.21
maritime II		−0.74	<b>0.76</b>	0.36	−0.22
urban		0.07	−0.30	<b>0.43</b>	−0.17
tree		0.65	−0.68	−0.46	<b>0.23</b>
palm		−0.54	<b>0.91</b>	0.17	−0.37
forest		0.58	−0.38	−0.56	<b>0.09</b>

The representativeness of the most typical location-specific emojis of the Italian regions was further assessed in a different way. For each region, the percentage deviation from the overall Italian average was calculated for all four geographical attributes (see Table 4). It can be assumed that there is an above-average deviation for the geographical attribute that is representative for the most typical emoji of the respective region (e.g., above-average elevation span in a region with the most typical emoji indicating a montane environment). In 11 out of 21 cases (marked green in Table 4), the geographical attribute with the maximum positive deviation corresponds to the emoji. In five cases, the deviation of the representative geographical attribute is at least positive, although not the positive maximum value (marked yellow in Table 4).

The data contained in Table 4 are also illustrated as a map in Figure 13 applying specifically developed diagrams based on a combination of regular and reversed wing charts [52]. For the two regions Sicilia and Campania, the emoji showing an erupting volcano was calculated as the most typical location-specific one, and these are the regions where Mount Vesuvius and Mount Etna are located. Campania does not have an above-average elevation span, but Mount Vesuvius seems to be very dominant. The emoji showing Mount Fuji is typical for the regions Abruzzo and Veneto. There are no volcanoes in these regions, but both have an elevation span of around 3000 m. Therefore, this emoji was apparently interpreted as a mountain with its top covered in snow, although there is another one with the official Unicode name “snow-capped mountain”.



**Table 4.** Typical location-specific emojis and the percentage deviation of the four geographical attributes from the overall Italian average per administrative region (bold: respective representative geographical attribute; green: deviation of representative geographical attribute is above overall Italy average and the maximum; yellow: deviation of representative geographical attribute is above overall Italy average, but is not the maximum; red: deviation of representative geographical attribute is below overall Italy average).

Percentage Deviation from Overall Italy Mean					
Region	Elevation Span	Coast Ratio	Urban Population	Forest Coverage	Location-Specific Emoji
Abruzzo	<b>3.4%</b>	−6.4%	−17.1%	0.6%	
Alto Adige	<b>19.3%</b>	−31.0%	−5.3%	9.4%	
Basilicata	− <b>11.7%</b>	−23.4%	−5.3%	−10.7%	
Campania	− <b>15.8%</b>	16.3%	29.0%	−5.1%	
Piemonte	<b>35.8%</b>	−31.0%	−0.8%	2.1%	
Sicilia	<b>12.0%</b>	69.0%	−0.1%	−24.6%	
Trento	<b>19.2%</b>	−31.0%	−3.9%	25.7%	
Umbria	− <b>14.1%</b>	−31.0%	5.6%	10.9%	
Valle d’Aosta	<b>36.7%</b>	−31.0%	−27.2%	−5.1%	
Veneto	<b>10.7%</b>	−5.2%	−5.9%	−11.4%	
Calabria	−9.8%	<b>54.6%</b>	−10.3%	5.8%	
Emilia-Romagna	−11.7%	− <b>17.5%</b>	11.1%	−5.9%	
Friuli-Venezia Giulia	0.9%	− <b>14.2%</b>	4.6%	9.1%	
Liguria	−11.9%	<b>11.6%</b>	25.0%	35.5%	
Marche	−4.9%	<b>2.5%</b>	−13.6%	−2.7%	
Puglia	−32.6%	<b>36.5%</b>	−5.6%	−28.1%	
Lazio	−5.9%	1.4%	<b>25.8%</b>	−3.3%	
Lombardia	25.5%	−31.0%	<b>13.8%</b>	−8.6%	
Molise	−12.0%	−20.9%	−10.6%	<b>0.1%</b>	
Sardegna	−18.9%	<b>69.0%</b>	−9.2%	− <b>6.9%</b>	
Toscana	−14.2%	<b>12.7%</b>	1.0%	<b>13.2%</b>	



**Figure 13.** Map of typical location-specific emojis and percentage deviation of four geographic attributes from overall Italian average per administrative region.

Additionally, for Italy, a calculation of the typicality of the location-specific emoji within the two sub-datasets relating to sunset and sunrise was performed. Comparing the location-specific emojis calculated as typical for each region for sunrise and sunset, they match at 52%. Looking at the groups montane, maritime I, urban, tree and palm from Table 3, there is 81% correspondence. This suggests that the perceived objects in the environment hardly differ between these two events, although this reveals nothing about the activities carried out.

## 6. Discussion and Conclusion

Prior to a critical discussion of the results, the contributions of this paper will be outlined. One contribution is emojis themselves, as well as an investigation of what they can convey in comparison to hashtags, and how they represent spatial context and associated activities. The methodological contribution is the specially developed typicality measure, with the purpose of normalization and elimination of bias. This measure can emphasize

relative meanings, and the sub-dataset required for the calculation can be formed both thematically and spatially. An evaluation of the typicality measure is conducted in the form of a case study and a comparison with reality. The case study is carried out both on a global level, which has only been realized in a few studies so far, and on a more detailed scale, always using a privacy-aware HLL data structure. Various visualization methods were used to illustrate the results: word/emoji clouds, maps and a specially developed type of diagram.

The preceding analyses show that hashtags and emojis contained in social media posts have a differing information content. Hashtags provide more general information about the event or topic addressed in the post, whereas emojis, although available in smaller number than hashtags, are much more multifaceted and can indicate emotions, activities and locations. Since the analyses were carried out at two scales, the comparison showed that the more fine-grained level allowed for more detailed and nuanced insights, as well as a comparison with real geographic settings. Although the scale of an Italian province is still such that a single emoji can hardly represent its entirety, typical or distinctive attributes of a province could be recognized. An analysis on an even more detailed level is possibly limited by the available number of social media posts. In the case of Italy, for example, an analysis of the capital city Rome should not be problematic, as numerous posts are available, which would probably again turn out to reveal deviating location-specific emojis. In the study presented, the focus was on emojis that reflect the user's geographical surroundings, but certainly other categories could also be found that have a spatial reference and can thus be examined for their spatial correlation.

Emojis can convey a lot of ambiguity. Depending on the context, a particular emoji can have a very different meaning. For example, the water wave emoji can represent a flood disaster, but also a day at the beach. In our case, it can be assumed that this ambiguity is largely excluded by the predetermined context and thus a thematic restriction.

When looking at the results, it is also important to bear in mind that only four languages were considered in the data query, which only represent one cultural user group. Accordingly, the posts in the dataset are found where these languages are spoken or where the people who speak these languages travel to. This explains the Europe- and North America-focused distribution of the data in Figures 4 and 5, meaning that users are locals in some places and tourists in others. The activities and perceptions described in relation to sunrise and sunset are therefore not transferrable to other cultures. Furthermore, the posts studied for Italy are not written in the country's native language and are therefore very likely to be mainly created by tourists. That said, it cannot be excluded that local Italians post in English, as it is a universal language in social media.

Nevertheless, it can be stated that relatively simple analyses reveal obvious things, and that this approach should thus be extendable to other events. The former could at least be ascertained for the used dataset with georeferenced Instagram posts on the topic of sunrise and sunset. In this case, it is therefore an event that the user him-/herself witnesses and documents photographically. It is questionable how topics or events that cannot be seen in the photo are reflected in emojis, or how this is manifested on purely text-based platforms such as Twitter.

In conclusion, it can be said that emojis can be considered contextual indicants in location-based social media posts, as claimed at the beginning. This suggests that they are also likely to be a rich source of information at other levels, which can be analyzed in an uncomplicated and largely language-independent way.

**Supplementary Materials:** The values underlying Figures 1, 2, 6–8 and 12 are available online at <https://www.mdpi.com/article/10.3390/ijgi10060407/s1>.

**Author Contributions:** Conceptualization: Eva Hauthal; data curation: Alexander Dunkel; funding acquisition: Dirk Burghardt; investigation: Eva Hauthal and Alexander Dunkel; methodology: Eva Hauthal; resources: Eva Hauthal and Alexander Dunkel; writing—original draft: Eva Hauthal; writing—review and editing: Eva Hauthal, Alexander Dunkel and Dirk Burghardt. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the German Research Foundation as part of the priority programme “Volunteered Geographic Information: Interpretation, Visualisation and Social Computing” (VGIscience, priority programme 1894). Open Access Funding by the Publication Fund of TU Dresden.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data available on request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Boyd, D.; Crawford, K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Inf. Commun. Soc.* **2012**, *15*, 662–679. [\[CrossRef\]](#)
2. Karami, A.; Kadari, R.R.; Panati, L.; Nooli, S.P.; Bheemreddy, H.; Bozorgi, P. Analysis of Geotagging Behavior: Do Geotagged Users Represent the Twitter Population? *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 373. [\[CrossRef\]](#)
3. Petutschnig, A.; Resch, B.; Lang, S.; Havas, C. Evaluating the Representativeness of Socio-Demographic Variables over Time for Geo-Social Media Data. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 323. [\[CrossRef\]](#)
4. Emoji Timeline. Available online: <https://emojitime.com/> (accessed on 13 April 2021).
5. Bai, Q.; Dan, Q.; Mu, Z.; Yang, M. A Systematic Review of Emoji. *Front. Psychol.* **2019**, *10*, 2221. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Hu, X.; Tang, J.; Gao, H.; Liu, H. Unsupervised Sentiment Analysis with Emotional Signals. In Proceedings of the 22nd International Conference on the World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; Schwabe, D., Ed.; ACM: New York, NY, USA, 2013; pp. 607–617.
7. Novak, P.K.; Smailović, J.; Sluban, B.; Mozetič, I. Sentiment of Emojis. *PLoS ONE* **2015**, *10*, e0144296. [\[CrossRef\]](#)
8. Evans, V. *The Emoji Code: How Smiley Faces, Love Hearts and Thumbs up Are Changing the Way We Communicate*; Michael O’Mara Books Limited: London, UK, 2017.
9. Hakami, S.A.A. The importance of understanding emoji: An investigative study. *Res. Top. HCI* **2017**, 1–20.
10. Harris, R.B.; Paradice, D. An Investigation of the Computer-mediated Communication of Emotions. *J. Appl. Sci. Res.* **2007**, *3*, 2081–2090.
11. Pavalanathan, U.; Eisenstein, J. Emoticons vs. Emojis on Twitter. *arXiv* **2015**, arXiv:1510.08480.
12. Er, M.J.; Liu, F.; Wang, N.; Zhang, Y.; Pratama, M. User-Level Twitter Sentiment Analysis with a Hybrid Approach. In *Advances in Neural Networks—ISNN 2016*; Cheng, L., Liu, Q., Ronzhin, A., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 426–433.
13. Illendula, A.; Yedulla, M.R. Learning Emoji Embeddings using Emoji Co-occurrence Network Graph. *arXiv* **2015**, arXiv:1806.07785.
14. Rodrigues, D.; Prada, M.; Gaspar, R.; Garrido, M.V.; Lopes, D. Lisbon Emoji and Emoticon Database (LEED): Norms for emoji and emoticons in seven evaluative dimensions. *Behav. Res. Methods* **2018**, *50*, 392–405. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Sari, Y.A.; Ratnasari, E.K.; Mutrofin, S.; Arifin, A.Z. User emotion identification in twitter using specific features: Hashtag, emoji, emoticon, and adjective term. *J. Ilmu Komput. Inf.* **2014**, *7*, 18. [\[CrossRef\]](#)
16. Seyednezhad, S.M.M.; Fede, H.; Herrera, I.; Menezes, R. Emoji-Word Network Analysis: Sentiments and Semantics. In Proceedings of the Thirty-First International Florida Artificial Intelligence Research Society Conference (FLAIRS-31), Melbourne, FL, USA, 21–23 May 2018; Brawner, K., Rus, V., Eds.; AAAI Press: Palo Alto, CA, USA, 2018; pp. 271–274.
17. Suttles, J.; Ide, N. Distant Supervision for Emotion Classification with Discrete Binary Values. In *Computational Linguistics and Intelligent Text Processing, Proceedings of the 14th International Conference, CICLing 2013, Samos, Greece, 24–30 March 2013*; Gelbukh, A., Ed.; Springer: Berlin, Germany, 2013; pp. 121–136.
18. Ayvaz, S.; Shiha, M.O. The Effects of Emoji in Sentiment Analysis. *Int. J. Comput. Electr. Eng.* **2017**, *9*, 360–369. [\[CrossRef\]](#)
19. Wijeratne, S.; Balasuriya, L.; Sheth, A.; Doran, D. EmojiNet: An Open Service and API for Emoji Sense Discovery. In Proceedings of the Eleventh International AAAI Conference on Web and Social Media, Montréal, QC, Canada, 15–18 May 2017; AAAI Press: Palo Alto, CA, USA, 2017; pp. 437–446.
20. Ai, W.; Lu, X.; Wang, N.; Huang, G.; Mei, Q. Untangling Emoji Popularity through Semantic Embeddings. In Proceedings of the Eleventh International AAAI Conference on Web and Social Media, Montréal, QC, Canada, 15–18 May 2017; AAAI Press: Palo Alto, CA, USA, 2017.

21. Barbieri, F.; Ronzano, F.; Saggion, H. What does this Emoji Mean? A Vector Space Skip-Gram Model for Twitter Emojis. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), Portorož, Slovenia, 23–28 May 2016; Calzolari, N., Ed.; European Language Resources Association: Paris, France, 2016; pp. 3967–3972.
22. Wagner, A.; Marusek, S.; Yu, W. Emojis and Law. *Soc. Semiot.* **2020**, *30*, 396–414. [CrossRef]
23. Kelly, R.; Watts, L. Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships. *Exp. Technol. Approp. Unanticipated Users Usage Circumst. Des.* **2015**, *2*.
24. Miller, H.; Klüvner, D.; Thebault-Spieker, J.; Terveen, L.; Hecht, B. Understanding Emoji Ambiguity in Context: The Role of Text in Emoji-Related Miscommunication. In Proceedings of the Eleventh International AAAI Conference on Web and Social Media, Montréal, QC, Canada, 15–18 May 2017; AAAI Press: Palo Alto, CA, USA, 2017; pp. 152–161.
25. Miller, H.; Thebault-Spieker, J.; Chang, S.; Johnson, I.; Terveen, L.; Hecht, B. “Blissfully happy” or “ready to fight”: Varying Interpretations of Emoji. In Proceedings of the Tenth International AAAI Conference on Web and Social Media, Cologne, Germany, 17–20 May 2016; AAAI Press: Palo Alto, CA, USA, 2016; pp. 259–268.
26. Riordan, M.A. Emojis as Tools for Emotion Work: Communicating Affect in Text Messages. *J. Lang. Soc. Psychol.* **2017**, *36*, 549–567. [CrossRef]
27. Tigwell, G.W.; Flatla, D.R. Oh that’s what you meant!: Reducing emoji misunderstanding. In *MobileHCI '16, Proceedings of the 18th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, Florence, Italy, 06–09 September 2016*; ACM: New York, NY, USA, 2016; pp. 859–866.
28. Barbieri, F.; Ballesteros, M.; Saggion, H. Are Emojis Predictable? In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, Valencia, Spain, 3–7 April 2017; Lapata, M., Blunsom, P., Koller, A., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017.
29. Barbieri, F.; Ballesteros, M.; Ronzano, F.; Saggion, H. Multimodal Emoji Prediction. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, LA, USA, 1–6 June 2018; Walker, M., Ji, H., Stent, A., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 679–686.
30. Illendula, A.; Manohar, K.V.; Yedulla, M.R. Which Emoji Talks Best for My Picture? In Proceedings of the 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), Santiago, Chile, 3–6 December 2018; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2018; pp. 514–519.
31. Kim, J.; Gong, T.; Kim, B.; Park, J.; Kim, W.; Huang, E.; Han, K.; Kim, J.; Ko, J.; Lee, S.-J. No More One Liners. *ACM Trans. Soc. Comput.* **2020**, *3*, 1–25. [CrossRef]
32. Barbieri, F.; Kruszewski, G.; Ronzano, F.; Saggion, H. How Cosmopolitan Are Emojis?: Exploring emojis usage and meaning over different languages with distributional semantics. In Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016; ACM Association for Computing Machinery: New York, NY, USA, 2016.
33. Barbieri, F.; Espinosa-Anke, L.; Saggion, H. Revealing Patterns of Twitter Emoji Usage in Barcelona and Madrid. *Front. Artif. Intell. Appl.* **2016**, *288*, 239–244.
34. Guntuku, S.C.; Li, M.; Tay, L.; Ungar, L.H. Studying Cultural Differences in Emoji Usage across the East and the West. In Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019), Munich, Germany, 11–14 June 2019; AAAI Press: Palo Alto, CA, USA, 2019; pp. 226–235.
35. Ljubešić, N.; Fišer, D. A Global Analysis of Emoji Usage. In Proceedings of the 10th Web as Corpus Workshop, Berlin, Germany, 12 August 2016; Cook, P., Evert, S., Schäfer, R., Stemle, E., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2016; pp. 82–89.
36. Hauthal, E.; Burghardt, D.; Dunkel, A. Analyzing and Visualizing Emotional Reactions Expressed by Emojis in Location-Based Social Media. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 113. [CrossRef]
37. Na’aman, N.; Provenza, H.; Montoya, O. Varying Linguistic Purposes of Emoji in (Twitter) Context. In Proceedings of ACL 2017, Student Research Workshop, Vancouver, BC, Canada, 30 July–4 August 2017; Ettinger, A., Gella, S., Labeau, M., Alm, C.O., Carpuat, M., Dredze, M., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 136–141.
38. Danesi, M. *The Semiotics of Emojis*; Bloomsbury Publishing: London, UK; Oxford, UK; New York, NY, USA; New Delhi, India; Sydney, Australia, 2017.
39. Ge, J.; Gretzel, U. Emoji rhetoric. *J. Mark. Manag.* **2018**, *34*, 1272–1295. [CrossRef]
40. Pohl, H.; Domin, C.; Rohs, M. Beyond Just Text: Semantic Emoji Similarity Modeling to Support Expressive Communication. *ACM Trans. Comput.-Hum. Interact.* **2017**, *24*, 1–42. [CrossRef]
41. Statistica. Available online: <https://www.statista.com/statistics/253577/number-of-monthly-active-instagram-users/> (accessed on 3 June 2021).
42. Marine-Roig, E.; Martin-Fuentes, E.; Daries-Ramón, N. User-generated events in tourism: A new must-not-miss phenomenon. *Sustainability* **2017**, *9*, 2250.
43. Dunkel, A.; Löchner, M.; Burghardt, D. Privacy-Aware Visualization of Volunteered Geographic Information (VGI) to Analyze Spatial Activity. *Int. J. Geo-Inf.* **2020**, *9*, 607. [CrossRef]
44. Flajolet, P.; Fusy, É.; Gandouet, O. HyperLogLog: The analysis of a near-optimal cardinality estimation algorithm. In Proceedings of the Conference on Analysis of Algorithms: AofA 07, Nice, France, 17–22 June 2007; pp. 127–146.
45. Emojitracker. Available online: <https://emojitracker.com/> (accessed on 3 June 2021).



- 
46. Princeton University WordNet. Available online: <https://wordnet.princeton.edu/> (accessed on 13 April 2021).
  47. The World Factbook. Available online: <https://www.cia.gov/the-world-factbook/> (accessed on 13 April 2021).
  48. Atlas Big. Available online: <https://www.atlasbig.com/> (accessed on 13 April 2021).
  49. Copernicus Land Monitoring Service. Available online: <https://land.copernicus.eu/> (accessed on 13 April 2021).
  50. European Environment Agency. Available online: <https://www.eea.europa.eu/> (accessed on 13 April 2021).
  51. Eurostat. Available online: <https://ec.europa.eu/eurostat/> (accessed on 13 April 2021).
  52. Schnabel, O. Diagrammformen in Karten. Ph.D. Dissertation, Institut für Kartografie, ETH Zürich, Zürich, Switzerland, 2007.