

# Emojis as indicators of spatial-temporal-thematic developments in geo-social media

Samantha Levi <sup>a,\*</sup>, Eva Hauthal <sup>a</sup>, Sagnik Mukherjee <sup>a</sup>

<sup>a</sup> *Institute of Cartography, Technische Universität Dresden – levibsamantha@gmail.com – eva.hauthal@tu-dresden.de, sagnik.mukherjee1@tu-dresden.de*

\* Corresponding author

**Keywords:** geo-social media, location-based social media, emoji, spatial-temporal analysis, typicality

## Abstract:

Because social media plays an increasingly large role in the lives of over half the world's population, it serves both as a platform through which users express thoughts and ideas and as a powerful source of data. Geo-social media allows for the spatial contextualization of this data, providing valuable insights into both individual and collective reactions to events, products, and individuals. Within geo-social media posts, emojis are increasingly used to serve both emotional and semantic functions. Research has shown that emojis are language-independent indicators of both a user's emotions and context, which can help bypass errors arising from slang, sarcasm, spelling mistakes, and grammatical errors. While the connotative meaning of an emoji may be ambiguous, the denotative meaning hardly ever is. Emojis can therefore be thematically contextualized by hashtags occurring in the same post.

The overarching research objective for this study is to investigate the extent to which emojis can be used to identify topics of discussion on social media as well as their spatial-temporal evolution in a dataset that is originally not thematically pre-filtered. This research aims to 1) detect spatial-temporal change in emoji usage over time, 2) investigate whether significant changes in emoji usage over time and space correspond with significant events, and 3) assess the suitability of existing geovisualization techniques and adapt selected methods to generate static or interactive maps. The results of this study will help determine the usefulness of emojis as indicators of spatial-temporal-thematic developments in geo-social media and illustrate the necessary considerations to be made when working with such data. Resulting visualizations will highlight emojis that have significant spatial or temporal variations in use.

The dataset used in this study consists of approximately 4 million Twitter posts spanning across Europe in the year 2020. Data preprocessing has been conducted to ensure each post in the dataset is geotagged and contains at least one emoji and one hashtag. The data will be analyzed both in its raw form and in the privacy-aware data format HyperLogLog (Dunkel et al. 2020). The conceptual framework of Dunkel et al. (2019), which describes reactions to events in geo-social media based on spatial, temporal, topical, and social facets, will be used to structure the data analysis.

A variety of statistical measures exist to potentially detect change in emoji usage over time and space; however, not all of these measures generate significant results. Absolute and relative frequencies, for example, allow for only surface-level analysis of emoji usage. In order to analyze the data more deeply, emoji usage must be normalized. The typicality measure, first introduced by Hauthal et al. (2021), allows for the calculation of normalized differences of relative frequencies, in this case effectively identifying the most characteristic emojis within a designated subset of a larger dataset.

Due to the massive nature of the dataset, a combination of SQL and Python programming languages will be used for spatial data exploration, analysis, and visualization. After dividing the dataset into spatial and temporal sub-datasets of varying granularities, an iterative analysis of results will be conducted during which various emojis are tested for their typicality – in other words, for their representativeness within each sub-dataset. This method will be repeated until instances of unusual or otherwise significant emoji usage are identified. For these instances, a form of ground-truthing will be conducted by which co-occurring hashtags are investigated to provide thematic context for the use of each emoji. If the hashtags seem to have topical consistency within the subset, it is more likely that topical context has some effect on emoji usage. If there is no topical consistency, it may be concluded that spatial-temporal changes in emoji usage happen independently of topical context. This ground-truthing may help to explain why certain emojis are more popular in some spatial and temporal locations than others.

**References**

- Hauthal E., Dunkel A. & Burghardt D. (2021): Emojis as Contextual Indicants in Location-Based Social Media Posts. *ISPRS International Journal of Geo-Information* 10 (6): 407.
- Dunkel A., Andrienko G., Andrienko N., Burghardt D., Hauthal E. & Purves, R. (2019): A conceptual framework for studying collective reactions to events in location-based social media. *International Journal of Geographical Information Science* 33 (4): 780.
- Dunkel A., Löchner M. & Burghardt D. (2020): Privacy-Aware Visualization of Volunteered Geographic Information (VGI) to Analyze Spatial Activity. *ISPRS International Journal of Geo-Information* 9 (10): 607.