Quantitative Content Analysis of maps on Twitter using network visualization and pre-trained machine learning models

Edyta P. Bogucka [a, *], Liqiu Meng [a]

[a] Technical University of Munich, Chair of Cartography, e.p.bogucka@tum.de, liqiu.meng@tum.de
* Corresponding author

Keywords: Content Analysis, Visual Analysis, Network Mapping, Machine Learning

Online maps are mirrors of a constantly evolving society. Tracking maps shared in social media and their underlying design patterns may help us sense which aspects of the world are represented and who, what and where is absent. The latest research on map content utilizes the quantitative content analysis (QCA). This process starts from collecting a dataset of maps and assigns each map a set of codes from previously established codes (categories) definitions. Fisch (2020) used QCA to evaluate the vividness of maps of climate change, while Zuo et al. (2019) used it as a tool to compare the cartographic design and interaction elements of two new genres of maps - story maps and map-based dashboards.

However, there are several constraints in using QCA for online map studies. Firstly, the content analysis is mostly executed manually by one or many coders and as such is time consuming, sample-limited, and hardly reproducible, unless the researchers share their coding scheme with other fellows. Secondly, the classification of map content is assumption-driven. The research questions about the data sample are set apriori and replicate already known theoretical dimensions, e.g., map classifications established in previous research (Muehlenhaus, 2011). Finally, maps published online tell more than the explicitly mapped contents. They are embedded in a digital environment - depending on their distribution platform, they are enriched with metadata of hashtags, timestamps, likes, or comments. As a result, online maps exist as parts of a larger network, in which a single image may become closer to other images, for instance because it comes from the same media organizations or is posted with the same hashtag.

We tackle these limitations by implementing a semi-automated QCA of online maps. Instead of starting with the assumptions about the design of maps shared in social media, we use the pre-trained machine learning models to assign labels to maps and classify their content. By doing so, we keep cartographers at the beginning and the end of the value loop as map producers and as the ones who explore, explain, and make sense of these new map networks.

We curated a collection of maps shared on Twitter by several media organizations, including Die Zeit Online Germany), The New York Times (US) and The Daily Charts by The Economist (United Kingdom). The images were scraped from the official Twitter accounts (@zeitonline, @nytimes, @ECONdailycharts) using the Twitter Media Downloader extension for the Chrome web browser. The start date of the map collection was different for each newsroom (Zeit Online, The Daily Charts − 01.01.2012, The New York Times − 01.01.2016), but the end date was fixed to 01.06.2021. The query resulted in 1100 scraped maps and their metadata accompanying tweet text, tweet replies, retweets, and likes.

The semi-automated QCA was performed using two visualization techniques: (1) a keyword-image co-occurrence network built with the Google Vision API and Gephi, and (2) a color similarity grid built with the ImageSorter. For the first technique, we connected the folder with scraped maps to the Google Vision API. The computer vision tagging service returned between $4 - 10$ keywords (tags) for each map. The extracted keywords and maps were then loaded into Gephi to create a bipartite network with two types of nodes − keywords and images. The size of the keyword node was additionally adjusted based on the keyword frequency. The ImageSorter is an image browsing application that clusters similar images by their color, name, size, or creation date and allows to trace a position of a single image within the image clusters. We used this application to load the map samples and find dominant maps (more frequently shared), map repetitions, map variations and design outliers.

The QCA has revealed the following preliminary findings: The share of maps among all visual contents in social media evolves dynamically. In 2016 the @ECONdailycharts released 162 maps (18,3% share in their total visual media), @nytimes published 264 maps (1,5% share), while @zeitonline only 16 maps (1,7% share). In contrast, the @nytimes published 246 maps (8.4% share) and the @zeitonline followed with 156 releases (6.7% share) in 2020, a year marked by the Covid-19 pandemic. Such a difference leads to immediate questions such as "what were the underlying driving forces of these dynamics? In which contexts and why the map representations were preferred over other visualization types?". More generally, we want to share our thoughts on several open questions related to the changing social media and spatio-temporal contexts:

(1) The circulation of images in social media is steered by sorting and recommending algorithms. How do they "interpret" the content of maps and what keywords are the most prominent as their "interpretation" output (Figure 2)? Which map design styles predominate and how do map outliers stand out?

(2) Maps are tied to the zeitgeist and reflect cultural interests, trends, and concerns as show in Figure 2. Which new maps would pop up next in visual threads on social media? Will they be triggered by events or make new events happen? Would they be governed by local or national contexts? What design guidelines will they follow?
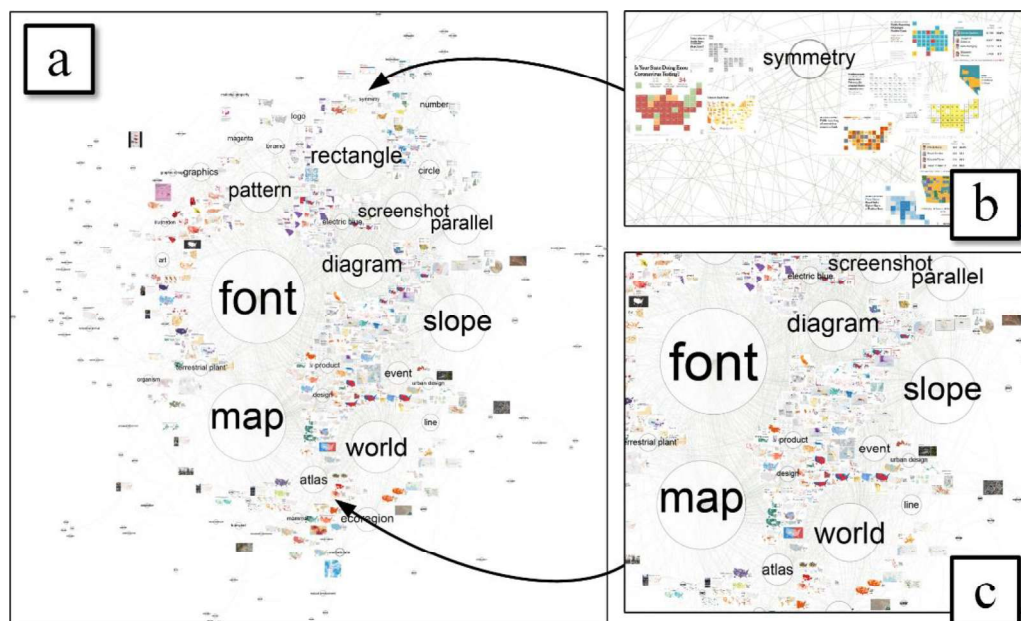

Figure 1. The keyword-image co-occurrence network for maps published online by The New York Times between 2018 - 2021 (a). The gridded cartograms form a cluster of "symmetry" (b), while most of the map samples lies in the closed cluster between the "font – diagram – slope – world – map – atlas" keywords (c). The outliers include image maps and satellite imageries.
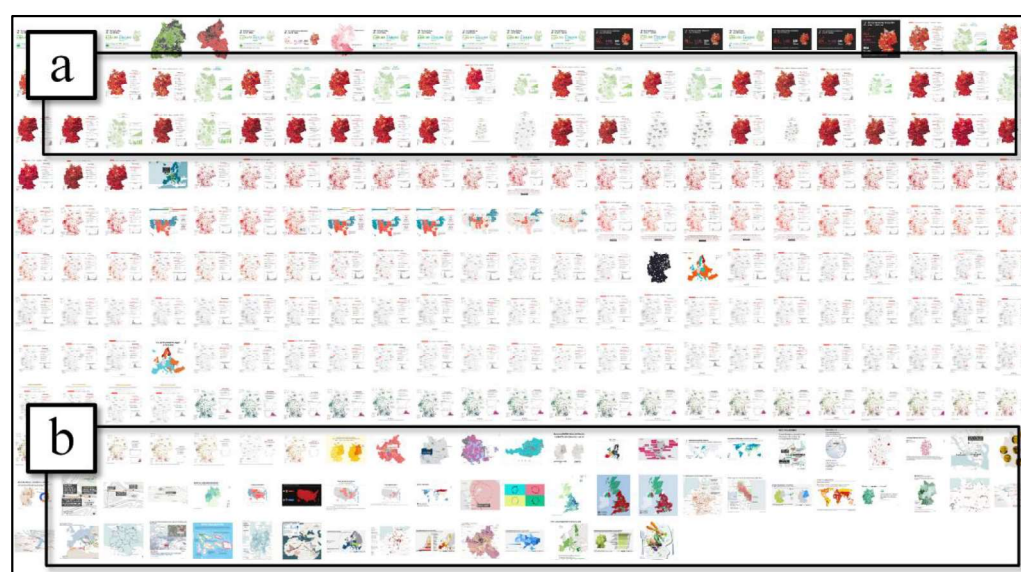

Figure 2. Image grid of maps published in Die Zeit Online starting from 2012 (bottom) up to mid-2021 (top). The cluster (a) shows the repetitive "red" maps of the COVID-19 virus spread, interspersed with the "green" maps of vaccination progress in Germany. Cluster (b) reveals the higher diversity of forms, colors and topics in the pre-pandemic times.

References
Fish, C. S. (2020) Cartographic content analysis of compelling climate change communication. Cartography and Geographic Information Science, 47:6, 492-507, DOI: 10.1080/15230406.2020.1774421
Muehlenhaus, I. (2011) Another Goode Method: How to Use Quantitative Content Analysis to Study Variation in Thematic Map Design. Cartographic Perspectives, (69), 7-30, DOI: https://doi.org/10.14714/CP69.28
Zuo, C., Ding, L., Bogucka, E. P., Meng, L. (2019) Map-based Dashboards versus Storytelling Maps. Proceedings of the 15th International Conference on Location Based Services, DOI: doi:10.34726/lbs2019