CLUSTER ANALYSIS FOR THE STUDY OF ONLINE VISUAL COMMUNICATION

Matteo Magnani[1], Matias Piqueras[1], Alexandra Segerberg[2], Davide Vega[1] and Victoria Yantseva[1]

[1] InfoLab, Department of Information Technology, Uppsala University, Sweden (e-mail: `matteo.magnani@it.uu.se`, `matias.piqueras@it.uu.se`, `davide.vega@it.uu.se`, `victoria.yantseva@it.uu.se`)

[2] Department of Government, Uppsala University, Sweden (e-mail: `alexandra.segerberg@statsvet.uu.se`)

ABSTRACT: Cluster analysis is a fundamental tool for the study of online communication. In this contribution we focus on the task of clustering online communication networks containing images. We provide an overview of the available approaches to cluster images and how to use image clustering as part of a social data science process. Then we present an approach to cluster online communication networks based on image clustering, that we apply to the study of climate change communication.

KEYWORDS: Online communication, image clustering, networks.

1 Cluster analysis and communication studies

Several online data analysis tasks such as the detection of coordinated inauthentic behaviour used to amplify online disinformation and the study of online polarisation and political persuasion require the summarisation of large numbers of social media posts. Therefore, cluster analysis is a fundamental tool for the study of online communication.

When we consider the type of data generated by people communicating online, visual content plays a fundamental role in all the main social media platforms. Images often carry an important part of the information contained in a social media post; in general, images are also associated to increased spreading, and can be shared across languages (Joo & Steinert-Threlkeld, 2018; Magnani *et al.*, 2013). This is true for image-based (Instagram), video-based (YouTube), and micro-blogging (Twitter) platforms. However, most of the existing computational studies of online communication, including those based on clustering methods, have focused only on either the networks of interaction (e.g. replies and retweets), to discover communities, or the text contained in social media posts, to compute topic models.

In this contribution we provide a road map towards the usage of image clustering to summarise online communication networks, starting from an overview of the available approaches to cluster images. Recent advances in deep learning have added more options to the traditional tools used to extract features from images, providing a number of alternative clustering approaches that still require a thorough comparison.
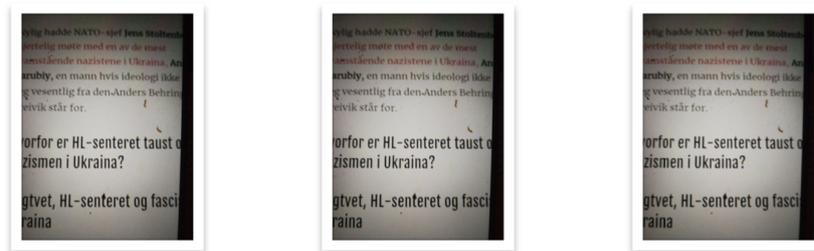
(a) A cluster with identical images independently posted by different accounts.

(b) A cluster with images from the same service/source.

(c) A cluster with images related to the same news.

Figure 1. *Three different types of clusters, from Deliverable 5.1, NORdic observatory for digital media and information DISorder (NORDIS)*

The focus of this part of the presentation is on clustering social images (for example, images shared on social media) and how to use image clustering as part of a social data science process (Magnani & Segerberg, 2021; Chen *et al.*, 2021; Giordano *et al.*, 2021; Zhang & Peng, 2022). Figure 1 shows an example of how the same approach, in this case based on clustering the colour histograms of the images, can produce clusters with different functions:

a cluster collecting identical images, a cluster of images coming from the same source, and a cluster of images showing pictures from the same event.

Then we present an approach to cluster online communication networks based on image clustering. This is based our previous work (Vega & Magnani, 2018; Vega & Magnani, 2019) on clustering networks with temporal and textual information (Figure 2), and consists in applying image clustering to organise the posted images into groups, assigning a label to each group or sets of groups to characterise the theme of the interactions, and using these themes to define thematic multiplex networks of social interactions. In these networks, edges represent interactions (e.g. replies, or retweets), and each layer of the multiplex network only includes interactions happened around that theme. Such networks can themselves be clustered using algorithms for multiplex networks (Magnani *et al.*, 2021).

Finally, we conclude by showing an application of image-based communication network clustering to study online visual climate change communication. This is based on a collection of tweets using the #COP*xx* hashtag, with *xx* being the number of the Conference of the Parties (COP) meeting, e.g. COP21 being the meeting held in Paris in 2015. The objective of this case study is both to highlight the importance of different features of image clustering methods within this type of research (e.g. scalability and explainability), and to showcase the possibilities and limitations of this research design.

References

CHEN, YAN, SHERREN, KATE, SMIT, MICHAEL, & LEE, KYUNG YOUNG. 2021. Using social media images as data in social science research. *New Media & Society*, Aug., 146144482110387.

GIORDANO, GIUSEPPE, PRIMERANO, ILARIA, & VITALE, PIERLUIGI. 2021. A Network-based Indicator of Travelers Performativity on Instagram. *SOCIAL INDICATORS RESEARCH*, **156**, 165–179.

JOO, JUNGSEOCK, & STEINERT-THRELKELD, ZACHARY C. 2018. Image as Data: Automated Visual Content Analysis for Political Science. *arXiv:1810.01544 [cs, stat]*. arXiv: 1810.01544.

MAGNANI, MATTEO, & SEGERBERG, ALEXANDRA. 2021. On the conditions for integrating deep learning into the study of visual politics. *In: 13th ACM Web Science Conference*.

MAGNANI, MATTEO, MONTESI, DANILO, & ROSSI, LUCA. 2013. Factors Enabling Information Propagation in a Social Network Site. *Pages 411–426 of:* ÖZYER, TANSEL, ROKNE, JON, WAGNER, GERHARD, &
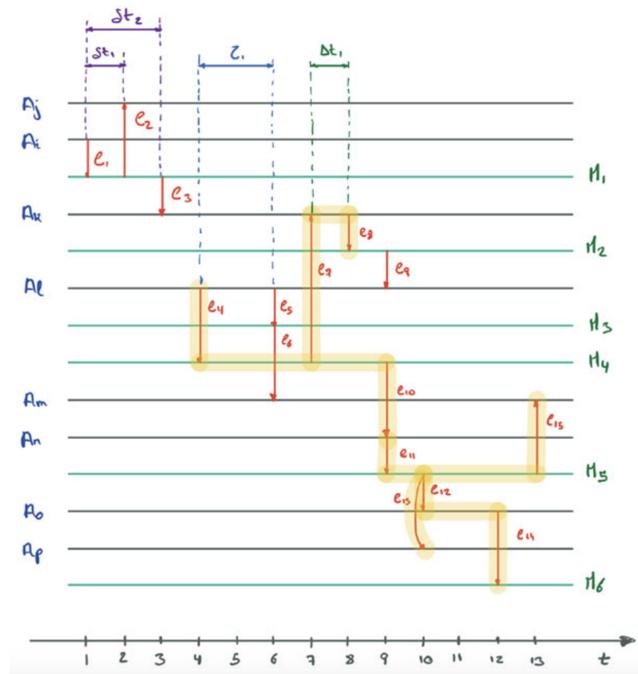
Figure 2. *A model for the analysis of networks and content, from Vega & Magnani, 2019, with A_x representing social media users and M_x representing the content they exchange*

REUSER, ARNO H.P. (eds), *The Influence of Technology on Social Network Analysis and Mining*. Lecture Notes in Social Networks. Vienna: Springer.

MAGNANI, MATTEO, HANTEER, OBAIDA, INTERDONATO, ROBERTO, ROSSI, LUCA, & TAGARELLI, ANDREA. 2021. Community detection in multiplex networks. *ACM Computing Surveys*, **54**(3).

VEGA, DAVIDE, & MAGNANI, MATTEO. 2018. Foundations of Temporal Text Networks. *Applied Network Science*, **3**(1), 25. _eprint: 1803.02592.

VEGA, DAVIDE, & MAGNANI, MATTEO. 2019. Metrics for Temporal Text Networks. Springer, Cham.

ZHANG, HAN, & PENG, YILANG. 2022. Image Clustering: An Unsupervised Approach to Categorize Visual Data in Social Science Research. *Sociological Methods & Research*, Apr., 004912412210826.