Geovisual analysis of VGI for understanding people's behaviour in relation to multifaceted context

Natalia Andrienko [a,b,] *, Gennady Andrienko [a,b], Siming Chen [a], Dirk Burghardt [c],
Alexander Dunkel [c], Ross Purves [d]

[a] *Fraunhofer Institute IAIS, Sankt Augustin, Germany, {Natalia.Andrienko|Gennady.Andrienko|Siming.Chen}@iais.fraunhofer.de*
[b] *City University London, UK, {Natalia|Gennady}.Andrienko.1@city.ac.uk*
[c] *TU Dresden, Germany, {Dirk.Burghardt|Alexander.Dunkel}@tu-dresden.de*
[d] *University Zurich, Switzerland, Ross.Purves@geo.uzh.ch*

* Corresponding author

Abstract:

Volunteered Geographic Information (VGI) in the form of actively and passively generated spatial content offers extensive potential for a wide range of applications. Realising this potential however requires methods which take account of the specific properties of such data, for example its heterogeneity, quality, subjectivity, spatial resolution and temporal relevance. The creation and production of such content through social media platforms is an expressive aspect of human behaviour, and as such influenced strongly by the co-occurrence of events and context external to the social media. In this project we are developing geovisual analysis methods which show how actors interact in location based social media (LBSM), and how their interactions influence, and are influenced by, their physical and social environment and relations.

In the first phase of the project, we developed and demonstrated a conceptual model enabling the extraction, analysis and visualisation of events and reactions to events in LBSM. A central element of this model and its implementation is the integration of spatial, temporal, thematic and social dimensions, or *facets*, combined with an explicit link between events and reactions. We have developed a conceptual model of collective reactions in LBSM [1] which includes a task matrix underpinning our methodological efforts. A key output of this conceptual model, and the resulting task matrix was the acknowledgement of the importance of exploring multiple dimensions in LBSM reactions to events, namely the spatial, temporal, thematic and social which relate to where, when, what and who questions which can be posed of such data.

The conceptual model formed a basis for our research on bridging the gap between visually-driven analysis and visual communication, or story telling [2]. Findings and results of the analysis often need to be communicated to an audience that lacks expertise in visualization and analysis methods. This requires analysis outcomes to be presented in simpler ways than that are typically used in analysis supporting systems. Not only analytical visualizations may be too complex for target audiences but also the information that needs to be presented. Analysis results may consist of multiple components, which may involve multiple heterogeneous facets. Hence, there exists a gap on the path from obtaining analysis findings to communicating them, within which two main challenges lie: information complexity and display complexity. We address this problem by proposing a general framework for story synthesis, in which the analyst creates and organises story contents from analysis results. Story synthesis includes selecting and assembling findings and arranging them in meaningful layouts that take into account the structure of information and inherent properties of its components (facets). Paper [2] proposes a facet-based generic framework for story synthesis which can be applied to different kinds of VGI and LBSM data.

To introduce our concepts, we use an example based on the IEEE VAST Challenge 2011 [3], requiring analysis of the circumstances of an epidemic outbreak in a fictive city Vastopolis. The data are geographically referenced microblog messages, some of which include keywords indicating disease symptoms, such as fever, chills, sweats, aches and pains, coughing, etc. The time span of the data is 3 weeks. An analyst needs to find out when and where the outbreak started and how it developed. The analyst uses a visual analytics system providing multiple types of interactive visual displays and supporting database queries and data transformations. Fig.1 shows how analysis artefacts are managed. In the course of the analysis, the analyst has obtained a set of findings (labelled F1-F5), which include the outbreak start time, the spatial clusters and the times of their existence, the differing sets of frequent keywords associated with the clusters, the location and time of the truck crash, and the ways of spreading and temporal development of two diseases. As the next natural step, these findings need to be communicated to any interested audience, but not as disjoint information pieces but as an integrated story. The pieces need to be arranged in appropriate ways revealing the relationships between the information pieces, such as temporal and spatial relationships. Figures 1C and 1D show examples of arrangements that might be created by the VAST Challenge analyst for conveying temporal and spatial relationships between the findings. Another kind of relationship the analyst may wish to reflect is the differences between the symptoms of two diseases that were discovered in the course of the analysis. For this purpose, the analyst may juxtapose the lists of the keywords

corresponding to the central-eastern and south-western clusters. Analysts should be able to create and edit such arrangements in order to construct understandable and interesting stories.
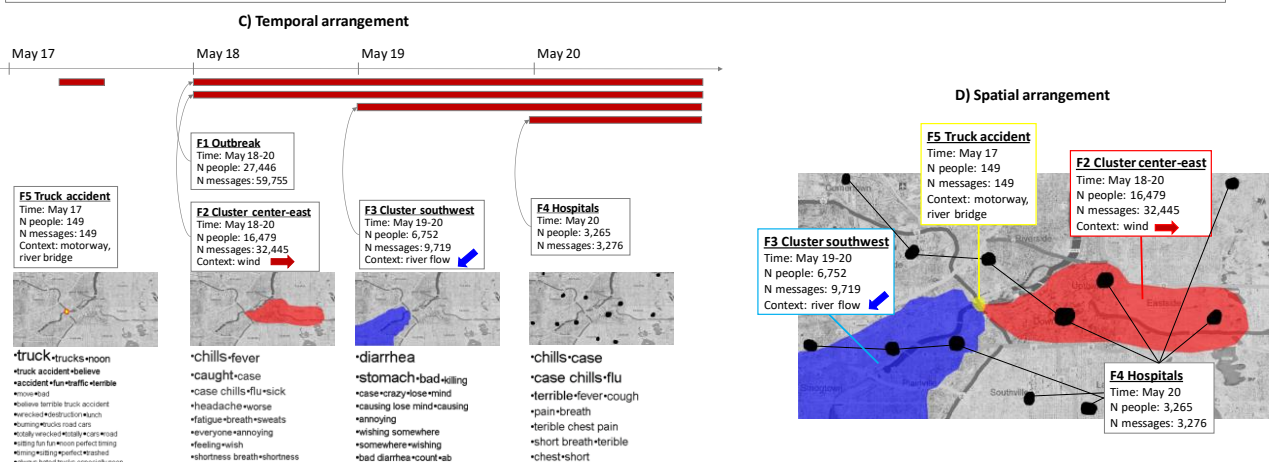

Figure 1. The motivating example of VAST Challenge 2011 analysis outcomes: A) a structure suitable for representing the findings; B) the findings collected in a storage medium; C) the findings arranged along a time line; D) the findings arranged spatially on top of a background map.

A process of story synthesis includes the following activities: aggregate and summarize (as a means of simplification and achieving a desired level of detail), embed details (enable drilling down into aggregates), arrange (put information pieces in a meaningful layout), show facets (exhibit information structure), and annotate (include explanations and comments). Information facets play an important role in story synthesis. They need to be presented to story recipients to enable proper understanding of information. However, heterogeneous facets, such as space, time, population, semantics of message texts (represented by keywords or topics), etc. may be hard to present simultaneously while keeping the display simple and easy to understand and avoiding information overload. Such facets may be represented in complementary views providing different perspectives on the information. The task matrix introduced in paper [1] suggests taking into account only two facets at once. Inherent properties of information facets can be used for meaningful arrangement of story slices and for aggregation. Thus, temporal and spatial arrangements, as in Fig.1 (C, D), exploit the inherent properties of time (temporal ordering and distances) and space (spatial distances, neighbourhood, and relative directions). Paper [2] describes an example of analysing expressions of people's reactions to political events and processes, such as the Brexit, in LBSM and organizing analysis findings in stories with the use of various facet-based layouts.

In the further work, we extend the research scope to studying reactions as a component of behaviour (along with human activities and emotions), incorporating external social and physical context to better allow events to be related and compared. This will not only include development of new analysis methods and workflows but also definition of new analysis tasks and, respectively, new types of analytical results. These extensions will require further work on finding, on the one hand, effective representations for analytical visualizations, on the other hand, expressive and easily understandable representations for communication of analysis findings. Besides, a general problem to be tackled is how to incorporate analyst's input, such as background knowledge and context information that is not reflected in available data, in both analytical visualizations and stories presenting analysis results. It would be interesting to go beyond mere use of textual annotations towards representing such inputs in a visual form, which needs to be distinguishable from the representation of the data.

References

[1] A.Dunkel, G.Andrienko, N.Andrienko, D.Burghardt, E.Hauthal, R.Purves. A conceptual framework for studying collective reactions to events in location-based social media. International Journal of Geographical Information Science, 2019. https://doi.org/10.1080/13658816.2018.1546390

[2] S.Chen, J.Li, G.Andrienko, N.Andrienko, Y.Wang, P.H.Nguyen, C.Turkay. Supporting Story Synthesis: Bridging the Gap between Visual Analytics and Storytelling. IEEE Transactions on Visualization and Computer Graphics, 2019 (accepted)

[3] G. Grinstein, M. Whiting, K. Liggett, and D. Nebesh, "IEEE VAST Challenge 2011" http://hcil.cs.umd.edu/localphp/hcil/vast11/