

Visual Analytics for Integrated Evolution of Physical and Cyber-Events: A Case of the World Cup in Social Media

Siming Chen*, Xing Gao, Jie Li†, Gennady Andrienko, Natalia Andrienko‡

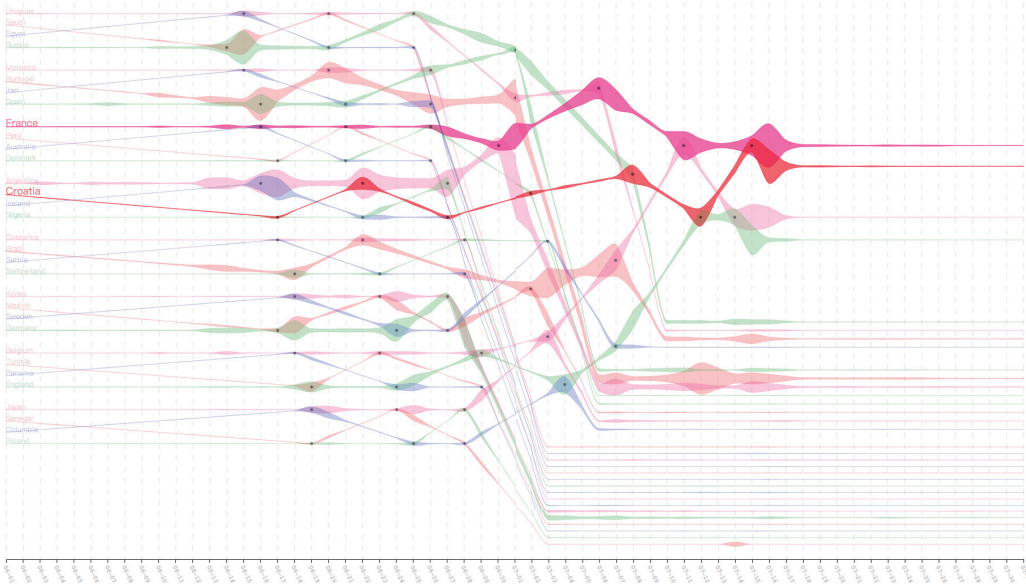


Figure 1: DualFlow visualization. Lines represent entities in the physical world (teams participating in the World Cup). Line widths encode the amounts of the corresponding social media discussions. Two finalists of the World Cup, France and Croatia, are highlighted for tracing their relationships to the corresponding social media discussions.

ABSTRACT

Significant events, such as large sport tournaments or presidential elections, develop over a long time in the physical world. In parallel, people's reactions to these events expressed in social media evolve in the cyber world. We are interested in exploring dynamic relationships between these two processes. We propose a *DualFlow* visualization, which uses a storyline metaphor for presenting the evolution of a real-world event arousing discussions in social media together with the evolution of these discussions. We illustrate our approach by example of social media discussions of the football World Cup 2018. DualFlow visualizes the course of the whole championship including games of all 32 teams and shows the evolution of the related reactions of social media users, i.e., how the interests of the users were distributed among the teams and how this distribution changed over time. Analysts can brush a period and select teams to explore the changing focus patterns of social media users as the tournament goes on. Major keywords of the discussion are shown on demand for explaining the content of the discussion.

*S. Chen, G. Andrienko, N. Andrienko are with Fraunhofer Institute IAIS, Germany. {siming.chen, gennady.andrienko, natalia.andrienko}@iais.fraunhofer.de. S. Chen is also with Bonn University, Germany. G. Andrienko, N. Andrienko are also with City University London, UK.

†X. Gao and J. Li are with Tianjin University, China. jie.li@tju.edu.cn

‡The work is supported by National NSFC project (Grant number 61602340) and it is also supported by German Priority Research Program SPP 1894 on Volunteered Geographic Information, EU project SoBigData, and Fraunhofer Cluster of Excellence Cognitive Internet Technologies.

1 INTRODUCTION

In a stadium, teams are fighting hard while thousands or millions of people are watching the game at the stadium or on TV or are listening to live broadcast. Many of these people express their impressions in messages posted in social media services such as Twitter. They keep discussing the game, players, goals (if any), mistakes, referees, rumors, etc. before, during and after the match. Physical-world events (**PWE**) and cyber-world events (**CWE**) happen in parallel, motivating our research to investigate the relationship patterns in their evolution. Specifically, we want to address the following research questions: 1) How do social media react to physical-world events? 2) How does the social media reaction change when the physical-world event changes?

Answering these questions is challenging because of the following reasons. First, the relationships between PWE and CWE are complex since a PWE may involve multiple interacting participants, threads, and/or sub-events, and a corresponding CWE is also a complex event that can be seen as a distribution of a large number of elementary events (i.e., individual posts) over the components of the PWE that changes over time. Second, not only the distribution of the elementary events changes but also users' opinions. To address these challenges, we propose a novel visualization DualFlow, visualizing related events from physical- and cyber-worlds in an integrated manner. By interactive filtering and selection, analysts can explore event evolution and variation of opinions, focusing on changes patterns.

Storyline visualizes evolution of relationships among entities in a story [4], which is the starting point of our approach. Social media visual analytics is an important research domain with a rich state-of-the-art [3]. However, the existing works either only addressed the cyber-world, namely people's reactions to some events that were expressed in social media [2], or investigate the evolution of the

physical-world events, such as football games [1]. Different from the existing works, we address the integrated analysis of evolving physical- and cyber-world events and relationships between them.

2 DUALFLOW VISUAL ANALYTICS

We propose DualFlow, a visualization that integrates PWE and CWE and shows the evolution of the events. To further enable the exploration of the relationship of entities in PWE and CWE, we propose a visual analytics workflow in which the analysis is supported by interactive selection and highlighting techniques.

In the illustrating case, we acquired two data sets, namely the match events of WorldCup 2018 (PWE), and the explicitly related Weibo messages, specifically, the messages having the hash tag “#WorldCup” (CWE). Weibo is the largest Chinese Social Media Service, similar to Twitter, with billions of active users. We acquired 67,871,823 related messages during the month of the World Cup. After data cleaning and removal of advertising accounts and bots, we have 7,653,986 messages from 3,297,502 social media accounts.

DualFlow Overview To support the analysis of the evolution of multiple entities within a PWE, we use a visual metaphor similar to the storyline [4]. In DualFlow (Figure 1), each entity (in the World Cup example, each team) is a line going from left to right along the x-axis which represents physical time. When two or multiple entities interact in the physical world during some period, their lines converge and go parallel along that period. In the World Cup example, the converging lines correspond to matches between teams. It visualizes the structure of the tournament and all games of all 32 teams. The social media messages mentioning each of the teams are aggregated at the daily resolution. The widths of the lines indicate how many Weibo messages mention the country or the players of the country. The color is designed to differentiate the teams in the group rounds. In Figure 1, though Croatia reached the final of the World Cup, it started gaining noticeable attention in the social media only after the knock-out rounds. In the final rounds both France and Croatia gained great attention in the social media. This fact is indicated by wide lines in the ending phase of the tournament.

Exploration of Event Dynamics We designed several interaction mechanisms for enabling further exploration. Analysts can select time and teams of interest by brush, which supports comparison of dynamic social media popularity patterns for the selected teams. We can conduct visual comparison for a single team (period-wise comparison) and across multiple teams (team-wise comparison). In our example, teams with initial high expectations to win (Portugal, Germany, Brazil, and Argentina) were popular in the beginning. After these expected winners were eliminated, they and their players were still mentioned in the social media. The peak patterns found in the exploratory study can be identified in the Argentina line.

Dual-space Analysis By clicking on a chosen team and brushing a time interval, we select the subset of social media users that discussed this team at the selected time. Social media users can be metaphorically considered as bridges that connect teams. We show which proportions of the total volumes of the discussions concerning the other teams are covered by the selected users. We use yellow highlighting curves with controlled opacity to overlay on the original lines (Figure 3). To enable thematic interpretation, we extract the keywords and hashtags of the selected messages. We calculate Term Frequency-Inverse Document Frequency (TF-IDF) value of keywords and hashtags and visualize them with a wordle (Figure 2).

One highlight in the illustrating case was an unexpected result of the game between Iceland and Argentina (Figure 2). Due to this event, Iceland gained much greater attention in the social media than usual. These social media users had high expectation on Argentina and Messi, but the performance of Iceland won their applause. From the wordle, we see that social media users think “Iceland is a miracle”. In this case, we see how a significant sub-event within a PWE (a game) can shift interests of social media users, change their

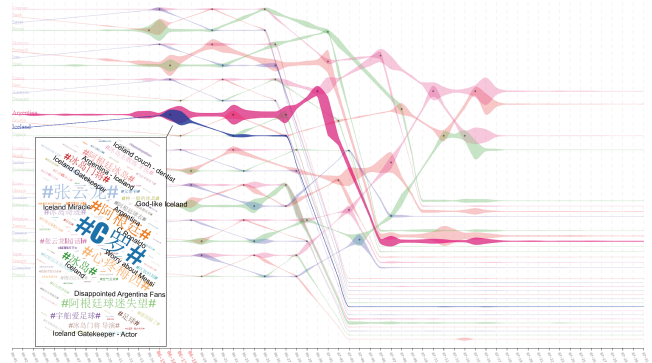


Figure 2: The wordle of hashtags for a selection of an entity and a time period. Iceland gained a peak attention in social media.

opinions, and incite particular expressions of their reaction.

In the same tournament, Germany was attracting a lot of attention at the beginning but got an upset loss. Thus, we are interested in how social media users shifted their attention from Germany to others teams. We select the period of Germany’s active participation during the group stage and retrieve the social media users who discussed this team. We can see which other teams were interesting to the people who talked about Germany (Highlighted in yellow). Some of them shifted their interests to other teams and commented on further games. Thus, we could identify the hidden relationships among entities based on developments in both PWE and CWE.

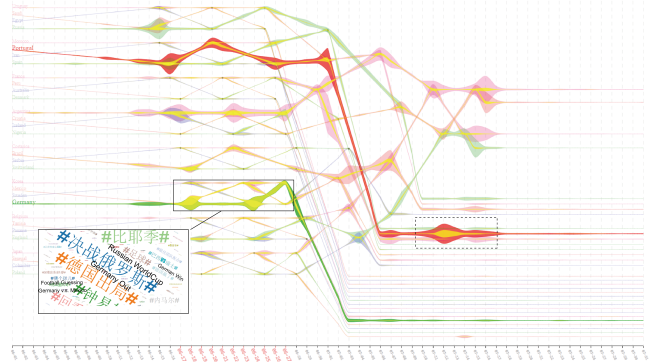


Figure 3: People who discussed Germany from June 17 to June 27 are selected. Their participation in discussions concerning other teams during the whole world-cup period are highlighted in yellow.

3 DISCUSSION AND CONCLUSION

Events in physical- and cyber-worlds affect each other and are worth to be investigated together. We do not know yet the driving forces and mechanisms of the complex relationships between PWE and CWE. This work is a starting effort to address these issues. We want to extend our approach by varying the level of temporal granularity, investigating the discussion topics, and assessing sentiments.

REFERENCES

- [1] G. Andrienko, N. Andrienko, G. Budziak, J. Dykes, G. Fuchs, T. von Landesberger, and H. Weber. Visual analysis of pressure in football. *Data Mining and Knowledge Discovery*, 31(6):1793–1839, 2017.
- [2] S. Chen, S. Chen, L. Lin, X. Yuan, J. Liang, and X. Zhang. E-Map: A visual analytics approach for exploring significant event evolutions in social media. In *Proc. of IEEE VAST*, pages 36–47, 2017.
- [3] S. Chen, L. Lin, and X. Yuan. Social media visual analytics. *Comput. Graph. Forum*, 36(3):563–587, 2017.
- [4] Y. Tanahashi and K.-L. Ma. Design considerations for optimizing storyline visualizations. *IEEE TVCG*, 18(12):2679–2688, 2012.