

COLORS OF THE WORLD CUP

Colors of the World Cup: visualizations of images shared on Twitter during the 2014 World Cup

Fábio Goveia, Lia Carreira, Lucas Cypriano, Tasso Gasparini, Johanna Honorato, Veronica

Haacke, Willian Lopes

University of Espírito Santo, Brazil

Professor in the Department of Social Communications
in the Federal University of Espírito Santo, Brazil and
coordinator of the Laboratory of Studies on Image and
Cyberculture.

This research was supported Research Support Foundation of Espírito Santo – Brazil (FAPES)

Correspondence concerning this article should be addressed to Fábio Goveia, Department of
Communications, Federal University of Espírito Santo, Av. Fernando Ferrari, Goiabeiras,
Espírito Santo, Brazil. E-mail: fabiogv@gmail.com. Corresponding author Lia Carreira -
liacarreira@gmail.com; Lucas Cypriano - lucascypriano@gmail.com; Tasso Gasparini -
tassogasparini@gmail.com; Johanna Honorato - johannahonorato@gmail.com; Veronica Haacke
- veronica.rhaacke@gmail.com; Willian Lopes - willianlopes42@yahoo.com.br

COLORS OF THE WORLD CUP

Abstract

The 2014 World Cup, held in Brazil from June 12 to July 13, had great impact on social networking sites. Seeking to better understand content sharing flow and to explore methodological possibilities for studies with large volumes of images data, the Image and Cyberculture Studies Lab (Labic), of the Federal University of Espírito Santo, created the project Colors of the World Cup - a project that presents interactive image visualizations of contents shared on Twitter during the football tournament. Throughout this paper, we will present the project's methodological process and propose a brief analysis of these images as a way of documenting the social and cultural movements and events that appeared on Twitter during the World Cup.

Keywords: image, visualization, Twitter, color

COLORS OF THE WORLD CUP

Colors of the World Cup: visualizations of images shared on Twitter during the 2014 World Cup

The large amount of information produced by each individual through mobile devices can be considered one of the most significant marks of contemporary society. On the one hand we are producing these data, on the other hand we are also sharing, circulating and consuming constantly these information. Digital social networks are among the most commonly used ways to consume and circulate personal and public news. These platforms have become privileged interaction zones, with more and more exchanges of information and experiences.

This new universe opens up numerous opportunities for knowledge and relationships - that once seemed nearly impossible to happen – and also makes it possible to develop research analyzing these interactions and weaving more precise understandings of the social behavior of individuals.

In this scenario new experiences among users emerge and important tools of activism and change get created. In addition to the significant role of social networks in the political field, it should be also be highlighted the significant amount of information produced in relation to other large social events (sports, cultural, artistic). Data from these social networks related to these events can contribute significantly to different studies, such as consumer behavior.

From this overview, the analysis of the ways in which such content spread through the networks is necessary for a better understanding of how these social movements develop and behave, in and out of these digital environments. Furthermore this analysis can help to identify potential patterns or future occurrences. Thus the big data studies appear as a field of study that has been highlighted in recent years, especially due to the expansion of this large amount of

COLORS OF THE WORLD CUP

information circulating in the environment of digital social networking, and that has been used for several studies, as the one presented here.

Seeking investigate new methodological frameworks, the Laboratory of Research on Image and Cyberculture (Labic), started the World Cup Colors project, which sought to analyze the production of images that occur in social digital networks during the 2014 World Cup held in Brazil. The event took place between June 12 and July 13, 2014, with a great impact on social networks. On this occasion, we have seen emerging a large number of images that were critical or made fun of the event, giving way later to a more favorable outlook. This change was attributed to the network called "Boleira" that is, the network of those fan profiles that accompany the real-time games, the backstage and the memes about the World Cup. They significantly contributed to the expansion of the online debate to a more positive and less critical bias.

Before the Cup a strong tension was expressed by various popular protests in several Brazilian cities. The first analysis of the images on social networks revealed that this tension ultimately was displaced by the accounts of the football fans which dominated the sharing of images on Twitter. No other perspective occupied the media spaces network in the same way as the football fans profiles. At one point, even some activists that were complaining about the FIFA impositions to the country, and corruptions with the event, started communicating about the games and became part of this football narrative.

In the following, we present the methods that guided the development of this research, as well as web application development, launched on October 6, 2014 in order to provide viewing modes for large sets of images collected during the World Cup.

Methodology

Manovich and Hochman (2013); Manovich et. al. (2011); Crandall et. al. (2009 e 2015) are some of the authors that work in this area, from social networks data mining to development of big data image visualizations. The challenges proposed by them were also part of this research. In this research were collected images shared on Twitter. In this case, data was collected through World Cup¹ related terms such as specific keywords, hashflags² of each country, terms of each match (for example, #BRAXGER used for the game between Brazil and Germany), and the names of the major soccer players, such as "Neymar," "Messi" and "Podolski". Here, the aim was to collect, from June 12 to July 13 of 2014, the largest number of tweets related to the event, using precise terms and related keywords.

The tweets published during this period were collected from a script called Marcus, developed in Python by Labic in partnership with the computer scientist André Panisson. This tool searches the tweets containing text terms previously selected and stored in a remote server database (MongoDB³). After this first mining, every 15 minutes of collect by Marcus script another script called Crawler⁴, also developed by Labia, accessed the database, captured the tweets of that period containing links (possible images), eliminates those images links resulting from external sites, considering in this way just the images that was posted directly in the tweet. After this the Marcus script selected the 100⁵ most shared images during the selected period - this mean, the images links most repeated during the selected period, in this case each 15 minutes. Thus the script accessed each link collected, saved the image attached and generated a .csv table listing the link of the image and the saved file for further analysis.

With the image dataset ready in hand, it was then possible to proceed with the data processing and image analysis, done in two simple steps: first the extraction of chromatic data and, second, image comparison in order to eliminate repeated ones. Therefore, at the end of each day's data mining process, a third script also developed by the lab called AISI (Automatic Identifier of Similar Images) swept the dataset, extracted and compared the image's colors metadata (such as hue, brightness and saturation, as well as each image's histogram), and identified those that were similar to each other, avoiding repetition. When similarities were identified, these similar images are considered as single ones, adding their frequencies (i.e., the amount of time they were posted) as to give them their correct weight in the dataset. Without this procedure, each one of these similar images, published in different tweets, wouldn't have the

COLORS OF THE WORLD CUP

impact they actually have on these networks through online sharing - a very important characteristic of social media content.

With this extraction method, we obtained a total of 30 million tweets, of which 2 million contained links with images, and a total of 42,000 images posted directly on Twitter. Among these, there were 17,000 unique images that were then used to create the visualizations. With these data, it was possible to launch a web application called Colors Of the World Cup that facilitates reading this diversity of content by organizing it according to previously collected data, such as their chromatic characteristics and frequency of sharing.

The online interface of the application was developed using the JavaScript programming language, and the D3⁶ library. From the .csv files generated by the Crawler and AISI, the engine retrieves information to generate the views and display them in the web application. With this data four different visualizations are generated: chromatic calendar⁷, chromatic timeline⁸, chromatic mosaic⁹ and images mosaic¹⁰.

The chromatic calendar (figure 1) brings daily charts, containing circles as representations of each image from its predominant color. In this view they are counted only the shares obtained by the image in each day. The circle (that represent a unique image) position on the X axis is defined by the predominant hue of the image that its represents, while the position in the Y axis is defined according to the number of shares that the image obtained during that period; the radius of each circle brings the total number of shares that the image accumulated over time; and the color of the circle is defined by the predominant hue of the image. Hovering the mouse over each circle shows the image, and click in the view starts an animation that shows the behavior (sharing number) of images throughout the day in 15-minute intervals.

The chromatic timeline (figure 2) is similar to chromatic calendar, but now shows on the same graph the total volume of the images during all days of the extraction period, and with this allows to be observed the behavior of each image along all the days. Same as the chromatic calendar the X axis position and the circle color is defined by the predominant hue of the image that its represents. But now the Y axis position is defined according to the number of shares that

COLORS OF THE WORLD CUP

the image obtained during all one day, and the radius of each circle brings the total number of shares that the image accumulated over the days. Clicking the graph shows an animation with the behavior of images in each day. You can also pause or fast forward the animation to view the images more closely every day, and skip to specific days.

The chromatic mosaic (figure 3) is a more straightforward view that brings the images represented by smaller circles with their respective predominant color and organized by its color hue. And hovering the mouse over each circle shows the original image.

Finally, the image mosaic (figure 4) is the last visualization, in this is displayed the own images organized by their predominant hue color value, with a option of zoom in and out to have a more overview of all images or a more close visualization for a specific image. It is important to note that this is the only visualization of the application that uses the stored images¹¹ from the dataset, while the others visualizations access the original link of the image to show those in real time. This means that images in these others views will gradually fail to appear, since their source links may in time cease to exist. Thus, these views reflect the vital movements in networks.

Likewise, downloading and storing the images is also a way to keep these vivid images as memories of a movement or major event. Working with these large datasets is to work with their transitions and movements over a period of time, and create views is one way to realize these processes and make them visible.

Analysis

The Colors of the World Cup visualizations aims to highlight certain features of the shared content on the network during the 2014 World Cup. Through them, it is possible to highlight, from a wide range of content, key aspects about the way of images sharing in social

COLORS OF THE WORLD CUP

network during the collection period. By analyzing these views, an intense chromatic variation of the images can be observed. A large volume of green tone images is noticeable, partly because the green color of the grass (since most of the photos with this predominant color tone are scenes that happened on the field) and partly because the presence of some Brazilians flags. There is also a considerable number of images in blue shades composed of uniform images such as the Argentina selection uniform with predominant blue colors, and a large number flags and advertising materials. In contrast, the more yellowish region, there is a large volume of images of materials related to the Brazilian national team, notably the yellow shirt traditionally used in games.

In the area with more red tones, we see mostly pictures of people, due to skin tones approaching beige and red. There is still a small part with pink predominant tone images, composed largely of images published with filters (like the ones used in the Instagram application), which pull the image tone to pink.

Beyond this evident chromatic variation, shared oscillations by day can be observed during the World Cup. This shows that the impact of the event on the networks did not occur in a homogeneous way. On the contrary, in a single day there was a very large amount of images with radically different levels in popularity on the network. The high frequency of an image in a day did not imply a similar level of popularity the next day, as happened with the eagle image (figure 5) with the US flag on 22 June¹², being the third most shared image of the entire Cup, obtaining total 22,098 posts just on the first day of sharing. This eagle image represented the discussions taking place on Twitter during the US game with Portugal, highlighting the broad participation of the North American and other supporters throughout this day. This demonstrates,

COLORS OF THE WORLD CUP

therefore, that the image sharing modes during the World Cup were related to the specificity of the events of each day.

The most shared image throughout the period appeared on the network a few days before the end of the World Cup, on 5 July. It is an image of Neymar (famous soccer player) (figure 6) kissing the ball, with a yellow mount in the background. The image was shared too by player David Luiz after Neymar suffered a spinal injury and cumulated a total number of 38,490 shares. This represents the support of fans and players in relation to their concerns for Neymar and was a very present image from the days of the event, showing the sentimental side of the players. Similar images in which players appear crying or celebrating a victory or defeat reinforce the emotional tone that the shared images had during the World Cup.

Thus, these images underline the existence of an emotional outlook: images shared by the network reflect the feelings of social networking users in relation to the championship and the participating teams. Images with scenes of the competition, as photo of players being emotional in the field, were used to express the feelings of users. This is most clear in the images most shared between the 8th and 9th of July, which brought scenes of sadness resulting from the defeat of Brazil to Germany by 7-1, and in the second shared image (figure 7) between the 12th and 13th June.

The wordcup shared images also have strong memetic content, ie, were created to become memes (images with the potential to multiply and propagate quickly, and can be given new meanings in the sharing process). These memes dominated the images on the networks that have formed around topics related to the World Cup, especially during the event first week, and remained largely shared together with pictures of other events that drew public attention (such as the elimination of teams considered "strong" or the Brazilian defeat to Germany, for example).

COLORS OF THE WORLD CUP

Thus, these visualizations made possible a rich analysis with the identification among other characteristics, of the existence of a chromatic rhythm. The images pulses and dances, even if at times, out of step and rarely able to maintain a constant frequency. This rhythm is symptomatic of the incessant and extremely fast content sharing process, characteristic of our time, that makes images suddenly appear, rise and quickly fall, preventing them to remain on the spotlight for too long. Thus, we can say that these social network images compete with each other for a spot to be seen, to be liked and to be shared - before they lose their sparks. And thus we can conclude, by watching these images through the visualizations created for this projects, that the images on Twitter posted during the World Cup were, above all, short-lived devices fighting for visibility.

Final Considerations

Our studies with large quantities of images still have a long way to go. With the current research, Labic has tried to bring question as well as contribute with new and already existing image visualizations methods. In a time where these data are easily produced and shared on every moment, initiatives that try to record these MOVEMENTS are even more needed, so narrative paths and rhythms can be discussed, despite the quick and uncertain life of the artifacts under consideration. Through Colors of the World Cup, we can observe the traces of images published in a network and try to determine its behavior, and by way of its vestiges comprehend its movements, apparitions and reapparitions, and longevity afterall.

However, there are still obstacles that tend to discourage this kind of research. The processing power needed for the conceptualization and execution of projects of this magnitude is very large: many files are used, and the large volume of images can lead machines to have a very

COLORS OF THE WORLD CUP

low performance. A stable and high speed internet connection is also an important issue to insure that the image crawling can proceed without interruption.

Another constant problem faced in this kind of study is “bots”, robot-profiles that automatically share content and almost always in a massive way. This kind of automatic process can make specific network's threads or subjects seem to have a high relevance, figuring between the most shared ones, though really their organic circulation is low: they are not shared by the network's "common" users. For this reason it is crucial to have a watchful eye for bots, so they don't taint the analysis.

Colors of the World Cup provides an experimental method for data extraction in real time and introducing new image visualizations possibilities, offering a contribution to the cyberculture and image study fields and presenting novel ways to analyze image data.

COLORS OF THE WORLD CUP

References

- Crandall, D. J., Backstrom, L., Huttenlocher, D., & Kleinberg, J. (2009, April). Mapping the world's photos. In *Proceedings of the 18th international conference on World wide web* (pp. 761-770). ACM.
- Lee, S., Maisonneuve, N., Crandall, D., Efros, A. A., & Sivic, J. Linking Past to Present: Discovering Style in Two Centuries of Architecture. In *IEEE International Conference on Computational Photography*.
- Haacke, V. , Ribeiro, A. et. al. Do “Não Vai Ter Copa” para a “Copa dos Memes”: Uma análise das imagens memes mais compartilhadas durante a Copa do Mundo FIFA 2014. Brazil, Foz do Iguaçu: Intercom, 2014.
- Hochman, N., & Manovich, L. (2013). Zooming into an Instagram City: Reading the local through social media. *First Monday*, 18(7).
- Manovich, L. (2012). How to compare one million images. *Understanding digital humanities*, 249-278.

Footnotes

¹ Terms were used that made reference to the main event, as "world cup", "copa2014", "brasil2014", "worldcup"; the side events directly related to WordCup as "Funfest", "wordcup opening"; and the organizations responsible for the event, such as "fifa" and "cbf".

² Hashflags were a special kind of hashtag used on Twitter during the World Cup. Were hashtags related to the countries participating in the World Cup,

³ Open source database, developed in C ++ programming language. The system generates an information store in JSON documents, a data file type that works independent of any specific programming languages. More information available at: <http://www.mongodb.org/>

⁴ Script developed in Java, that accesses the selected links and downloads of these the images found. Available code: <https://github.com/ufeslabic/crawler>.

⁵ For to do the extraction of all images published would require a long time, it would preclude the 15-minute time window..

⁶ D3 (Data Driven Documents) is a JavaScript library designed to handle large data files for online visualizations. More information available at: <http://d3js.org/>.

⁷ Available at: <http://labic.net/coresdacopa/calendariocromatico/>

⁸ Available at: <http://labic.net/coresdacopa/timelinecromatica/>

⁹ Available at: <http://labic.net/coresdacopa/mosaicocromatico/>

¹⁰ Available at: <http://labic.net/coresdacopa/mosaicoimagens/>

¹¹ Because of the large size of the image and to possibility zoom the view, the ZoomIt tool, developed by Microsoft for viewing very large size images on the Internet has been used. More information available at: <http://zoom.it/>.

COLORS OF THE WORLD CUP

¹²In that day there was an unexpectedly tough match between Portugal and the United States, which ended in a 2-2 draw, which had great impact on the web.

COLORS OF THE WORLD CUP

Figures



Figure 1. Chromatic calendar: days 12-20 of June. Best viewed through the site.

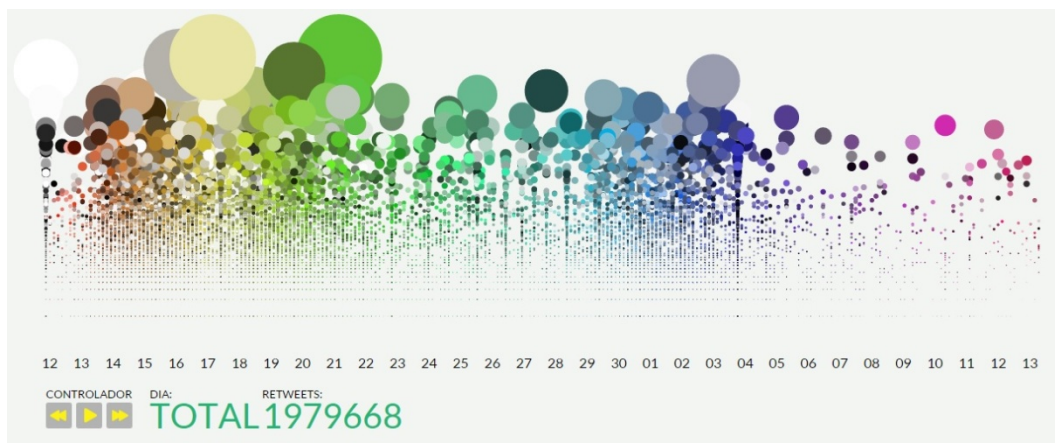


Figure 2. Chromatic timeline. Best viewed through the site.

COLORS OF THE WORLD CUP

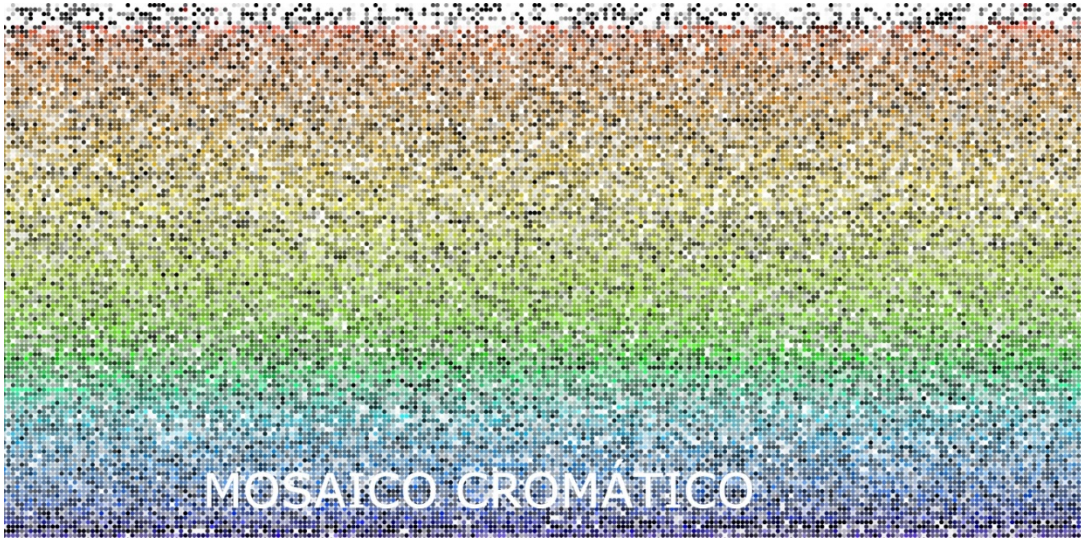


Figure 3. Chromatic Mosaic. Best viewed through the site.



Figure 4. Images Mosaic. Best viewed through the site.

COLORS OF THE WORLD CUP



Figure 5. Most shared image of June 22. Author (s) unknown (s).



Figure 6. Image composition with Neymar player, was the most shared in the period of the World Cup. Author (s) unknown (s).



Figure 7. David Luiz (player) photo being consoled was the second most shared image in the day after the Brazilian defeat to Germany. Author (s) unknown (s).