

# *Cultural networks and the future of cultural analytics*

Juan-Luis Suárez, University of Western Ontario  
Modern Languages and Literatures & Computer Science  
The CulturePlex Lab  
London, Canada  
[jsuarez@uwo.ca](mailto:jsuarez@uwo.ca)

Ben McArthur, Simon Fraser University  
School for International Studies & The CulturePlex Lab  
Vancouver, Canada  
[bmcarthu@sfu.ca](mailto:bmcarthu@sfu.ca)

Adriana Soto-Corominas, University of Western Ontario  
Modern Languages and Literatures  
The CulturePlex Lab  
London, Canada  
[asotocor@uw](mailto:asotocor@uw)

**Abstract**—The objectives of this paper are to point out the different trends prevalent in the analysis and study of culture by digital means, to define the concept of “cultural network”, and to explain why this concept can help us to understand the past and to solve social and economic issues facing our modern world.

**Keywords**—cultural networks; cultural analytics; graph databases; SylvaDB; culturomics

## I. CULTUROMICS: OBJECTS

The term "Culturomics" was officially coined in 2011, with the publication of an article titled "Quantitative Analysis of Culture Using Millions of Digitized Books" in the journal *Science* [1]. The article defines "Culturomics" as "the application of high-throughput data collection and analysis to the study of human culture," pointing out that though the study focused solely on books, there is huge potential for Culturomics techniques to be applied to all cultural products. A team of researchers based at Harvard University created a multilingual corpus of some 5 million digitized texts, which were then subjected to computational analysis in an effort to pinpoint linguistic and distributional patterns across different subsections of the dataset. By undertaking such a task, the researchers intended to demonstrate that there are many unexplored opportunities to apply quantitative methods to the analysis of cultural products and patterns.<sup>1</sup>

The main characteristics of Michel et al.'s work are the study of text found in books, the fact that their applications of Culturomics techniques are focused on the past, and the lack of social contextualization provided by the study, at least in terms of the idea of the social network. With respect to the first element, the study of texts, the authors focus on the study

of n-grams (in this case, meaning *contiguous strings of n characters that can be pulled from a digitized text*) that, in turn, are provided to researchers through the Google Ngram project. The study of texts using n-grams constitutes an approach to cultural granularity that is radically new and disruptive for most researchers who typically use traditional humanistic methodologies [3]. Granularity, size of information, as well as time and linguistic extension are combined in order to reach a scale that is unprecedented in our approach to human history and culture.

The studies of Culturomics that focus on only one type of cultural object go beyond just texts, although it is true that textual research has quantitative prevalence [4]. Our lab's research team has focused its efforts on other cultural artifacts, particularly paintings. In a recent article titled "A Quantitative Approach to Beauty: Perceived Attractiveness of Human Faces in World Painting," we carried out a study of the evolution of beauty in the representation of human faces in more than 120,000 paintings of different periods of art history [5]. The main goal of this study was to determine whether a single canon of beauty had prevailed between the 13th and 20th centuries or whether it had changed over time. With the help of several measures (averageness, symmetry, and orientation), this study proved that the representation of human faces had not remained constant through history.

Similar studies focused on other types of objects [6] share with Michel et al. a methodological combination by which advanced computational techniques (such as the combination of facial recognition and feature extraction with machine learning) allow us to take a more precise and extensive look at human history from the perspective that a specific cultural object offers.

## II. CREATORS AND NOTABLE PEOPLE

Our fascination with human culture stems from the combination of sheer admiration for creations of the human spirit -Beethoven's *9th Symphony*, Cervantes' *Don Quixote*, or Hokusai's *The Great Wave*- and for the geniuses that made them; those who occupy a special position in humankind

<sup>1</sup> Other articles using the term and methodology since this was published include: Kalev Leetaru's "Culturomics 2.0: Forecasting Large-Scale Human Behavior Using Global News Media Tone in Time and Space," Gao et al.'s "Culturomics Meets Random Fractal Theory: Insights into Long-Range Correlations of Social and Natural Phenomena over the Past Two Centuries," and Klaas Willems' "Culturomics' and the Representation of the Language of the Third Reich in Digitized German Books."

thanks to their creativity and ability to capture the essence of humanity.

Yu et al., for example, approach the study of the geniuses of human culture production from a thoroughly contemporary perspective [8]. For them, the concept of impact across linguistic and cultural boundaries is the most important one. In order to quantify that impact, they have tried to determine who are the most relevant figures of human history -a domain that until recently was mainly localized, since it was usually only studied within a specific linguistic or political boundary-measuring what figures are more recognized across these boundaries.

With the creation of their “Pantheon” dataset, Yu et al. made use of quantitative methods and manual curation in order to link several different metrics within the world of cultural production. In order to do so, the researchers created a classification system wherein individuals are organized based on their contributions to the world of culture. The project’s data was drawn from Freebase and Wikipedia, including all 277 language editions of the online encyclopedia, before being grouped based on “time periods, cultural domains, and geographies.”

The researchers also introduce a new metric in the paper, the Historical Popularity Index, which determines the importance of one’s contribution to culture based on pageviews and the age of the individual in question within the dataset. This ranking system can also be configured to reflect individuals’ significance within a certain time period, cultural domain, or geographic area.

Schich et al. also focus on individuals in order to develop a data-driven macroscopic perspective, trying to bring two observation scales together so that they are unified under one unique framework, which they call the “network Framework of cultural history” [9]. What is truly special about this framework is not the individuals that are found in the corpus (all of whom are artists or other notable individuals), but the irregularities that are observed in the behavior of these individuals when confronted with the statistical regularities that are found in the dataset.

The authors focus fundamentally on the dates and places of birth and death, as well as the concept of migration in order to analyze a group of notable individuals of the last 2,000 years who moved within Europe and between Europe and North America. Their analysis is extremely important, because it is the first time calculations on how far in average creators moved from their birth places, have been made on such a global scale, and because it is an indicator that helps us understand how human culture is transmitted.

Distance and the studies that show remains of human culture in places that are far from where they were initially produced [10], suggest that although the mobility of notable individuals is important in terms of concentration of talent, there are other considerations that had not been taken into account until this moment: It is not notable individuals who contribute the most to dispersion in the geographic reach of culture and ideas. Rather, we need to identify other types of human groups and demographic cross-sections (the *weak links*

of cultural networks) whose role in the dispersion, recreation, and adaptation of ideas across cultural ecosystems allow us to better understand these phenomena.

### III. CULTURAL NETWORKS

A specific strand of digital research into human culture focuses on the study of particular cultural objects -texts, books, artistic objects, or music- with an emphasis on topics that are related to textuality and language. Another strand focuses on human beings who created these artifacts. It is crucial that the study of objects and people be connected in a way that ensures this emerging body of research is relevant beyond the academic domain of digital humanities. Likewise, researchers must be able to exploit all of its scientific, political and economic potential to understand different questions regarding the past and the present of cultural evolution, migrations, world flow of ideas, and collective behaviors around social media [11].

With this purpose in mind, we propose the concept of “cultural networks”. A cultural network is a multimodal network in which at least two types of the nodes represent “persons” and “objects”, where the links between nodes are semantically loaded with information that is contextually relevant to the domain of research. A cultural network is an abstract construction that connects creators with their own objects, and these objects with other human beings that have come into contact with them by means of proximity, reading, contact, influence, reception, or whichever link is necessary to understand the human context that the network is reconstructing -the past- discovered -the present, or anticipating -the (near) future.

A cultural network differs from a social network (at least in the way they are most often thought of) in that the latter is composed of nodes directly linked by edges that represent sociologically relevant connections between human beings. A cultural network is more complex in that the human relationships it contains are mediated by objects and phenomena that are considered cultural in a given group or community. This means that while the minimum distance between two human beings in a social network is equal to one –“John” ‘is the father of’ “Mary”–, the minimum distance between two human beings in a cultural network is always two links and one node –“Picasso” ‘was influenced by’ “The Family of Philip IV” ‘painted by’ “Velazquez”. In many cases, distances may be much longer and relationships much more complex within cultural history. Such inherent complexity in cultural networks is an integral part of their value to humanity. These cultural networks also help us to better understand the trajectories by which creators have historically sought out inspiration in a varied ensemble of both contemporary or previous creations and artists. Finally, it constitutes one of the nuclei of human cultural transmission beyond the aesthetic domain.

The publishing network of literary works in Spanish in the 17<sup>th</sup> century, known as the Spanish Golden Age, is an example of cultural network. This specific network may be constructed from a very specific cultural object: the “preliminary”



cultural networks. In terms of research that may be applied to both computer science and disciplines that make extensive use of the network as a conceptual tool -neuroscience, genomics, culturomics- they depend on the progress on this research to be able to better analyze large networks. Even more importantly for cultural networks, the next frontiers lies on our ability to integrate and study groups of networks that were initially conceived separately due to the different nature of the entities studied, but that need to be integrated for higher-level research questions to be answered. An example of this type of supercomplex network is the cultural network that integrates all the cultural networks of a specific historical period, that is, of books, art, music, commerce, etc. in order to have a general but precise perspective of cultural production in that given historical time and place.

#### IV. WHY ARE CULTURAL NETWORKS IMPORTANT?

Another type of supercomplex cultural network is the one that is being formed by large multinationals who deal with social media -Amazon, Google, Facebook, Baidu, Mixi. In these cases, the client information that they collect does not necessarily serve a historical purpose, but deals with us, common human beings that leave traces of behaviors through our use of different digital and mobile platforms. This, in turn, is related to other people, as well as commercial and cultural objects that they interact with. The aims of these companies relate strongly to their respective commercial strategies, as they use information to understand to identify the trends that inform our behaviour, to predict actions, and, crucially, to relate all this information to their own business agenda.

Through Boyd and Richerson's approximation to culture as information that affects people's behavior, we will understand that both the information that human beings produce with respect to art and high culture as well as that which is derived from their interactions with mainstream or pop culture [14] offer us a new window to peek into history, evolution, and current events. Indeed, our ability to understand these processes through big data analysis and cultural networks constitute an advance to implement what Dan Sperber labeled "the epidemiology of representations" [15]. Developing a large-scale research program based on the concept of cultural networks encompassing multiple political and geographic areas as well as historical periods, types of cultural artifacts and representations, would initiate a paradigm shift that could shed light on human history and the relationships between cultural and geographic areas. Similarly, we would be better equipped to understand the dissemination of ideas and cultural phenomena and, in relevant cases (such as in art fairs, universal exhibitions, or Olympic games), we would be able to apply this knowledge to improve the management of current events.

With this in mind, we need to integrate our knowledge of objects and people, change the scale of our observation of culture, and take advantage of the new possibilities for data storage and analysis that Big Data techniques offer. The frontiers of our understanding of human beings, history, and cultural evolution are being drawn by cultural networks that

we are all connected to. They link us to the past in mysterious ways, but also offer us the opportunity to predict our future.

#### REFERENCES

- [1] Jean-Baptiste Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books," *Science (New York, N.y.)* 331, no. 6014 (January 14, 2011): 176–82, doi: 10.1126/science.1199644.
- [2] John Bohannon, "Google Opens Books to New Cultural Studies," *Science* 330, no. 6011 (December 17, 2010): 1600–1600, doi:10.1126/science.330.6011.1600.
- [3] Leetaru, Kalev. "Culturomics 2.0: Forecasting Large-Scale Human Behavior Using Global News Media Tone in Time and Space." *First Monday* 16, no. 9 (August 17, 2011). See also Nanne van Noord, Ella Hendriks, and Eric Postma, "Toward Discovery of the Artist's Style: Learning to Recognize Artists by Their Artworks," *IEEE Signal Processing Magazine* 32, no. 4 (July 1, 2015): 46–54.
- [4] Javier De La Rosa and Juan Luis Suárez, "A Quantitative Approach to Beauty. Perceived Attractiveness of Human Faces in World Painting," *International Journal for Digital Art History* 1, no. 1 (June 26, 2015): 112–29.
- [5] A. Weiss and R. James, "An Examination of Massive Digital Libraries' Coverage of Spanish Language Materials: Issues of Multi-Lingual Accessibility in a Decentralized, Mass-Digitized World," in *2013 International Conference on Culture and Computing (Culture Computing)*, 2013, 10–14, doi: 10.1109/CultureComputing.2013.10.
- [6] Anne Burdick, ed., *Digital Humanities* (Cambridge, MA: MIT Press, 2012).
- [7] Amy Zhao Yu et al., "Pantheon: A Dataset for the Study of Global Cultural Production," *arXiv:1502.07310 [physics]*, February 25, 2015, <http://arxiv.org/abs/1502.07310>.
- [8] Maximilian Schich et al., "A Network Framework of Cultural History," *Science* 345, no. 6196 (August 1, 2014): 558–62, doi:10.1126/science.1240064.
- [9] Juan Luis Suárez et al. "Towards a digital geography of Hispanic Baroque art". *Literary and Linguistic Computing*, first published online August 13, 2013. Juan Luis Suárez et al., "The art-space of a global community: the network of Baroque paintings in Hispanic-America." *Proceedings of the International Conference on Culture and Computing 2011* at Kyoto University, Japan. 2011. Juan Luis Suárez et al. "Sustaining a Global Community: Art and Religion in the Network of Baroque Hispanic-American A Model for the Study of Cultural Complexity in the Atlantic World." *South Atlantic Review* 72:1 (2007): 31-47. See also Lautaro Núñez, Martin Grosjean, and Isabel Cartajena, "Human Occupations and Climate Change in the Puna de Atacama, Chile," *Science, New Series*, 298, no. 5594 (October 25, 2002): 821–24; Takeshi Inomata et al., "Earl Ceremonial Constructions at Ceibal, Guatemala, and the Origins of Lowland Maya Civilization," *Science* 340, no. 6131 (April 26, 2013): 467–71, doi:10.1126/science.1234493; Michael Bawaya, "A Chocolate Habit in Ancient North America," *Science* 345, no. 6200 (August 29, 2014): 991, doi:10.1126/science.345.6200.991.
- [10] Alex Pentland, *Social Physics: How Good Ideas Spread- the Lessons from a New Science* (New York: The Penguin Press, 2014).
- [11] Erika Ortega, Juan Luis Suárez, and David Brown, "Redes Textuales: Bases de Datos En Grafo Para Estudios Literarios," *Insula: Revista de Letras Y Ciencias Humanas*, no. 822 (2015): 22–25.
- [12] Juan Luis Suárez Javier de la Rosa, "SylvaDB: A Polyglot and Multi-Backend Graph Database Management System." (International Conference on Data Technologies and Applications, Reykjavík, Iceland, 2013), [http://www.researchgate.net/publication/256439478\\_SylvaDB\\_A\\_Polyglot\\_and\\_Multi-backend\\_Graph\\_Database\\_Management\\_System](http://www.researchgate.net/publication/256439478_SylvaDB_A_Polyglot_and_Multi-backend_Graph_Database_Management_System).
- [13] Peter J. Richerson and Robert Boyd, *Not by Genes Alone: How Culture Transformed Human Evolution* (Chicago: University of Chicago Press, 2005).
- [14] Dan Sperber, "Anthropology and Psychology: Towards an Epidemiology of Representations," *Man, New Series*, 20, no. 1 (March 1, 1985): 73–89, doi:10.2307/2802222.