Cultural Phenomena at Multiscale and Multiresolution: Powers and Perils

Tina Eliassi-Rad (Northeastern University)
Leon Gurevitch (Victoria University of Wellington)
Maja Gutman (Institutum Studiorum Humanitatis (Alma Mater Europaea))
Katherine Kinnaird (Brown University)
John Laudun (University of Louisiana)
Pablo Suarez-Serrato (National Autonomous University of Mexico UNAM)
Melvin Wevers (Utrecht University)

Introduction

Culture Analytics seeks to understand cultural phenomena as inherently multi-scale and multi-resolution. Cultural phenomena are multi-scale in that they span various ranges of time and/or exist in scalable spaces. They are multi-resolution in terms of their self-similarity across time and space. To address these complexities, Culture Analytics harnesses methods drawn from Machine Learning, Applied Mathematics, and Data Science, and weds them to problems in the study of culture, drawn largely from the Humanities and the Social Sciences.

Humanities and Social Science theories have long acknowledged, and often accounted for, the dynamic nature of human culture.[1] While the many interrelated fields represented by these broad disciplines have created several centuries of productive and meaningful research, the opportunity now exists to operationalize many of the theories that form their basis. A major focus of Culture Analytics is, consequently, to begin an open dialogue concerning the implementation of the abstract tools and models from mathematics and related disciplines that can address questions concerning culture.

The Arrowhead problems in Culture Analytics (see White Paper: Introduction to Culture Analytics) raise several questions focused specifically on the application of multi-scale and multi-resolution techniques: What elements, or dimensions, of cultural phenomena can be measured, and what are the appropriate metrics? What kinds of models, not only descriptive but also predictive, could be developed based on these metrics?

[1] During the Culture Analytics program Aaron Clauset, Dan Jurafsky, and Maria Binz-Scharf among others approached cultural expressions (university hiring, food, work practices) as dynamic and complex processes.

Main Theoretical Components of this Topic

Large amounts of cultural data are now available online.[2] Already, a great deal of work is being done to develop more sophisticated ways to correlate diverse dimensions of cultures and explore hierarchies describing the culturally productive ways in which people in groups interact (cf Journal of Cultural Analytics). Culture Analytics takes this multi-scale, multi-resolution approach to be foundational and seeks to be a synthetic and contiguous sphere of activity. Many of the current efforts across the Humanities, the Social Sciences, the Mathematical Sciences, and the Data Sciences have largely relied on various kinds of borrowings. We contend that Culture Analytics requires a more integrated approach to addressing these complex problems (cf White Paper: Collaboration and Translational Data Science for the Study of Culture). In Workshop 3, for example, Ingrid Daubechies demonstrated that image analysis could detect differences between original paintings and their forgeries, ultimately revealing that a difference in the brush strokes between originating artists and forgers could be attributed to intentionality. Intriguingly, it is the forger's strokes that are more considered, controlled, and regular due to the conscious nature of replication. Her collaborative work with artists, art historians, and computer scientists is exemplary of the multi-scale, multi-resolution approach we see as the key to a successful Culture Analytics.

Mathematical Possibilities for Quantifying Cultural Phenomena at Multi-scale and Multi-resolution

It has become something of a commonplace that culture, and cultural data, exist at a number of levels. The usual contrast has been close reading versus distant reading (Moretti), but other terms such as macroscope (Tangherlini) or macroanalysis (Jockers) reveal that the binary is less interesting than the ability to scale up and down, achieving the resolution that best works for the data or topic at hand.[3] While this notion of scale has become a regular part of discussions in the digital humanities and social informatics, there has been less attention paid to how cultural phenomena on the micro level might reveal a level of heterogeneity that disappears at the macro level. To understand the multiple identities and multiple worlds within which individuals interact as well as how small communities and identities cohere into larger wholes, Culture Analytics approaches the micro and macro aspects of culture on both a continuous scale and resolution.[4] Such a focus requires a more sophisticated understanding of the relevant

---

[2] Note that despite the "bigness" of data, our observations of cultural phenomena are often partial. (Alas, we are not omniscient.) This partial observability leads to biased and incomplete data; and consequently to skewed and sometimes even incorrect results. Eliassi-Rad's talk in Workshop 3 covered ways to enhance incomplete data.

[3] Moretti, Franco. 2013. *Distant Reading*. Verso. Tangherlini, Timothy. 2013. The Folklore Macroscope: Challenges for a Computational Folkloristics. *Western Folklore* 72:1 (Winter 2013): 7–27. Jockers, Matthew. 2013. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press.

[4] *Don't ask where I'm from, ask where I'm a loca*l by Taiye Selasi; DeLanda, Manuel. *A New Philosophy of Society: Assemblage Theory and Social Complexity*. London: Continuum, 2006; Appadurai, Arjun.

methods and methodologies as well as a more nuanced recognition of the culturally motivated questions and theories.

This ability to shift scales has clear ramifications for the application of cultural understanding to real world problems, such as those related to the environment, social unrest, and public health. To be effective, for example, improving public health requires efforts that function on multiple scales. Attempts to reduce tobacco use should focus on the individual (nicotine patches), the family/social unit (counseling interventions that promote constructive dialogue), and government institutions (tobacco control legislation). Without an understanding of the interaction between scales and common patterns across scales, the mechanisms implemented to affect change are likely to be incomplete and insufficient.

An intriguing idea related to culture is that these multi-resolution phenomena reflect self-similarities and fractal structures. A visual metaphor for such a phenomenon is a Sierpinski carpet (Figure 1a). Cultural processes might not always be best explained using such strict formalizations. For instance, those drawn from chaotic mixing (Figure 1b) might be more apt. Alternatively, consider the development of complex fractals as a result of fluid flows (cf IPAM Long Program on Turbulence). This mixing model approximates processes of virality, diffusion, and self-similarity in chaotic systems and may have considerable applicability to cultural phenomena, aligning well with the theories of culture found in the Humanities and Social Sciences (e.g. social constructivism, critical theory, structuralism, phenomenology, symbolic interaction theory and game theory).

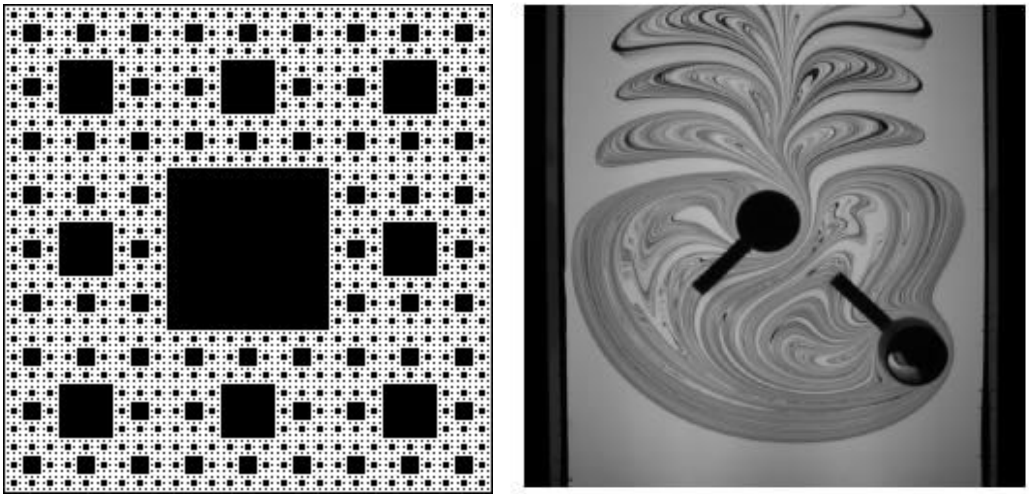

Figure1: (a) Multi-resolution (self-similarity): An Example of a Sierpinski Carpet. (b) An example of Chaotic Mixing, for a Visual Analogy (from http://iramis.cea.fr/Images/astImg/993_1.jpg).

Modernity At Large: Cultural Dimensions of Globalization. Minneapolis: Univ Of Minnesota Press, 1996; Robertson, Roland. "Glocalization: Time-Space and Homogeneity-Heterogeneity." In *Global Modernities*, edited by Mike Featherstone, Scott Lash, and Roland Robertson, 25–44. Theory ; Culture & Society. London: SAGE Publications, 1995.

Leveraging Mathematics for Culture Analytics

Feature extraction, dimension reduction, feature engineering, representation learning, and topological data analysis are a few of the different areas in Applied Mathematics and Machine Learning that can be leveraged in Culture Analytics. Dimension reduction and feature extraction are two areas that facilitate the examination of cultural data at various levels and resolutions, providing insights on the patterns that exist at the micro-, meso-, and macro-levels. Both techniques seek to limit the number of dimensions or features used to understand data, while not losing crucial information about that data. Importantly, these techniques apply across all cultural data. So, for instance, in Workshop I, Lindsay King demonstrated an interface for 2-dimensional exploration of Vogue magazine covers across several time and hue scales.

We stress that computational processes applied to the data-driven analysis of culture at scale must include the domain expertise of scholars from fields traditionally focused on the study of culture. Such translational collaborations will push to innovations in dimension reduction and feature extraction techniques (cf White Paper: Collaboration and Translational Data Science for the Study of Culture). In an intriguing presentation during Workshop 3, Aaron Clauset discussed work using network science on university employment patterns, exposing that North American universities tend to hire each other's students resulting in a "homeland" versus "colonies" dynamic. This work is an excellent example of using network science at several scales to illuminate patterns at both macro and micro scales, while at the same time identifying phenomena in our own close homogeneous groups (university departments) of which we are unaware, and that might require thoughtful consideration.

Examples from social networks revealed the pervasiveness of bot accounts (cf Menczer BotOrNot),[5] while studies of censorship and self-censorship in weblogs were discussed by Margaret Roberts.[6] The work presented by these two groups has led to a new collaboration in studying the effect of entire networks of bot accounts that flood Twitter, thereby censoring the communication channel that "protest hashtags" provide. These bot attacks have profound social impact, and have reportedly influenced elections in Latin America recently.[7] The multi-scale nature of these investigations is revealed when comparisons are made at increasing levels of resolution, starting with small locations and moving up in scale through cities, countries, regions and ultimately the whole world. This approach allows for cultural comparisons to be made at a quantitative level.

Feature engineering and representation learning are essential components of any approach in Culture Analytics. Feature engineering is the process of (often manually) creating

---

[5] BotOrNot: A System to Evaluate Social Bots (WWW Developers Day 2016) Clayton A. Davis, Onur Varol, Emilio Ferrara, Alessandro Flammini, Filippo Menczer, arXiv:1602.00975 [cs.SI]

[6] King, Gary, Jennifer Pan, and Margaret E. Roberts. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* (2013). copy at http://j.mp/LdVXqN

[7] "How to Hack an Election, Andrés Sepúlveda rigged elections throughout Latin America for almost a decade. He tells his story for the first time." Jordan Robertson, Michael Riley, and Andrew Willis. March 31, 2016 (http://www.bloomberg.com/features/2016-how-to-hack-an-election/ accessed 6/9/16).

features for machine learning algorithms and relies heavily on domain expertise.[8] The teams that recognize the need for domain expertise move rapidly to the head of the line, as domain experts provide clear direction on feature engineering. Representation learning is the process of automatically transforming data into representations that can be effectively used by machine learning algorithms; domain experts here can provide input into this process through data selection and priors on the representations. Part of the excitement around deep learning comes from the effectiveness of representation learning for several important tasks such as object recognition and speech recognition. Aram Galstyan's talk in Workshop 3 included aspects of representation learning for social media data.

Topological data analysis is a method to explore the various scales at which a cultural phenomenon may take place in a manner that does not depend on a specific distance or metric. This technique allows for different topological features of a given space (datasets, point clouds, vector representations, etc) to be distinguished and recorded in a precise fashion (Zomorodian-Carlsson, Ghrist).[9] In this view, a network gives rise to a space called a simplex, and its persistent homology is computed. Multiple timescales can be observed by using perturbation theory to analyze nonlinear dynamical systems (Strogatz). Looking at networks again, theorems from symbolic dynamics can quantify properties of diffusion of culture (memes, virality, and misinformation, for example) in terms of mixing, entropy and ergodicity[10].

An important aspect of Culture Analytics is measuring similarity across scales and resolutions. Eliassi-Rad's presentation during the Tutorial week covered a guide for selecting algorithms that measure similarity in complex (social) networks at micro-, meso-, and macro-levels.

Conclusion

We see one of the first steps for Culture Analytics to be the development of a coherent, central core of methods that can act as both a resource and a guide for practitioners. One could imagine a series of workshops or other pedagogical interventions that would help educate researchers from the humanities and social sciences about the relevant mathematics and machine learning methodologies and, just as importantly, educate mathematicians and data scientists about the relevant theoretical aspects of the study of culture. By bringing these groups of domain experts together, the opportunity is ripe for creating entirely new approaches and theories of cultural dynamics. While there have been a number of theories developed about the

---

[8] Matthew Jockers - Sentiment Analysis as Proxy for Plot; David Blei - Understanding History using Topic Modeling; Taylor Arnold - Recommender Systems for Digital Archives.

[9] Edelsbrunner, H., Letscher D., and Zomorodian, A. Topological persistence and simplification. *Discrete Comput. Geom*. 28 (2002), 511–533. Afra Zomorodian and Gunnar Carlsson. 2005. Computing Persistent Homology. *Discrete Comput. Geom.* 33, 2 (February 2005), 249-274. DOI=http://dx.doi.org/10.1007/s00454-004-1146-y. R. Ghrist, "Barcodes: The persistent topology of data," *Bull. Amer. Math. Soc.*, 45(1) 61-75.

[10] Hasselblatt, Katok, *Introduction to the Modern Theory of Dynamical Systems*, *Cambridge University Press, 1996.*

way culture works, Culture Analytics offers the chance to operationalize these theories, thereby making them available for testing. Consequently, Culture Analytics will not only significantly advance how we understand culture but also the way we understand the development of ideas, itself a cultural process.