Collaboration and Translational Data Science for the Study of Culture

Maria D'Orsogna (CSU Northridge)
Sandra Alvaro (Autonomous University of Barcelona)
Blake Hunter (Claremont McKenna College)
Mila Oiva (of Turku)
Isabel Meirelles (OCAD University)
Clarisse Bardiot (University of Valenciennes)
Ana Jofre (OCAD University)
Lev Manovich (The Graduate Center, CUNY)

Introduction:

The unprecedented growth of cultural data created, collected and shared digitally over the past decade presents enormous opportunities for a better understanding of cultural trends, popular interests and human needs. Extracting information from heterogeneous and, at times, poorly defined data sources and then applying this information in a cogent way presents significant challenges, but holds the promise of advancing the state of the art across many of the core disciplines of Culture Analytics: the Humanities, the Social Sciences, the Library Sciences, Applied Mathematics, Physics and Computer Science. Meaningful advances in the study of culture can derive from a deliberately translational approach to the data-driven study of culture at scale, engaging in a vigorous transdisciplinary approach to problems that range from archiving historical cultural artifacts (see White Paper: Cultural Archives), to helping individuals navigate the cultural complexities of the dating scene (see White Paper: Cultural Phenomena at Multiscale and Multiresolution: Powers and Perils), from tailoring culturally-sensitive medical interventions for at risk patients to creating personalized advertisement, while recognizing the demands of users as they navigate these complex collections (see White Paper: Interfaces and User Experience Design for Culture Analytics).[1]

Challenges in culture analytics abound and the enterprise can only succeed if deep teams that traverse these core disciplines are created, taking advantage of the domain competencies of all the team members, thus sidestepping one of the greatest potential traps of data science: imagining that everything is a nail once one has constructed an algorithmic hammer. Indeed, once an issue is identified, one must ask the right questions, acquire and bring

[1] All examples are taken from the lectures presented at the Culture Analytics long program of 2016 at IPAM: Ingrid Daubechies (Duke University) "The Master's Hand: Can Image Analysis Detect the Hand of the Master"; Workshop 4: Understanding History through Topic Modeling. Joint work with David Blei (Statistics & Computer Science), Allison Chaney (Machine Learning), Hanna Wallach (Computer Science), and Matt Connelly (History); Workshop 3: Network Structure and Dynamics of the Scientific Workforce Aaron Clauset (University of Colorado); Workshop 1: Maximilian Schich (University of Texas at Dallas) Figuring out Art History - Towards a Systematic Science of Art and Culture.

structure to the available but ever-evolving data, so that a meaningful analysis grounded in a substantial understanding of culture can be carried out. Results must be presented in an intelligible and user-friendly way across disciplines that reveals the value added by approaching the problem from the translational data science approach of culture analytics.

During the Spring of 2016 NSF's Institute for Pure and Applied Mathematics (IPAM) at UCLA embarked on one of the first large-scale experiments of this kind, bringing together forty professionals from a variety of backgrounds for a residential period of three months to share experiences, foster collaboration, and spark cross-disciplinary dialogue. An additional 115 speakers and participants from academia and industry came to participate in workshops for varying lengths of time to share their expertise and to explore future research collaborations.

The 2016 Culture Analytics long program has allowed us to establish interdisciplinary collaborations and to lay the foundations for the creation of a common outlook and toolkit across disciplines. Among the participants were literary scholars, mathematicians, computer scientists, educators, librarians, medievalists, geographers, political scientists, artists, designers, sociologists, physicists, art historians, cultural historians, statisticians. As a group, we realized that we often tackled similar questions from different perspectives, and found it constructive to learn from each other's methods and backgrounds, through critique and thinking out of the box. Our interactions show that translational research between the Humanities, Social Sciences, Engineering and the Mathematical Sciences can be greatly beneficial in identifying questions, honing approaches, and finding new ways to overcome obstacles. Hard scientists were able to offer methodological strength to the complex culture-grounded research questions offered by experienced Humanities scholars and social scientists. Importantly, the collaborative engagement across these disciplines allowed for new, significantly more sophisticated questions to be formulated.[2]

As an example, we created a diverse working group of 10-15 people to analyze trends emerging from the art and science *Leonardo Journal* published by the MIT press (Cambridge, MA) for which raw data was available. The group met at the same place and time during the non-workshop weeks of the long program. As a group, we educated each other on our respective methodologies, and were able to find common language and understanding. We proposed specific research questions, dissected and discussed these problems in a cooperative and constructive manner, and began formulating more substantive research questions. Those of us who had specific skills, such as topic modeling analysis, data cleaning or script coding, gave tutorials and documented that work in detail, sharing files on google docs. As a result of this collaboration, the group has put together a multi-year, international grant application to the Transatlantic Program of the Digging into Data challenge (NEH, CLIR and NSF as the core US partners).

[2] Workshop 3 - A relational practice approach to the study of social networks Maria Binz-Scharf City University of New York (CUNY)

Challenges in Translational Data Science for the Study of Culture:

Culture analytics reaches across many disciplines. As noted, a core challenge is studying cultural artifacts and culturally-informed behaviors at the very large scales now possible, being cognizant of the complexity, heterogeneity, and velocity of these data resources. While there are clear benefits to addressing the data-driven analysis of culture through a multiscale / multiresolution model of culture (see White Paper: Cultural Phenomena at Multiscale and Multiresolution: Powers and Perils), such an approach necessitates a translational disciplinary perspective that is fundamentally reliant on collaboration across traditionally unrelated disciplines. The level of collaboration necessary for successful Culture Analytics is a key challenge in this burgeoning area.

Humanistic + Scientific methods:

For well over a century, the discourse on cultural studies has been divided across the frameworks informed by interpretive methodologies (hermeneutics) as opposed to empirically driven methodologies. The former contends that a relativist approach is necessary and that phenomena and observations must be interpreted in context. Meanings, social uses, symbolic values, historical eras, may all contribute to the same material being interpreted in different ways. The latter builds on the assumption that scientific, universal, quantitative models must be at the basis of cultural understanding. From this viewpoint, the processes that lead to the “truth” are not dissimilar from those used in the development of mathematical theorems or of the universal laws of physics. Can we use, bridge and integrate these approaches? If so, what methods can we use?

Data applications:

Cultural and technological artifacts mediate how we access and interact with the world. Can we design virtual and real-life user-interfaces that ease and increase participation, access, transparency, affordability, activism, exchanges of ideas (see White Paper: Interfaces and User Experience Design for Culture Analytics)? Can we use these tools to better our social and natural environment and help alleviate local and global socio-economic problems?

Scientific methods:

Science uses a number of methods: the laws of physics, statistical inference, and simulations. When and how can we apply them to traditional humanistic approaches? Which methods do we select? How do we ensure that end results make sense, that they are understandable to both communities, that the integrity of each discipline is maintained? Judiciously translating methods across fields is the crux of all interdisciplinary efforts, one that can lead to great rewards if tackled properly: new humanistic insight may emerge thanks to the use of quantitative methods; at the same time the development of new quantitative tools may be spurred by the unique challenges of problems in culture analytics.

Humanistic methods:

Culture is a dynamic process that not only involves the creation of expressive forms, but also the reception and interpretation of those expressions by a wide range of audiences. The same artifact may be interpreted differently by different groups of people and different points in time. How can we incorporate cultural context in the context of data-driven analysis? Can we develop methods that quantify context, or allow for a more relativist analysis? Possible examples include tracking the history of cultural artifacts (such as magazine covers reflecting standards of beauty) to highlight how these aesthetics change geographically, in time, and in response to current events. The proper analysis of evolving cultural structures calls for tools from both the sciences and the Humanities.

Group dynamics and workflows:

Due to its inherent cross disciplinary nature, analyzing culture requires the integration of diverse tools and knowledge that may be specific to the particular phase of the workflow. Among the most common stages of any work plan are Data Mining (calling for expertise in Mathematics and Computer Science); Visualization (calling for expertise in Data Visualization, Design, Art, Library Information Science); Analysis and Interpretation (calling for expertise in the Social Sciences and Humanities). A typical workflow, with other auxiliary stages, is depicted in Table 1 and Figure 1. Traditionally, some of the disciplines within these stages have been deeply siloed: for them, cross collaboration is a new challenge. However, even for fields that have been more open to cross-disciplinary work, such as Applied Mathematics that has benefitted from branching towards biology, physics or finance, interacting with the Humanities and Social Sciences opens new possibilities.

Sample Culture Analytics Workflow -- Art History of Late 19th Century France

1. Define an area of interest along with meaningful interrelated research questions from different disciplines
2. Devise domain data
  a. Data Collection (domains: Art History methods: Computer Science, Library/Info Science)
  b. Data Cleaning (Computer Science, Linguistics)
3. Modeling and Representation (Computer Science, Applied Mathematics, Art History)
  a. Data Mining (Mathematics, Computer Science, Art History)
  b. Visualization (Data Visualization, Art, Library Information Science)
4. Analysis / interpretation (Sociology, Art History, Anthropology)
5. Refine and Repeat

Table 1: Typical stages of a cultural analytics workflow. Note the wide variety of disciplines involved, and the need for continuous dialogue and feedback across stages.

Due to the high degree of specialization that seems to be requisite in each of these stages, it is tempting, and almost natural, for the respective experts to create insular working groups that are not necessarily integrated with one another. As a result, the process risks becoming fragmented into sub-processes having each their own language, tools and approach, rendering the collaboration into a series of “black boxes”, the inner workings of which are inaccessible to outside members. While some degree of specialization is necessary, the translational goal of Culture Analytics is to devise mechanisms for better integrating tools and expertises across disciplines without requiring individuals to master all the necessary methodologies.

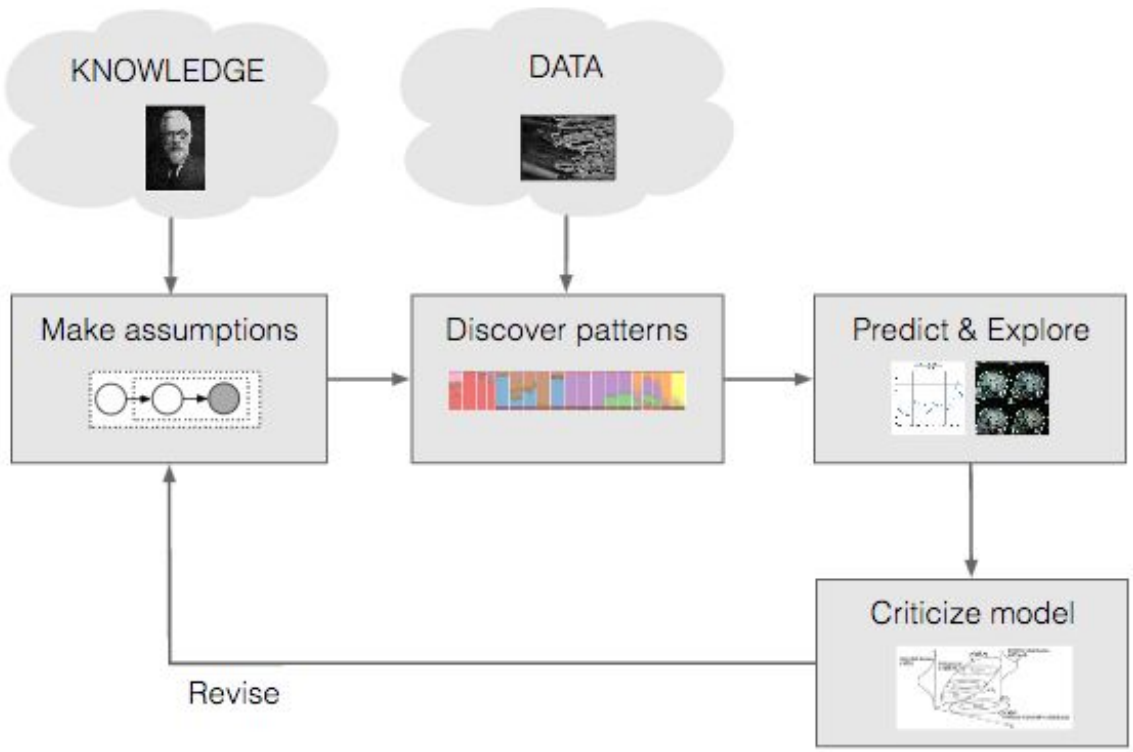

Figure 1: Schematic of a typical cultural analytics workflow as illustrated by Dr. David Blei
http://helper.ipam.ucla.edu/publications/caws4/caws4_12767.pdf

For Culture Analytics to move beyond the impressive initial success of the field, it is imperative to create clear opportunities for researchers to interact and learn from one another, supporting the essential iterative processes outlined above. Critically, researchers must have opportunities not only to identify appropriate and emerging methods to solve specific problems, but also to find collaborators *with whom* to formulate and solve these problems. Throughout the program, various lecturers pointed out these challenges and showed effective ways of

overcoming them.[3] Additional examples were presented related to successful collaborations between artists and researchers in different fields.[4]

Common Language:

Understanding the value of disciplines other than one's own is a central step in creating a truly interdisciplinary working group. It is important to not speak in jargon, to be humble and listen to one another. For example, similar terms may index different concepts in different disciplines, and vice-versa. These attitudes may be helpful in developing a common vocabulary, and hopefully a common language, to avoid misunderstandings and so that exchanges can be more direct, and for a better understanding.

Recommendations:

Below we propose some possible solutions to the challenges outlined above. Of these, some come from within specific disciplines, and can be implemented by researchers, while others may require social or institutional changes.

1. *Support mechanisms to establish common language and methods across disciplines*

One of the best ways to foster cross-cultural dialogue is education (See White Paper: Teaching Culture Analytics). Basic tools and approaches can be taught at different levels, from undergraduate and graduate seminars and classes, to extended workshops and programs for academic researchers, from online tutorials and videos to dedicated educational websites. Some of the goals of these educational efforts include:

- Training Humanities and social science scholars in fundamental scientific skills and approaches including statistics, data analysis, and fundamentals of computer science, including coding and data structures
- Training scientists in aspects of the Humanities and qualitative Social Sciences, particularly anthropology, sociology and geography, so they can acquire a fundamental understanding of the methods used in these fields that focus on the study of culture
- Training all culture analytics researchers in visualization and user interface design as means for exploring data and being better able communicate findings both to other experts and general audiences (See White Paper: Interfaces and User Experience Design for Culture Analytics).

---

[3] Workshop 1: Jacob Foster (Sociology, University of California, Los Angeles (UCLA)) : *Metaknowledge: Mapping, Measuring, and Modeling Science*; Yong-Yeol Ahn (Indiana University) Emotion, complexity, and cultural transmission; Workshop 2: Mauricio Giraldo (New York Public Library) Human-Computer Collaboration at NYPL Labs; Workshop 1: Lev Manovich, Culture Analytics Beyond text.

[4] Tutorials: Victoria Vesna (University of California, Los Angeles (UCLA)) Part 1: Bird Song Diamond: mapping the acoustic network of birds (with John Brumley); Workshop 1: George Legrady University of California, Santa Barbara (UCSB) Media Arts & Technology. *Making Visible the Invisible DataVis at the Seattle Public Library*; Workshop 2 : Philip Beesley (Philip Beesley Architect Inc.) Living Architecture Systems.

We are not advocating that scholars become experts in the other disciplines with whom they will work. What we are advocating is that researchers be exposed to, be curious of, and become well-versed in the terminology, approaches, and questions of the other disciplines that traverse the many disciplines of Culture Analytics. This cross-pollination will allow for better dialogue, cooperation and understanding, and will lead to more robust, intellectually sound, and meaningful results. No single group can turn their expertise to this complex data and expect success. Rather, through a true translational data science for the study of culture, Computer Scientists, Mathematicians, Physicist, Engineers, Social Scientists, Humanities Scholars all working together will be able to ask and answer the vexing questions of culture.

2*. Develop interdisciplinary collaborations*

As participants in the "Culture Analytics" Long Program, we found it very useful to have a physical location and ample time to interact. Several collaborations were formed and we hope to see them come to fruition in the months ahead. We thus recommend:

- Support for more opportunities for researchers to meet, interact, and foster the creation of interdisciplinary knowledge, such as summer schools and long programs similar in spirit to "Culture Analytics";
- Within these programs, create interdisciplinary social situations for researchers to discuss ideas on a regular basis (e.g. tea-time, small dedicated working groups);
- Encourage student and faculty exchanges (summer internships or visiting faculty within a research group);
- Develop in-residence programs for artists and designers to join research labs (both in the Humanities and in the sciences) to specifically develop culture analytics projects.

3. *Support the dissemination of interdisciplinary work*

Since Culture Analytics is a new, emerging and truly transdisciplinary field, there is a need for better outreach, awareness, and dissemination of interdisciplinary knowledge. We believe that this may be achieved by:

- Incentivizing the diffusion of interdisciplinary journals, conferences, workshops, with a specific emphasis on overcoming the barriers of traditionally siloed disciplines;
- Offering more grant opportunities to foster interdisciplinary research.
- Creating more opportunities for the general public to understand and appreciate the societal benefits of culture analytics, with specific examples of how our research efforts impact and hopefully better people's lives. One possibility would be to create interactive portals to encourage participation in culturally aware social, environmental and educational campaigns.

4. *Accessibility to data, tools, methods and analysis by educators and researchers*

Due to the great diversity of user backgrounds and purposes, tools must be designed (or redesigned) with the widest possible audience in mind and must be portable to accommodate different levels of expertise. We propose:

- The development open source, “user friendly” software. The goal is to offer basic commands and concepts given that one of the largest barriers many non computer scientists face in beginning to work with the data-driven analysis of culture and cultural phenomena is learning how to write code.
- Making data more accessible to researchers, given the needs and speed of twenty-first century research. Progress is at times hindered by copyright issues that vary from country to country. Encourage (and incentivize) open data for research, from large private companies, such as Google, Twitter and Facebook;
- Creating venues for researchers to share best practices both within academia and with the general public;
- Creating code and repositories, as well as demonstrator projects with reusable code and data, that are specific to culture analytics methods.

Conclusions and future outlook:

Many scholarly disciplines are currently confronted with opportunities and challenges brought on by unprecedented access to “big data”. Big data, however, can lead to just as big errors if a project gets off on the wrong foot. Culture Analytics is focused on bringing together teams from across all of the culture disciplines and all of the data science disciplines to ensure that we ask the proper questions, offer the proper analysis, using the proper tools and the proper methodologies. Understanding and making sense of cultural data, however large or small, requires thorough thinking and a wide range of methods and viewpoints. Collaborations are fundamental for a holistic, complete view of cultural data, cultural structures and social causalities.

In most data-driven cultural studies to date, researchers have been content with simply incorporating methods or perspectives from disciplines other than their own without true engagement with domain experts. Much greater leaps can doubtlessly be made by fashioning truly collaborative teams where ideas, methods, and questions are developed across disciplines without simplifying challenges and assumptions.

Learning and appreciating skills beyond one’s own discipline is time consuming and requires a sincere commitment. However the rewards are valuable for everyone involved, as evidenced by the enormously positive outcomes of the Culture Analytics long program. By joining forces, research questions will not run the risk of lacking accuracy or of missing the greater picture; the tools used will be state of the art and the approaches employed will guarantee optimal outcomes.

The “Culture Analytics” long program of 2016 at IPAM-UCLA was a first step in establishing a common language for the data-driven analysis of culture at scale, fostering interdisciplinary collaborations, and accessibility across disciplines. We hope that further progress will be made as members of the core group return to their home institutions and continue collaborating on joint projects initiated at IPAM.