

**Data as Critique:
New Computational Approaches to the Study of Culture
May 9, 2014
University of Chicago**

Abstracts:

Andrew Piper

Associate Professor of German Literature, McGill University

The Sociability of Detection

In this project we explore how social network analysis can provide new insights into the meaning and social functions of literary genres. Focusing on the genre of detective fiction, we are interested in asking how it is that a community arrives at a shared sense of truth. Rather than emphasize the work of a single, charismatic individual, we want to know what the social conditions are that allow for the arrival of shared truths. To do so, we have implemented new crowd-sourcing techniques that allow us to extract social networks from prose texts with a high degree of both speed and reliability. In addition, we have developed measures that try to understand the literary features specific to narratives of detection. In what ways can detective fiction be understood as a genre designed to help us conceptualize “communities of detection” and the social conditions through which new truths are generated?

Tom McEnaney

Assistant Professor of Comparative Literature, Cornell University

Provincial Cosmopolitans and the Literary System of the Americas

In an early preface to the vanguard Argentine journal *Sur* the editors described the public they sought to invoke: “South American, North American and European writers... will find in the pages of *SUR*... all of the problems concerning our spiritual life, our moral values and our intellectual facts. These problems will be our problems; this America ‘our America’ [nuestra América].” While *Sur*’s editors—Jorge Luis Borges among them—claimed the journal would study “our America,” bibliographic data of the period tell a different story in which poets who appeared in *Sur* rarely published elsewhere in Latin America. Employing network and stylometric analysis from the mining of data across thousands of poems, this talk explores what *Sur* can tell us about the larger literary system in Latin America. What makes *Sur*’s America both cosmopolitan—connected to writers across Europe and the United States—and strangely provincial or marginal within the literary network of Latin America? Comparing the journal’s collective style with the period’s most widely published poet—Juana de Ibarbourou, aka “Juana de América”—should provide a new means to conceptualize the gateways and barriers to poetic circulation during the vanguard period.

Dennis Tenen

Assistant Professor of English, Columbia University

System Administrator Syndrome

Why does participation in peer production communities sometimes decline with time? In this paper, I analyze data from one of the largest pirate libraries in the world, extracting a social network from the patterns of file sharing over time. My hypothesis, still speculative at the moment, is that barriers to participation grow in proportion to the perceived value of the common good. When the community is young and the archive is small there are few rules and regulations governing the acquisition of new material. As the library matures, the risk of a new contributor "messing things up" grows as well. At this point, we see the emergence of a new class of dedicated users, the "sys admins," who raise barriers to participation to protect their investment of time and resources. This dynamic, I suggest, is not unique to online peer production.

Jacob Foster

Assistant Professor of Sociology, UCLA

Finding Cultural Holes: How Structure and Culture Diverge in Networks of Scholarly Communication

Divergent interests, expertise, and language form cultural barriers to communication. No formalism has been available to characterize these "cultural holes." Here, we use information theory to measure cultural holes, and demonstrate our formalism in the context of scientific communication using papers from JSTOR. We extract scientific fields from the structure of citation flows, and infer field-specific cultures by cataloguing phrase frequencies in full text and measuring the relative efficiency of between-field communication. We then combine citation and cultural information in a novel topographic map of science, mapping citations to geographic distance and cultural holes to topography. By analyzing the full citation network, we find that communicative efficiency decays with citation distance in a field-specific way. These decay rates reveal hidden patterns of cohesion and fragmentation. For example, the ecological sciences are balkanized by jargon, while the social sciences are relatively integrated. Our results highlight the importance of enriching structural analyses with cultural data.

Jennifer C. Lena

Associate Professor of Arts Administration and Sociology, Columbia University

Why Listening is Overrated: A Relational Approach to Classifying Style and Excellence in Music

I am interested in how social relations produce classification systems that people treat as external to those relations. A relational or network approach to category construction can reveal resemblances “hidden” by classification systems. In this presentation, I provide an empirical justification for a relational approach to the study of culture and classification using two of my recent projects on music. In the first, a secondary analysis of musical history texts, I discover four social structures characterize 20th century musical communities, and these provide a more robust explanation for styles of music-making than do musicological genre categories. In the second, a quantitative study of rap’s musical content, I demonstrate that imitation fuels peer esteem and chart success, in spite of autonomous art discourses that hold imitators in contempt. These results explode the validity of “natural” theories of genre and of artistic excellence. In the wake of their destruction, I offer alternative specifications for the measure of genre and of peer esteem in creative fields.

Hoyt Long, Assistant Professor of Japanese Literature, Chicago

Richard Jean So, Assistant Professor of English, Chicago

Form in the Machine: Tracing Imitation and Influence in the American Haiku

What do computational techniques like text-classification and machine learning bring to a study of poetic form? Can they help us to discover and track form within massive bodies of texts? This paper addresses such questions through the case of the haiku. At the turn of the 20th century, the Japanese haiku embarked on a global journey that took it to France, England, Latin America, the United States, and various other locales before it circled back to Japan over two decades later. Its influence on Anglo-American modernism is well documented, but the form was also immensely popular amongst lesser known poets of the era, saturating the US poetry field by the 1920s. Here, we demonstrate how large-scale text analysis can help to understand the broader contours of this diffusion process. First by allowing for the detection of formal patterns across large corpora. And second by revealing where these patterns bleed into each other, generating regions of interpretive ambiguity. These regions tell us something of what it is to subject form to the principled logic of the machine, and what it is to read the machine into the social processes of formal imitation.