Who is Thomas Curtis Clark? Modernist Networks of Exclusion
Richard Jean So and Hoyt Long, University of Chicago
Session #233, MLA2014

Imagine if you will American literary modernism – a topic so studied and familiar by now – as not merely the names of an elite cohort of writers, a half dozen significant "little magazines" or a few influential schools of writing, but rather, an entire "network" of thousands of poems, hundreds of authors, and dozens of journals. What would it look like? Perhaps something like this: here are visual representations of the U.S. poetry field between 1915-1919 and 1920-1924 (images 1-2). It would look like thousands of nodes representing poets and journals, linked by thousands of edges representing published poems. In this last slide alone, it would consist of some twelve thousand poems, 2200 hundred poets, and more than 240 periodicals and newspapers. It would look like a vast web of interaction with hundreds of smaller entities sharing ties with dozens of larger organisms. Leveraging available bibliographic data about poetic production between 1915 and 1930 (i.e. which poets published what poems in which journals), we created these network images of U.S. modernism as it existed in empirical, macro-scale terms.

Scanning these network representations of U.S. modernism in the late teens and early 1920s, the images confirm some things we already know about the field. Journals such as *Poetry* magazine dominated the literary scene. Around them existed tightly knit clusters of well-known modernists, such as Pound and T.S. Eliot (image 3). Slightly further away we see the early rise of important minority literary groups, such as the Harlem Renaissance (image 4). But the maps also contain some less obvious truths, exposing fault-lines in our common understanding. For example, one striking aspect of this 1920-24 image, besides all that seems to be happening outside modernism proper, is

that the most prolific poet during this time, the Christian poet Thomas Curtis Clark, is also fairly marginalized in terms of where he published and who he was thus connected to in terms of publication (image 5). He is a sizable specimen within the larger ecology of modernism. But he is utterly alone – unconnected to poets in modernism's center.

Who was Thomas Clark? What does his paradoxical position within American modernism tell us about the status of authorial outliers and the logic of American literary modernism as a system? What does his exclusion – a seemingly structural excision from modernism's core – tell us about the nature of modernism's rules of inclusion, as well as literary modernism itself? In this paper, we combine traditional humanist modes of analysis and quantitative modeling to understand modernism both at the level of local, individual writers and texts and as a vast ecosystem of interacting, linked entities.

In seeking an answer to this mystery, we start at the simplest scale: the level of the text. Here is a poem, "Dead Saviors," that Clark wrote in 1921 and is representative of his general style and content (image 6). Surely there is something in this poem that marks Clark worthy of exclusion. But for the most part, this poem would not be out of place in *Poetry* or *Little Review*. Its topic – World War I – fits squarely with the period's poetic focus among both modernists and middle-brow poets. Its plain language and unadorned style puts him close to distinguished modernist poets, such as Vachel Lindsay and Carl Sandburg. And while his rhyme scheme places him at a distance from the growing popular trend of free verse, elite literary journals in the late teens, such as *Poetry*, still published a fair amount of rhymed verse. Since the early 1940s, literary critics such as Amos Wilder, and more recently, Joan Rubin, have argued that in both content and form, there existed a significant amount of traffic between modernists, such as Ezra

Pound and religious poets, such as Clark. Modernist poets often merely transmuted more "traditional" late 19th century tropes, such as "the savior," into a more modern and thus modernist idiom, and vice versa. We see this at play in Clark's poem: the savior figure obviously taps into an older popular religious discourse, yet is re-fashioned to address the distinctly modernist crisis of the war's affective catastrophe. To the close reading eye, this poem would not be out of place in *Poetry* magazine. But neither this nor any of his other poems satisfied editors like *Poetry* magazine's Harriet Monroe.

Perhaps historical context gives an answer. Is Clark's solitude in our network maps an indication that he lived in relative geographical solitude as well? Perhaps he lived in the middle of nowhere and had no direct contact with other poets. But no: Clark in fact lived in Chicago and he worked at the *Christian Century*, a journal located about ten blocks from *Poetry* magazine. And he was far from a solitary hermit. He went to the same events and social parties as Amy Lowell and Monroe, and he tried hard to curry favor with these important tastemakers. He even struck up a correspondence with Harriet Monroe (image 7) and submitted several poems to *Poetry*. All of his work was rejected by the journal though. Clark was no lone wolf, it turns out, but a conscious player who tried hard to break into the inner circle of *Poetry*. He incessantly wrote fan letters to famous poets, such as T.S. Eliot, in order to collect their signatures (image 8).

He proved a successful networker, in fact, and over the course of the 1910s, he became good friends with two of the most famous Chicago modernist poets: Vachel Lindsay and Carl Sandburg. Letters from both were warm: Sandburg sent hand written poems to Clark and Lindsay refers to him "fraternally" as his "dear friend" (images 9-10). The disjoint between Clark's social network and his literary isolation is a paradox. Take

another look at the U.S. modernist field in the early 20s. *Poetry* magazine's movers and shakers, such as Sandburg and Lindsay, dominate the entire right hemisphere, but Clark is entirely disconnected from his friends and perceived allies. This goes against what we read in their correspondence, where the three poets all seem to feel aesthetically close. It may be that Sandburg and Lindsay were just being kind, or that Clark was simply fooling himself. But how can we make sense of Clark's aesthetic exclusion from his own social network? What larger patterns of American modernism rule over Clark?

In the rest of this paper, we scale up our analysis using computational methods. Our premise is that analyzing localized moments of textual meaning and intention is inadequate for understanding how an entire system of authors and texts interrelate. The words and forms that authors employ have significance at the individual level, but they are also part of larger-scale, but still traceable, patterns of language use that tell us something about the position of these authors in a field of discourse. Exploring large corpora of texts *as data* rather than as individual, localized examples of expression, we believe we can discern patterns of textual relation that elude non-algorithmic approaches. These patterns may seem simple and basic from a literary critical viewpoint (i.e., what words a poet uses), but they encode meaningful socio-linguistic relations. What if we could take all of Clark's poems, for instance – all of his words, all of his rhetorical tics, his total *linguistic style* – and explore how his usage of them differs from the ways that other poets of American literary modernism used their words? What patterns would we see, and would they overlap with patterns of exclusion in other dimensions?

We decided to build an experiment using text classification and machine learning techniques using Python's Natural Language Processing toolkit. The basic idea is this:

we took all of Clark's poems from this period, about 110 in total, and created an abstract representation of them, what's called a "feature." We could make anything a feature really, like the appearance of the word "God" for instance, or a specific phrase. But we needed something more general. One useful representation is the "bag of words" approach. It takes every single text and represents it as a set of all the words in that text. Order is not taken into account, nor frequency, though there are ways one could do that. We then extract from this set more common words, like "is" or "a", as these are not likely to be a good marker of difference between poets. This bag of words is the feature for each poem. The aggregate of these features across Clark's corpus identifies a linguistic "genetic code" for his work, and assumes that there's something in this code that distinguishes him from other poets—an authorial style.

Next, we created this same textual representation for every poem published in *Poetry* between 1912 and 1940, about 9000 poems. The assumption here is that given a set of *Poetry Magazine* texts for a particular time period, say 1912 to 1913, there will be a group of words used often enough that they rise to the top and identify a genetic code for that time period. The *Poetry Magazine* style, let's call it. That's going to vary from year to year, naturally, but the idea is that it'll be distinct enough from Clark to make a meaningful difference. We chose *Poetry* magazine because we believe it represents the major trends and dispositions of U.S. modernism, and thus stands in for modernism at its most normative. We know that structurally, in terms of our publication networks of the field, these two sites of U.S. modernism, core and periphery, are highly discrete. Would an analysis of their linguistic codes confirm or unsettle that observation?

To figure out how distinct *Clark and Poetry* magazine were, we chose to use a machine learning, or text classification, approach. Specifically, for two year windows between 1912 and 1923, we took one-half of Clark's corpus and a random sample of the *Poetry* corpus (equal in size to the Clark) and "trained" the machine to learn which are the bags-of-words most likely to be Clark's poems and which are more likely to be the poems from *Poetry*. That is, we told the machine which was which by labeling each of the texts, and had it learn to associate these labels with their respective bags-of-words. We then took the remaining Clark corpus and sample *Poetry* corpus and removed these labels so the machine did not know which was what. Like a blind taste-test of sorts. We then had the machine attempt to guess correctly which poems were Clark's and which were from *Poetry* based on what it had learned from the "training" set. Our goal was to find if there truly is something that makes Clark "Clark" and *Poetry* magazine "*Poetry* magazine" based on machine pattern recognition. Is there something within his linguistic genetic code that authorizes his exclusion?

Here are some initial results (image 11): first, they indicate that the machine is able to reliably distinguish Clark from *Poetry* magazine, especially as *Poetry* matures. This tells us what we already know – that Clark and *Poetry* magazine are different – but it's useful in showing that the bag of words representation is powerful enough to locate significant differences. And here are the lexical items (both unigrams and bigrams) that tend to distinguish them: (image 12). Confident that it was providing useful information, we decided to extend this model to derive counter-intuitive insights, not just verification of what we know.

Most surprising is this fact that in the first half of the decade, 1912-1915 or so, the machine has difficulty distinguishing Clark from *Poetry* magazine, while as the decade wears on, it finds it much easier to classify Clark. This could just be a function of having more poems to choose from in the *Poetry* corpus, and so the machine does a better job of classifying Clark vs. *Poetry* when it has more texts. But there could be an aesthetic or a historical explanation behind this as well, and zooming back into the texts and history would be appropriate to refine these results. Our results are further useful in that they present a rather accurate index of *when precisely* a kind of distinction in style emerges between Clark and *Poetry* – the year 1915, specifically. The results tell us something concrete about the hardening of the difference between the two. Moreover, and most critically, the machine occasionally misclassifies a poem as Clark or *Poetry* and these errors are important because they provide examples of when we find the Clark "gene" in *Poetry* magazine. That is, the errors are moments in which Clark is where he should not be – his style is present in *Poetry* despite its general systematic exclusion of his work. Here is one example of a misclassified poem, a Clark poem misidentified as a *Poetry* poem from the year 1919, a year in which the machine has otherwise little difficulty in differentiating Clark from *Poetry* (image 13). It would be worth explicating this Clark poem to glean aspects of his work that *were* acceptable to *Poetry* magazine.

So far, however, our model tells us things of limited interest. Clark's choice of words excluded him from modernism's center. This was probably because he used a lot of overt religious language, and *Poetry* magazine tended to exclude this kind of poetry. Sometimes we can find Clark's "gene" in *Poetry*, but this too makes sense because we know *Poetry* did accept religious-inflected verse on occasion. Where the model gets

even more interesting, though, is when we scale up and examine more writers using the same method. We get a sense of how modernism performs as a large system, and how different poets measure against each other. We chose 7 poets active during the height of Clark's exclusion – Witter Bynner, Lowell, Sandburg, Paul Eldridge, Georgia Johnson, Mary Davies, and Sara Teasdale – and ran the same type of classification on their corpora of poems over the same period. We chose these poets because they represented a mix of "outliers" like Clark and "connectors" who represented the exact opposite. Here, again, is the network map of 1915-1919 with the 7 poets highlighted (image 14). And here are the classification results (images 15-16).

Most immediately, the results confirm our suspicion that the classifier does better as the decade moves on into the early 1920s. This is probably because of both a technical reason – the more poems the machine has to choose from, the better it can classify poets – and because of a historical reason – as the decade wears on, the American modernist field becomes more differentiated as poets become better at defining their styles. At the same time, the results contain several surprising discoveries. For example, the machine finds it very easy to distinguish Sandburg's poems from *Poetry* magazine, but this is quite odd because Sandburg represented a staple of *Poetry*. In a sense, we expect him to blur into *Poetry* magazine given his importance to the journal. It perhaps suggests that indeed what makes him so attractive to *Poetry* is because he didn't sound like anyone else in the journal. But how then do we explain Clark's exclusion? In terms of our textual classification approach, the machine finds that Clark and Sandburg are equally different from *Poetry* magazine, yet one is happily accepted by the journal and one is routinely rejected by it. This becomes all the more complicated when we consider that

Sandburg and Clark thought of themselves as aesthetic comrades. They saw no real differences between them and this self-perception is confirmed by our model.

There are other curious aspects to the results. Take for example what we find for Georgia Johnson, an early Harlem Renaissance African-American poet, who like Clark, never was able to publish in *Poetry* magazine. We can think of her as a fellow outlier to Clark, even though obviously in content and form, she differs greatly. What is startling is that the machine has trouble distinguishing Johnson from *Poetry* magazine, and here this is surprising. At least based on the words she was using, she isn't significant enough to stand out from the words being used in *Poetry*. Her language is more generic, in a sense, but not generic enough that the machine classifies any *Poetry* poems as hers. Compare this with Amy Lowell, a fixture in *Poetry* magazine during the late teens. The machine classifies her poems with roughly the same accuracy, but the errors are different. It is much more likely to find Lowell-like poems in *Poetry*, but not Johnson-like poems. All the same, we're still left with a profound question: what does it mean that *Poetry* magazine is less distinct from Johnson, a Black female poet, than from Clark, a white, male, middle class Chicago poet with direct connections to the journal? What does it mean that a machine-based algorithm that uses linguistic presence and patterns to classify poets will have as much trouble classifying a poem by Lowell, the "face" of *Poetry* magazine, as it does a poem by Johnson, a Black poet on the margins of modernism?

There are many questions still to be answered, but they are not questions we could have asked without macro-scale techniques. These techniques allow us to operationalize the theory that different poets, such as Clark, Lowell, and Johnson, who are typically seen as completely unrelated, worked within a shared system. In the mid-1920s, Clark wrote a

self-reflexive poem about his place in this system, titled "A Poet Departs." It may as well have been a valediction for his own career and life: "He lived amid shadows/He shunned open roads … Saw men in a mirror - /Men twisted and tortured by hard circumstance/ Keen was his glance/As he took note of men - /But he lived amid shadows." To writers like Clark, modernism could be a shadow world of exclusion and deprivation. But it was still a world they could feel a part of, a world that could be observed, a world with some kind of discernible logic or pattern. Pushed to the corners of that field, Thomas Clark felt such patterns keenly and oppressively. Macro-scale methods now offer a way to discern some of these patterns that defined modernism as a system or ecology. Patterns that tell us something important about how modernism, religious poetry, and Black poetry might have functioned through or against each other based on rhetorical disposition rather than identity-based categories, and in turn, how rhetorical forms functioned to produce social groupings and movements outside what we generally recognize as modernism. Utilizing Clark's scrapbooks and our new tools, we can now bring some light to these shadows.