A COMPUTATIONAL ANALYSIS OF TOPICS AND TONE IN AMERICAN LABOR AND MAINSTREAM NEWSPAPERS, 1909-1911

Vilja Hulden, LAWCHA 2019

Reading the newspaper is a curious phenomenon. It is part ritual, part pragmatism; part habit, part choice; part pleasure and part a chore. It is wrapped in multiple layers of stark connotations: reading the newspaper is an act of the informed, deliberating citizen, of a serious member of the community, of an individual with duties and aspirations. It is performed in settings that themselves are deeply invested with meaning (as well as highly gendered): the home, where the family father signals his responsibilities as well as his privileges by raising his newspaper as a shield against the mundane cares of the wife and children; or the commute to the workplace, where the sober, conscientious (male) workers make good use of the downtime between home and workplace (office) to catch up on world events.

These settings and connotations are also intimately bound up with class: they exude middle-class civic-mindedness and domesticity and ambitions of upward mobility. By the early twentieth century, journalism as a profession was increasingly coming to embrace these ideals, raising the flag of "objectivity" and, more and more frequently, rejecting an explicit partisan affiliation, though not necessarily with results that strengthened democracy.[1]

Yet the newspaper was not the exclusive domain of the educated middle and upper classes. Not only did papers ("yellow" and otherwise) attempt to cater to the large working-class audience, labor and working-class organizations—often complaining of the biases that most mainstream newspapers had against them—set up their own papers. The faith in rationality, in enlightening the masses, and in self-improvement that animated middle-class reformers also echoed among worker intellectuals, if in a somewhat different key. Labor was generally very dissatisfied with the portrayal of worker movements in the mainstream press, complaining, for example, that major events organized by labor unions went unremarked in the local press[2]. Organized labor therefore invested considerable resources in creating and maintaining newspapers devoted to providing a worker point of view.

There were hundreds of labor newspapers in the early-twentieth-century United States, even if one only considers bona fide newspapers and does not count the union trade press (that is, publications like the *American Federationist* or the *Shoeworkers' Journal*.) Many of those publications were short-lived, to be sure, but many a locality nevertheless had an organ providing news from a labor perspective on a daily or weekly basis, and these, along with non-English-language labor papers, offered a different set of news stories and different takes on stories to a significant slice of the American working class.

Using a set of computational methods, this paper takes a birds-eye view of how the content in a sample of labor papers compared to "mainstream" weeklies and dailies published in nearby towns. This kind of content analysis of course in no way replaces close reading and archival research when it comes to understanding the positions these papers took on various issues, the kind of constraints they labored under, or the tone they employed on particular questions (say, woman suffrage, or immigration.) Nor does it attempt to do so. Though sensitive to context in some ways, the computational methods cannot parse sentences and arguments in the way a human does.

---

[1] Richard L. Kaplan, *Politics and the American Press: The Rise of Objectivity, 1865-1920* (New York: Cambridge University Press, 2001).

[2] See e.g. Jon Bekken, "The Working-Class Press at the Turn of the Century," in *Ruthless Criticism: New Perspectives in U.s. Communication History*, ed. William S. Solomon and Robert W. McChesney (Minneapolis: University of Minnesota Press, 1993), 155.

Instead, computational content analysis focuses attention elsewhere: on the sheer frequency of words, on the associations those words carry in different bodies of material, on what is mentioned and what is not. Such analyses are in some ways very crude: mostly, they simply involve counting words, even if the counts are then put into the service of fancier algorithms. Yet despite—and even perhaps because of—that crudeness, they can underline for us some basic features of newspapers that may be quite important—especially given that many of us do not, in fact, read the newspaper in the civically-ideal way of perusing the news stories with careful consideration and cogitation.

DATA AND METHODS

The "data set" used in this paper consists of the full run of 16 newspapers for the years 1909-1911 (25,897 pages, about 56 million words), downloaded from the database Chronicling America.[3] The papers include 4 labor papers published in four different states (Minnesota, Washington, West Virginia, and Nebraska) and 12 mainstream papers (2-4 per state from towns nearby the place of publication of the labor papers.)

affil	*title*	*publication*	*place of*
DEM	Valentine	Weekly	Valentine, NE
DEM	Little Falls	Weekly	Little Falls,
GOP	Fairmont West	Daily	Fairmont, WV
GOP	Omaha Daily	Daily	Omaha, NE
GOP	Clarksburg	Weekly	Clarksburg, WV
GOP	Bemidji Daily	Daily	Bemidji, MN
GOP	McCook Tribune	Triweekly	McCook, NE
GOP	Colfax Gazette	Weekly	Colfax, WA
LABOR	Labor Journal	Weekly	Everett, WA
LABOR	Labor World	Biweekly	Duluth, MN
LABOR	Labor Argus	Weekly	Charleston, WV
LABOR	Wageworker	Weekly	Lincoln, NE
UNAFF	Princeton	Weekly	Princeton, MN
UNAFF	Leavenworth	Weekly	Leavenworth,
UNAFF	Ellensburg	Weekly	Ellensburg, WA
UNAFF	Lynden Tribune	Weekly	Lynden, WA

Table 1. Papers in the data set.

The Optical Character Recognition (OCR—the conversion of images of text into actual computer-readable text) quality of the material is quite poor, which makes some types of analyses (e.g. methods that require grammatical parsing) challenging. The analyses here make use of two methods that are fairly robust even given poor OCR: topic modeling and word embeddings.

Topic modeling, while algorithmically fairly complex, is in principle a simple enough method: it attempts to identify "topics" in a body of texts based on word cooccurrence in documents. The order

[3] https://chroniclingamerica.loc.gov/

of words within a document does not matter, nor does the user "suggest" topics; instead, the algorithm processes a set of documents to extract a user-defined number of topics, represented as a list of the words that most strongly characterize each topic. In addition to this list of topics, the topic modeler also produces a calculation of how prominent each topic is in each document, which enables one to examine the distribution of topics within different subsets of the material.[4] The main use of topic modeling here is to examine the differences in distribution of topics in the labor papers versus the mainstream newspapers.[5]

Word embeddings construct numerical representations of words as they are used in a corpus. The goal is often to produce a representation of the vocabulary such that words that often occur as neighbors, such as "wrote" and "book", are judged similar (first-order co-occurrence), but also that a second-order similarity is revealed: words that themselves may not be neighbors, but tend to occur next to the same neighboring words should be numerically close to each other, such as "wrote" and "said". The basic principle for achieving this is simple: define a context window, and for each word, collect counts of the words occurring within the word's context window. In other words, the method is based on the old adage that you shall know a word by the company it keeps.

The basic procedure is to first construct a matrix that has the vocabulary on the x axis as well as the y axis, and each cell contains a count of how often the x axis word has appeared in a context window that also contains the y axis word. For example, in Table XX, the word *cat* has never appeared in a context window that also contains *freedom*, but has appeared six times in a context that also contains *queen*. So our corpus might contain sentences like "the queen had a pretty white cat" or "the cat queen was evil" or "the queen's cat was fluffy" but not a sentence like "the queen guaranteed the freedom of her people."

	cat	queen	dog	freedom	watermelon
cat	0	6	3	0	0
queen	6	0	0	8	0
dog	3	0	0	2	1
freedom	0	8	2	0	0
watermelon	0	0	1	0	0

Table 2. Toy example of a word count matrix.

These counts are then transformed into vectors of numbers, with each word being represented by a different vector constructed from its context words (here, the vector for *queen* is 6 0 0 8 0.) Representing words as vectors allows one to perform various calculations on them: for instance, how close together are two vectors—i.e., how similar are the words represented by the vectors? This works remarkably well, and one can for example calculate analogies like

[4] There are a variety of topic modeling algorithms and software; the one used here is MALLET, Andrew Kachites McCallum, "MALLET: A Machine Learning for Language Toolkit," 2002, http://mallet.cs.umass.edu

[5] Chronicling America contains news material as pages rather than as separate articles. Because a page of newspaper text is not a particularly coherent "document" and thus word cooccurrences on a page is not particularly meaningful, the material used here was chunked into smaller units for topic modeling. Dividing material into articles reliably is a difficult problem; the method of chunking used here was the rather crude but relatively useful one of splitting texts by the occurrence of four consecutive capital letters, as news headlines are commonly in all capital letters.

```
man : woman :: king : ?
```

and get the answer `queen`. In other words, *queen* and *king* appear in similar context positions; in a sentence like "The crowd cheered the king, who waved from the balcony," one might easily substitute "queen" for "king."

The simple (though very large) matrix of context counts is then further crunched through probability and data compression algorithms to make it more robust and keep it to a manageable size while retaining the most important information.[6]

In this paper, word vectors are used to compare the 'meaning' of particular words—as revealed by their context—in the labor papers versus the mainstream newspapers.[7]

Topic analysis

After some experimentation, the topic model selected for analysis here here consists of 150 topics. One topic model was created out of all the material, but with filenames containing the identifying information of each paper. The results were then postprocessed to average the frequency of each topic in each newspaper title so as to enable comparisons across titles rather than only across individual documents. Since the most of the titles also had an explicit political/class affiliation (Democratic, Republican, or "labor"), it is also possible to draw comparisons between these broader categories.

The main focus here is on how the content in the labor papers differs from the mainstream content. Let's start by examining that content. That is, excluding the labor papers for the moment, what topics are the most common in these papers?[8]

[6] The specific methods used here are calculating positive pointwise mutual information (PPMI) and carrying out a Singular Value Decomposition (SVD) compression.

[7] Scholars have also used word embeddings to examine the shifting context of particular words—for example, how the word *gay* has changed meaning over the course of the twentieth century.

[8] That is, for each topic, after averaging the prominence of the topic in each newspaper title, take the average of the averages for all titles except the labor titles.

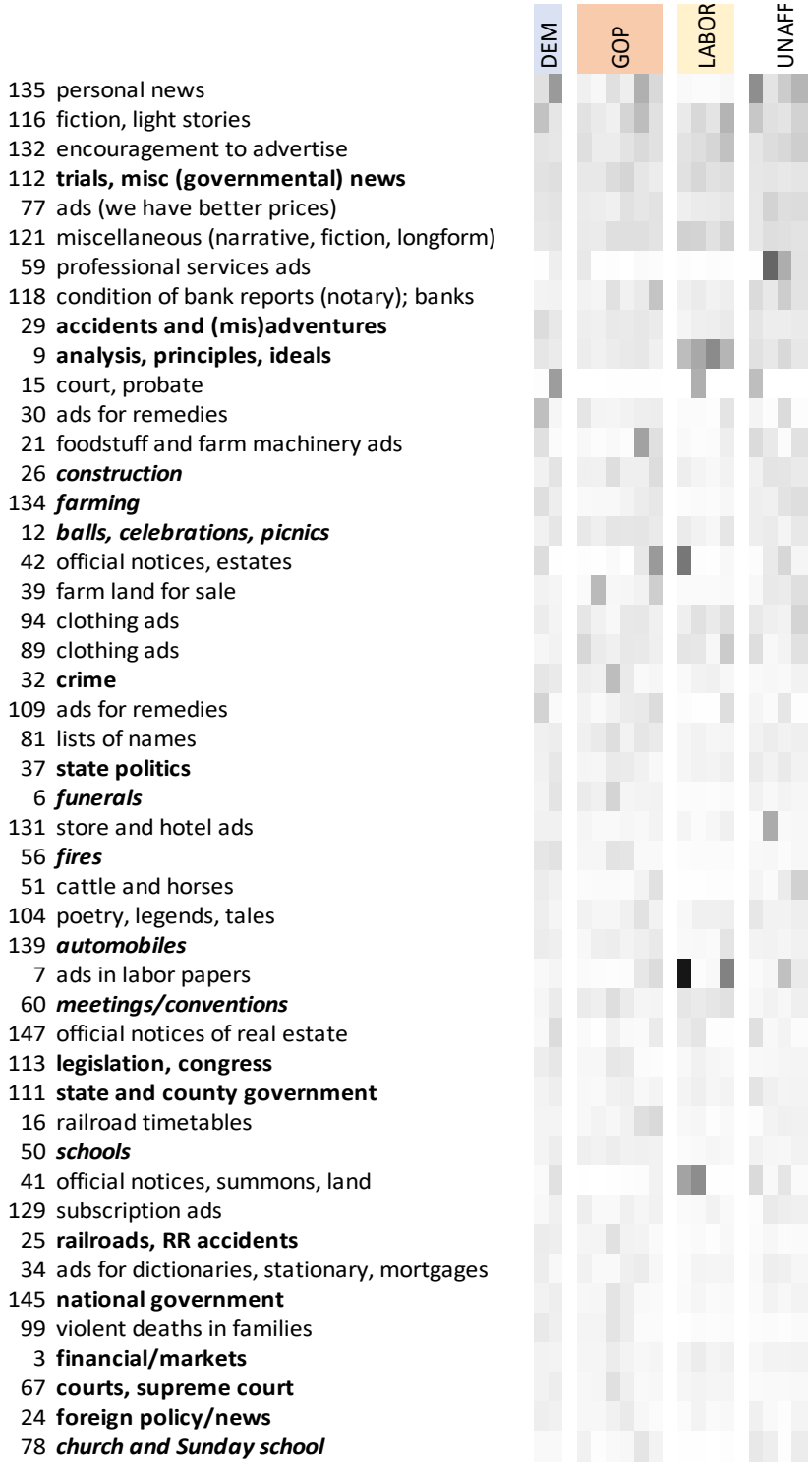

Figure 1. Top topics by average in all material except labor. News and editorial topics in **bold**, *topics containing a mix of notices/advertising and news/editorials in* ***bold italic.*** *Each square represents the strength of a topic in a particular newspaper title; the darker the square, the more prominent the topic was in that title.*

As Figure 1 shows, the topics that are at least in substantial part composed of news consist of roughly the kinds of material that one would expect: news about the government and politics and about accidents, crime, and violence, with a mix of (mostly locally based) stories and notices about celebrations, school and association goings-on.

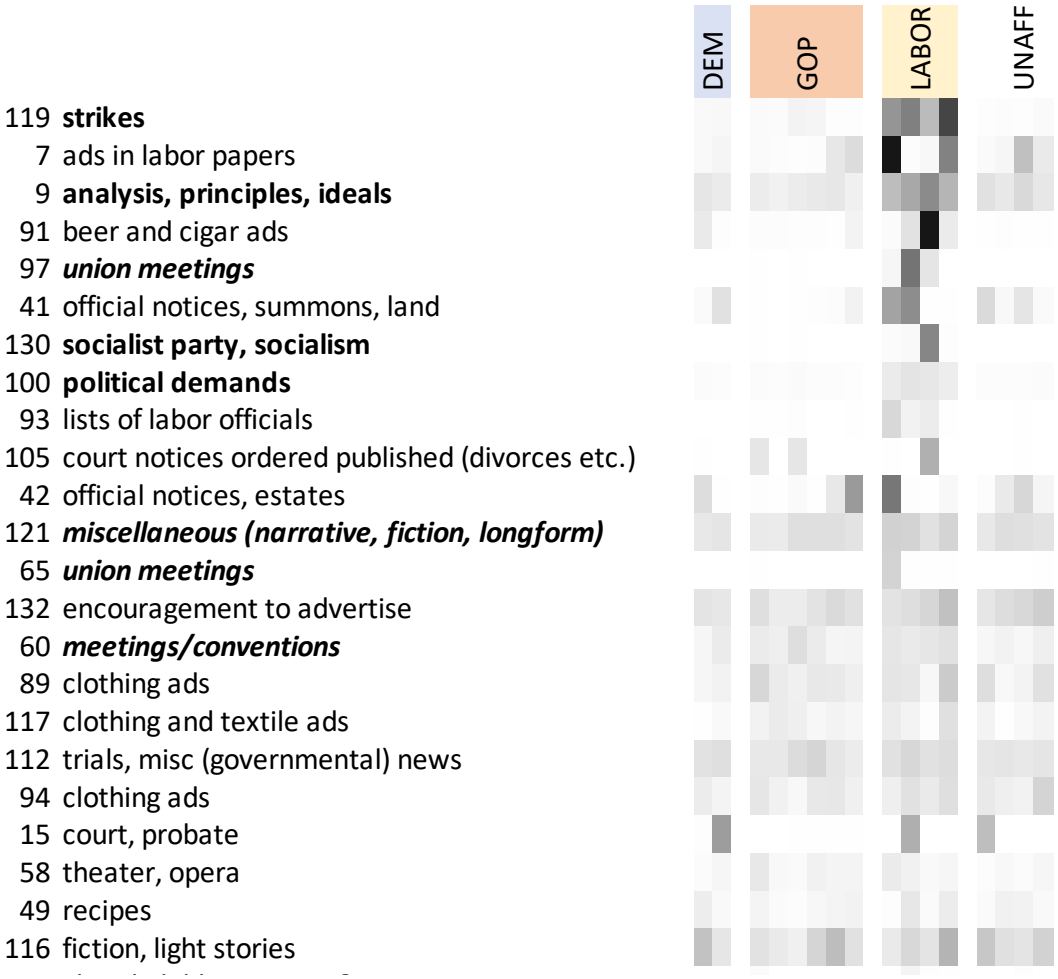

Figure 2: Topics most characteristic of the labor material.

If one compares this list to the topics in Figure 2, which shows the topics most characteristic of the labor material,[9] the most obvious difference is of course the prominence of news about strikes, which tops the list and is prominently represented in all the labor papers (whereas it comes somewhere about 2/3 of the way down when its prominence is averaged across all the other papers.) The other two sets of news material that are clearly much more prominent in the labor papers than in the mainstream papers are the "analysis, principles, ideals" category, news about labor and socialism, and the category of "political demands." The first contains a mix of editorial-type material, reports of speeches, et cetera, that cultivates the fairly flowery language often common in political rhetoric and laying out of first principles.[10] The second is fairly self-explanatory, and its clear prominence in the *Labor Argus* is hardly surprising, as the *Argus* had a history of supporting workers' efforts to self-organize politically

[9] That is, the topics that are prominent in the labor material but not in the mainstream material – calculated as the average of the topic in the labor material minus the average of the topic in all the non-labor material.

[10] The list of words most prominent associated with the topic is "`people men man great country life good public law power world government american time true things fact human free political`."

as well as economically, and became an explicitly Socialist paper in 1911 under the editorship of Charles H. Boswell.[11] The third contains material like Socialist Party platforms, political demands of state labor federations, and political demands of the American Federation of Labor (AFL).

The differences in the news topic composition, while not perhaps surprising, are nevertheless significant—in my view, in two somewhat contradictory ways. First, they underline how important the role of the labor press was in distributing news ignored or slighted by the mainstream press. The labor press provided the news about strikes that the mainstream press did not bother with; it made sure that there was a forum for distributing labor's political demands; (sometimes) it represented the viewpoint of the Socialist Party; it invested heavily in adding labor's vision to the mix of political rhetoric (represented by the "analysis, principles, ideals" topic.) Surely this was a crucial task. Yet it also brings us the contradictory second point. Namely, the labor press did not provide the daily news. It was explicitly *in addition to* rather than *in lieu of*. These topics, at least, seem to indicate that a person who wanted the usual newspaper fare of news about accidents, government, and local goings-on could not simply subscribe to the local labor paper, but would need to get the local paper for general news and the labor paper for labor news.

This focus on distributing the labor viewpoint to the exclusion of the "usual" job of a newspaper is even clearer when it comes to non-news material. In some ways, the thing that is most striking about the main topics in Figure 1 is *how much the newspaper is about other things than news*. The two most common topics are "personal news" (local society pages reporting things like the visit of Mrs. Kinnicaid's New York cousin or the fact that the Misses Nelson have gone to visit their aunt in Seattle) and fiction or other entertaining narratives. Both are more about amusement than news; and both, perhaps, are also about community, the first explicitly and the second because the fiction in the local newspaper surely provided material for chats at the greengrocer's or the post office. Of these, the personal news shines by its absence from labor papers (Figure 3), though there is some fiction and narrative; on the other hand, some other surely important categories like classified ads about housing, notices of church and Sunday school meetings, and advertisements by local doctors and other professionals are also absent.

[11] "About the Labor Argus," *Chronicling America*, https://chroniclingamerica.loc.gov/lccn/sn85059855/

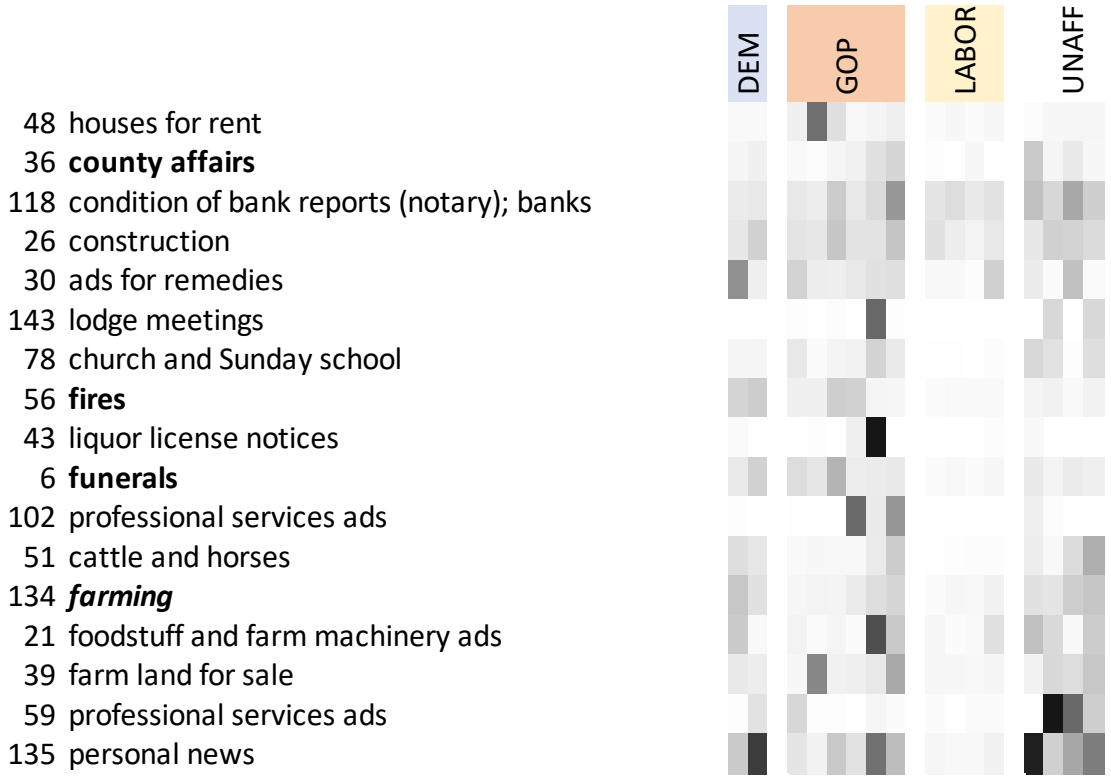

Figure 3: Topics least characteristic of the labor material.

One should not exaggerate the significance of these absences; as Figure 1 reveals, the labor papers did contain (versions of) much of the same material as the mainstream papers, and it was hardly a failing that often that material had an explicit focus on labor (for example, news and notices about celebrations that in mainstream papers could be anything from dress balls to school picnics were in labor papers mostly about Labor Day events.) And it does nothing, of course, to detract from the importance of providing labor news and working-class points of view on the news. Yet it is worth noting that this made the labor papers very different in their function from run-of-the-mill papers, which often flourished (or scraped by—note the prominence of "encouragement to advertise" topic in all the papers in the sample) at least as much because people needed them to get the train timetables and the notices about lodge and Sunday school meetings as they did because their readership was committed to staying informed about current political events.

WORDS IN CONTEXT

Topic models can provide a representation of *what* gets talked about, but they don't say much about *how*. One way to get at a representation of that "how" is to use word embeddings to represent the meanings of words—in this case, the ways in which their meanings differ between the mainstream newspapers versus the labor papers.

Once one has the word embedding vectors, one can try to make sense of them along a variety of analytical axes. One can measure the distance between two vectors representing the word in each set of material; one can look at which word vectors are close to the target word in each set of material; and one can examine what words most commonly surround the target word in each set of material. These are examined for a few key words in table 3.

There are of course various reasons why the "meaning" of a word (as defined by its context) might differ between two sets of material. Sometimes it is simply a completely different sense of a word that

word	*mat*	*sim*	*similar to*	*context word ex*
coats	main	0.96	suits, skirts, overcoats, jackets, tailored, dresses, worsted, capes, wool	suits, worth, long, price, ladies, mens, fur, rain, childrens, values, sweater
coats	labor		suits, jackets, skirts, 1250, dresses, priced, plush, 1950, worsted, overcoats	suits, ladies, price, sale, long, childrens, sweater, mens, new, fur, dresses
bodies	main	0.68	victims, ruins, burned, supposed, missing, lives, body, burning, believed, unknown	dead, found, recovered, desired, charge, commercial, director, victims, mine, ruins, burned
bodies	labor		body, organizations, central, various, representatives, affiliated, formed, civic, unions, clubs	central, labor, federation, state, american, officials, unions, embracing, local, minds, city
scab	main	0.39	parasites, germ, worms, eradicate, parasitic, germs, insect, preventive, scalp	apple, seed, treatment, solid, potatoes, disease, diseases, blight,
scab	labor		scabs, quit, hire, jobs, employ, hired, refuse, job, insist	union, labor, coal, clothes, fields, tobacco, cigars, nonunion, company
strikers	main	0.90	strikebreakers, nonunion, walkout, sympathizers, riot, rioting, foreigners	men, union, company, police, victory, committee, demands, city, meeting, aid
strikers	labor		strikebreakers, miners, scabs, refused, threatened, officials, bosses	men, company, today, meeting, car, police, return, , union, street, ranks, city, committee
bosses	main	0.82	politicians, dictate, faction, tammany, officeholders, congressmen	mine, foremen, fire, party, political, democratic, republican, state, machine, examinations
bosses	labor		operators, refuse, tactics, officials, threatened, strikers, contractors	political, strike, union, workers, foremen, mine, strikers, labor, party, public, fire
courts	main	0.87	decision, ruling, federal, supreme, jurisdiction, injunction, proceedings, legal	federal, state, practice, law, decision, case, supreme, court, county, lawyer, states, washington, circuit
courts	labor		judges, injunction, legal, supreme, justice, injunctions, federal, rights, decisions, contempt	labor, law, state, justice, federal, power, legislatures, states, people, judges, corrupt, supreme, injunction
capitalism	main	0.35	bulwark, injustices, molder, usurping, libertyloving, autocracy, wageworking, serfs, industrialism	active, britain, commonwealth, despotic, great, ills, impossible, indictment, institutions, legal, monopoly

capitalism	labor		greed, slavery, oppression, capitalist, slave, masters, tyranny, masses, slaves, ignorance	workers, force, preparation, greed, masses, insatiable, welfare, mental, poverty, socialism, political, class
capitalist	main	0.47	financier, magnate, philanthropist, millionaire, banker, dobbs, rockefellers, onetime, wealthy	banker, york, city, new, prominent, business, chicago, dead, man, died, estate, george, home, known, real, son, john, late
capitalist	labor		capitalists, capitalism, class, capitalistic, trusts, nation, masses, masters, slave, greed	class, system, working, rule, power, present, struggle, capitalist, press, workers, society, profits, oppression,
socialism	main	0.82	doctrines, doctrine, viewpoint, socialistic, morality, propaganda, believer, theories, ethical, aims	socialist, church, means, people, men, socialists, labor, party, capitalism, system, unionism, principles, vote, great
socialism	labor		economic, unionism, common, nation, socialists, democracy, masses, true, labors, capitalist	state, socialist, lecture, people, principles, antitrust, forms, human, party, public, church, city, country
socialist	main	0.79	socialists, dahlman, tammany, nominee, democrat, polls, reelection, hearst, bryans, nominated	party, candidate, mayor, vote, city, ticket, elected, member, milwaukee, berger
socialist	labor		party, socialists, political, campaign, issue, vote, press, democratic, socialism, republicans, votes	party, socialist, ticket, labor, vote, national, movement, press, city, milwaukee

dominates in one body of material than in another. In this case, for example, the word *bodies* in the mainstream press material is close in meaning to words like *victims, ruins, burned, supposed, missing, lives, body, burning, believed, unknown* — whereas in the labor material it is close to words like *organizations, central, various, representatives, affiliated, formed, civic, unions, clubs.*[12] In the mainstream material, the top five context words for *bodies* are *dead, found, recovered, two, men*; in the labor material the top five context words are *central, labor, federation, state, american*. Obviously, *bodies* in the mainstream material mostly appears in stories about accidents, whereas in the labor material it mostly appears in stories about organizations.[13] A similar example is *scab*: in the mainstream material, similar words include *parasites, germ,* and *worms*, while in the labor material similar words are all about strikebreaking: *nonunion, quit, hire*. (Interestingly, the context words seem to indicate that the labor material uses *scab* as a perfectly routine term, not only to berate: the top 25 context words contain no negative adjectives but rather are all business: *union, coal, clothes, tobacco,* et cetera.)

[12] That is, these words appear in similar contexts as *bodies*.

[13] As noted above in the discussion of the topic model, the labor papers lacked much of the regular news material prominent in the mainstream press, and the *bodies* example fits that.

Table 3: Sample words and their similar and context words. The "sim" column shows a similarity measure between the word's context vector in mainstream and labor material.

COMMENTS ON THE EMBEDDINGS

Strikers: Note the proximity to *riot*, *rioting*, and *foreigners* in the mainstream material. On the other hand, clearly it is the work stoppage sense of striking (rather than, say, the word's baseball meaning) that dominates mainstream material as well, perhaps attesting to the presence of labor strikes in the news despite the topic's low prominence in the topic model. Note also that the similarity measure is quite high; that is, the contexts do not differ that dramatically in the two sets of material.

Bosses: This is more of a case of two different senses mixing; in the mainstream material the political meaning clearly dominates, while the labor material emphasizes the employment context—as the context words reveal, both materials contain both senses, but the similar words would seem to indicate that mainstream material was much less likely to call employers "bosses."[14]

Capitalism: The labor material, unsurprisingly, is rather starkly critical of capitalism; note also that capitalism is discussed in the context of socialism in that material, which is not the case in the mainstream material. The mainstream material is more difficult to interpret. On the one hand, with the exception of *libertyloving* in the similar words and *great* in the context words, there are few indisputably positive words. On the other, the references to *serfs, autocracy,* and the like might be due to comparisons to a feudal past (presumably to the glory of capitalism.) However, *capitalism* simply does not appear much in the mainstream material, so the word vectors are not reliable: the word appears once in the mainstream material for roughly every 675 times it appears in the labor material, and indeed, only once for roughly every 50 times the mainstream material mentions *socialism*—a silence interesting in itself.[15]

Capitalist: Much more common that *capitalism* (appearing once roughly every 40 times it appeared in the labor material, and once roughly every 4 times that *socialist* was mentioned), *capitalist* also has a much more obviously different connotation in the two sets of material, evident in both the similarity measure and the differences in similar words and contexts. In the labor material, it is linked to *class, greed*, and *masters* and is mentioned firmly in the context of a critique of capitalism. In the mainstream material, it is linked to *philanthropy, magnates* and *rockefellers* – and the context words (besides underlining the role of New York City and Chicago as hubs of capitalist activity) seem to indicate that it was commonly used in obituaries to describe the deceased (note the words *dead, died, estate, son*), and thus presumably used as a term of respect and praise.

Courts: Unsurprisingly, the labor material similar words contain both *injunction* and *injunctions*, the main issue about which labor in this period worried when it came to courts. Although the similarity measure indicates that the contexts are not widely divergent, note how the mainstream material focuses on procedure within the courtroom (*practice, decision, case, lawyer*) whereas the labor material seems to put courts in a much larger context (*justice, power, legislatures, people.*)

Socialism and *socialist:* Somewhat surprisingly, the representation of *socialism* does not differ all that dramatically between the two materials; although the similarity measure for *socialist* seems to indicate more difference than for *socialism,* the similar words and context words do not quite bear this out. Clearly, both words are in both materials rather substantively tied to the Socialist Party rather than simply to a socialist ideology; this is particularly the case for *socialist*. Note, however, that the main

[14] That *bosses* represents politicians rather than employers in the mainstream material is also indicated by the fact that the vectors for "politicians" and "democrats" constructed from within the labor material are very close to the vector for "bosses" based on the mainstream material.

[15] *Socialism* is more common than *capitalism* in the labor material as well, but only by a factor of about 2.5, not 50.

material seems to discuss *socialism* in the same terms as *doctrine* or *propaganda*, while the labor material is more focused on the *economic* aspects and on *democracy* and the *masses*.

CONCLUSION

If we imagine "regular" reading as seeing the image in a tapestry, we could think of the computational analyses of text as examining the materials of which the tapestry is woven. Simple frequency counts as well as more complex methods like algorithmically extracted topics and word vectors that simulate the "meaning" of a word through its context underline the importance of the content beyond its specific purport.

Both the differences in topics and the differences in the ways a number of key words are used in the labor papers versus the mainstream papers underline the important work the labor papers did to advance a working-class interpretation of the contemporary reality. This is particularly clear in a few of the examples above, such as the near-complete absence of the very word *capitalism* from the mainstream material and the overwhelmingly positive connotation of *capitalist* as a term of respect accorded to prominent men in their obituaries, or the prominence of strike news in the labor papers compared to their quite limited presence in the mainstream papers. It is also evident in more subtle ways in the labor papers' emphasis on news and editorial material focusing on analysis, principles, and ideals, as well as in the diverging vocabularies around *strikes*, *courts*, or *bosses*.

On the other hand, this work of informing and educating the working-class public only paid off if the working-class public in fact subscribed to and read the labor papers, and if the papers survived. Labor papers faced many struggles, ranging from outright suppression to advertiser avoidance, as has been well documented by scholars. As we have focused on the civic-minded reader, however, what has been less remarked upon is the extent to which many American newspapers survived due not to their brilliant journalism but to their practical function as a community glue and source of entertainment. That labor papers had less ability to (and/or interest in) fulfill those community and entertainment roles may be more significant in their struggles than we have realized.