

Rallying the Mass Media: News Networks and the Geography of Political Campaigns

Santiago Cantillo-Cleves*
UC San Diego

October 7, 2026

[Most recent version here](#)

Abstract

Around the turn of the twentieth century, mass media expanded and politicians could increasingly reach voters without visiting them. This expansion may also have changed where campaign travel was most valuable, because a speech delivered in a well-connected place could reach a larger audience elsewhere. I study whether the expansion of news agencies, which collected news and supplied it by telegraph to subscribing newspapers, changed how U.S. presidential campaign speeches spread and where candidates chose to campaign between 1880 and 1916. Analyzing millions of newspaper articles, I construct data on campaign speeches, their coverage, and each newspaper's access to a news agency. When newspapers in both the county where a speech was delivered and another county subscribed to a news agency, the probability that the latter covered the speech increased by 1.5 percentage points, from a baseline of 0.4 percent, with larger increases at shorter distances. I use these estimates to construct each county's *media reach*, the expected number of outside voters reachable from a speech there. Exploiting the evolution of the news-agency network elsewhere while holding local access fixed, I find that a one-standard-deviation increase in media reach raised the probability that a county hosted a speech by 2.3 percentage points, from a baseline of 4.8 percent. Mass communication thus shaped both how political messages traveled and where politicians chose to deliver them.

Keywords: political campaigns; mass media; news agencies; information diffusion; text as data; economic history.

JEL codes: D72, D83, L82, N41, N71.

*Corresponding author: scantilocleves@ucsd.edu. Website: <https://santiago-cantillo.github.io>
I am grateful to Sam Bazzi, Gordon Dahl, Sara Lowes, and Craig McIntosh for their guidance and support.

“[...] I want to address that larger constituency which we reach through the newspapers, for it is a thousand times as numerous as any crowd that could assemble here [...].”

—William Jennings Bryan, August 8, 1900

1 Introduction

Political campaigns are strategic in what they say, when they say it, and how they disseminate their message. A less obvious choice is where to send that message from. Candidates cannot speak to every voter in person, so they rely on the media to carry their words beyond the places they visit. Before radio, local newspapers reported speeches given nearby, and news agencies, such as the Associated Press, collected those reports and sent them by telegraph to subscribing newspapers elsewhere. If the value of a campaign stop were only the crowd it drew, candidates would concentrate on the most populous places. Yet between 1880 and 1916, a candidate who made ten stops in a state typically skipped four of its ten most populous counties for smaller ones. The press may explain why. A smaller town well positioned in the news-agency network could offer a much larger audience than its population alone would suggest. Research on media and politics has emphasized the effects of *receiving* political information (e.g., Wang, 2025; García-Jimeno et al., 2022), but we know less about whether the media also shape where politicians choose to *send* their message from.

This paper studies how the expansion of news agencies changed where U.S. presidential candidates chose to campaign. Using text from millions of historical news articles between 1880 and 1916, I build a novel dataset that records where candidates spoke, where newspapers covered each speech, and which newspapers had access to a news agency. I first study how the news-agency network affected the geographic diffusion of campaign speeches. I show that a speech was more likely to be covered by newspapers in another county when newspapers in both counties had access to a news agency. I use these estimates to construct each county’s *media reach*, the number of additional voters in other counties that a speech given there could reach through the news-agency network. I then ask whether candidates responded to differences in reach, holding each county’s news-agency access fixed and leveraging changes in reach that come from *other* counties gaining access. I find that candidates directed more of their stops toward counties with greater media reach. Media networks can therefore shape who is exposed to a message even before it is transmitted, by changing where it is sent from.

The United States’ late nineteenth century combines a changing news-distribution system with unusually rich newspaper data. The telegraph had reached much of the settled country by the 1880s, yet covering and distributing national news still required costly coordination. News agencies supplied it by collecting reports centrally and sending shared reports to subscribing newspapers. Demand for these reports was widespread, and admission rules decided which newspapers had access. The Associated Press (AP) let an existing member veto a local competitor, so access often turned on which paper in a town had joined first, while the commercial agencies that entered after 1907, such as the United Press, sold their report to any newspaper that could pay (Le, 2026). At the same time, presidential candidates increasingly took to the campaign trail, so they were choosing where to appear while the audience reachable from each place was changing.

Newspapers were the main source of political news at the time, so their text contains most of

the political information voters had. I build the data from 438 million articles across 20 million newspaper pages (Dell et al., 2023). From the 1880 to 1916 campaigns I extract nearly 330,000 articles that mention a presidential candidate, and use a large language model (LLM) to detect campaign speeches and their locations. The LLM recovers 86 percent of the counties on Bryan’s 1896 itinerary, 88 percent of the counties it detects are in the record, and the number of counties visited per campaign in my data has a correlation of 0.98 with the counts of Pino and Salisbury (2025). Combining these labels with newspaper headquarters locations links each speech to the counties where newspapers reported it. Because agency copy reached every subscriber by wire the same day, while other channels took days, I infer a newspaper’s access from how often, in previous years, it reprinted a non-political story within a day of its first appearance elsewhere, using 400,000 stories reprinted 2.7 million times (Silcock et al., 2024). Over time, more newspapers printed same-day copies of articles from distant newspapers, consistent with the expansion of news agencies.

To understand how news agencies changed political behavior, I first examine how they changed the flow of political information. I estimate the change in the probability that a speech given in an origin county was covered by newspapers in a destination county once both counties had gained access to a news agency, similar to a gravity equation in the trade literature. However, news-agency access was not random. Growing counties may have both demanded more news and attracted more campaign activity, while some county pairs were persistently more likely to exchange information. I therefore identify the effect from the same county pair before and after both counties first gained access, accounting for all contemporaneous changes in either county and for persistent differences between pairs.¹ Once both counties had access, the probability that newspapers in the destination county covered a speech in the origin county increased by 1.5 percentage points, from a mean of 0.4 percent among unconnected county pairs. Although news agencies removed one of the supply barriers for the flow of information and made distant events easier to cover, geographic relevance continued to matter as the increase was concentrated within states and declined with distance.

I then translate these estimates into a county-level measure of *media reach*. For each potential speech location, I predict the increase in coverage probability from the pair’s distance and news-agency access, multiply it by the number of voters there, and sum across destinations. Media reach summarizes a county’s position in the evolving news-agency network, analogous to “market access” in studies of transport networks, where the value of a location depends on the size of the markets it can reach at low cost (Donaldson and Hornbeck, 2016). Two counties can therefore have the same population and news-agency access, yet face different outside audiences depending on how accessible the voters in nearby counties are. While candidates were unlikely to know the network precisely, contemporary accounts show that they courted the press to reach readers beyond the local crowd.

I now turn to candidates’ response. Because a county’s own access to news agencies may reflect its growth or political importance, I hold its *own* local media environment fixed and identify the effect of media reach on campaign location choices from changes in reach generated by *other counties* gaining news-agency access. The identifying assumption is that shocks to a county’s appeal as a campaign stop are uncorrelated with changes in its media reach generated by news-agency access elsewhere. I find that a one-standard-deviation increase in media reach, roughly 44,000 additional reachable voters, increased the probability that a county hosted a speech by 2.3 percentage points, roughly 47 percent of the mean among counties without news-agency access. Because the comparison

¹The remaining threat is a time-varying county-pair shock correlated with both joint access and news coverage.

is across counties within an election, the estimates identify the relative allocation of stops and rather than whether candidates campaigned more overall. To address the concern that regional shocks simultaneously expanded news-agency access and attracted candidates, I show that later changes in media reach do not predict current visits and that the results persist when I construct reach using only distant counties. Finally, because speeches are identified from newspaper text, greater media reach could mechanically increase detection. Bryan’s 1896 campaign records suggest that undetected speeches are, if anything, more common in high-reach counties, so measurement error would attenuate the estimated effect.

The paper’s central contribution is to show that media networks can shape the geography of political communication itself. A large literature shows that mass media affect which voters are informed and, in turn, how politicians allocate policy and effort (Strömberg, 2004a,b; Snyder and Strömberg, 2010; Strömberg, 2015; Prat and Strömberg, 2013). In this literature, politicians respond to the audience created by the media, but the geographic origin of political communication is typically not itself a choice margin. Other work treats the production of political messages as a choice, along two margins. Political actors choose *when* to act, timing attacks and executive orders to moments when the news is looking elsewhere (Durante and Zhuravskaya, 2018; Djourelouva and Durante, 2022). Candidates also choose *what* to say, selecting partisan or moderate messages in anticipation of media coverage (García Jimeno and Yildirim, 2025) and adapting manifesto content to broader electorates (Le Pennec, 2024). I study another margin, *where* the message is produced.

The campaign-allocation literature explains location choices through electoral pivotality and the value of the local crowd (Brams and Davis, 1974; Bartels, 1985; Stromberg, 2008; Chen and Reeves, 2011). These frameworks emphasize the electoral value of the place itself rather than the outside audience reachable from it through media networks.² I show that news agencies raised the relative value of appearing in better-connected locations, creating a spatial complementarity between mass media and in-person campaigning.³

These findings speak to how exposure to campaign communication is measured (DellaVigna and Gentzkow, 2010). Studies of campaign appearances typically assign exposure to the place where an event occurred, and often find mixed or short-lived local effects (Shaw, 1999; Snyder and Yousaf, 2020; Buggle and Vlachos, 2023; Assouad, 2025; Pino and Salisbury, 2025). In my data, newspapers in connected counties reported speeches their readers never attended, so exposure extends beyond the county where the candidate appeared. Designs based on the location of an event may therefore classify some exposed voters as unexposed, which would attenuate estimated effects. That candidates moved toward high-reach places suggests that campaigns themselves believed these distant readers were worth pursuing.

I also contribute to the literature on communication innovations and the spatial diffusion of political information. Wang (2025) studies how telegraph access between 1840 and 1852 changed newspapers’ coverage of national politics and political participation, while related work shows how communication technologies reduced spatial information frictions (Steinwender, 2018; García-Jimeno et al., 2022; Jensen, 2007; Allen, 2014) and how newspaper markets shaped the supply of political

²Althaus et al. (2002) noted that appearances concentrate in large media markets, but cannot separate the value of the outside media audience from everything else that makes those places attractive to campaigns.

³This echoes the urban literature, where cheaper telecommunications raised the value of face-to-face interaction (Gaspar and Glaeser, 1998; Storper and Venables, 2004; Agrawal and Goldfarb, 2008).

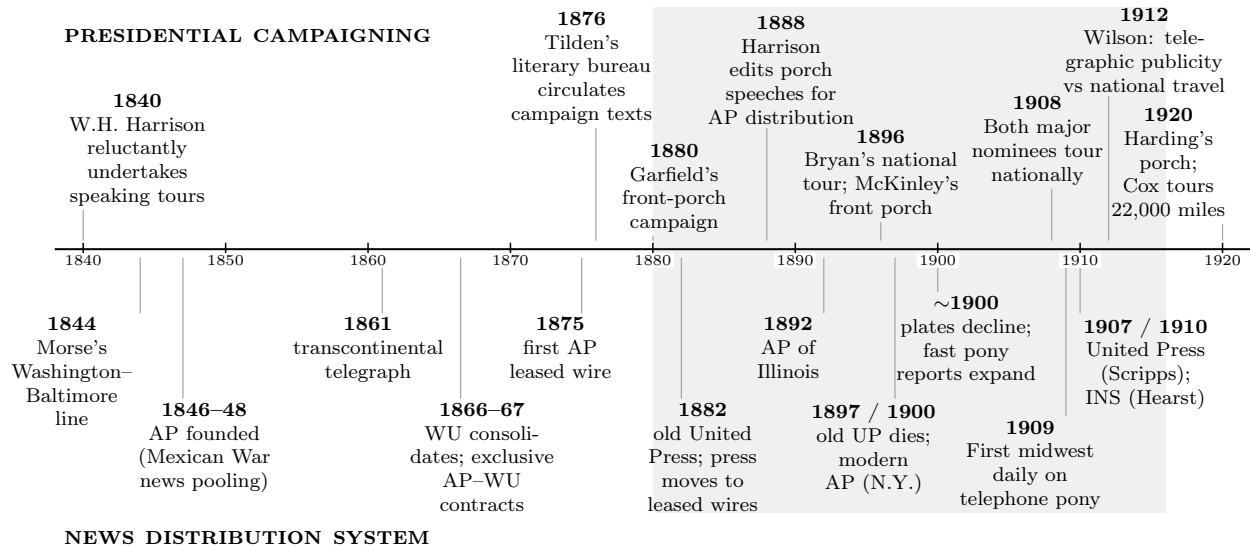
information (Gentzkow et al., 2011; Petrova, 2011; Gentzkow et al., 2015). I depart from this work in two ways. First, I study news agencies, an organizational layer built on top of the telegraph that determined which newspapers received national copy even after the wires were in place. Second, while much of this literature compares places that received a communication technology with places that did not, I show that the diffusion of political news between two places depended on access at both ends, which is what gives the *sending* location its value.

The paper proceeds as follows. Section 2 describes the historical setting of campaigns and the news distribution system. Section 3 describes the data and measurement. Section 4 presents the empirical strategy and Section 5 the results. Section 6 concludes.

2 Historical Setting: Campaigns and the News Distribution System

This section describes two parallel developments during the period studied. Presidential candidates became frequent public campaigners, while the news-distribution system expanded and allowed campaign appearances to be reported across the country. Figure 1 places both on a common timeline.

Figure 1: Presidential campaigning and the news distribution system, 1840–1920



Notes: The shaded band marks the period under study (1880–1916). The top track reports selected changes in the role of presidential candidates. The bottom track shows the organization of telegraphic news distribution. Sources: Ellis and Dedrick (1997); Hilpert (2015); Schwarzlose (1990); Rosewater (1930); Lee (1937); Hochfelder (2012).

2.1 Campaign Norms and Media Strategies

Nineteenth-century presidential campaigns could be intensely organized and participatory, but nominees generally avoided appearing to solicit votes themselves. Candidates were expected to have made their positions clear before being nominated and, after that, to maintain a posture of dignity and reserve, since openly soliciting votes was portrayed as evidence that a candidate lacked the character required for the presidency. Instead, parties, local campaign clubs and surrogate speakers organized

political events to promote their ticket, while candidates would work behind the scenes and increasingly communicated through public letters, acceptance statements, and other documents intended for publication. However, over time candidates faced competing expectations, as they should not appear eager for office, but voters increasingly demanded to know their positions and judge their character (Ellis and Dedrick, 1997).

One increasingly durable way to reconcile these demands was through mediated and stationary forms of campaigning. In 1876, Samuel Tilden established a “literary bureau” that circulated his speeches and campaign documents while he publicly maintained the customary posture of reserve. In 1880, James Garfield received organized delegations and addressed them from his Ohio home, inaugurating what became known as the “front-porch” campaign. Benjamin Harrison developed this strategy further in 1888. A campaign committee would schedule delegations to visit him and would review Harrison’s proposed introductory remarks. Harrison would then walk from his home to listen to their greetings, give his remarks and afterward send his edited responses to the Associated Press for publication the following day. The front-porch, rather than a retreat from campaigning, was a controlled campaign site that combined a live audience, organized local participation, and national newspaper publicity (Ellis and Dedrick, 1997).

The 1896 campaign greatly expanded two competing versions of presidential campaigning. William McKinley received organized delegations in Canton and delivered roughly 300 front-porch addresses. McKinley’s stationary campaign was still organized for national press reproduction, as press representatives reported almost every speech, and McKinley later thanked the Associated Press for “reporting and transmitting so fully the news from Canton” (Harpine, 2000; Associated Press, 1918). Meanwhile, William Jennings Bryan traveled approximately 18,000 miles and delivered about 600 speeches. Representatives of press associations and major newspapers also travelled with him or covered his campaign, turning his train into a moving news-production operation (Bryan, 1896, pp. 612–614). Touring then became increasingly normal for challengers, although incumbent presidents remained more constrained. In 1900, for example, incumbent McKinley remained largely above the campaign while his vice-presidential nominee, Theodore Roosevelt, conducted an extensive railroad tour. Roosevelt initially preferred one or two major appearances per state and expected newspapers to transmit his message to the remaining population, but state party committees demanded many local appearances. In the end, Roosevelt visited 480 communities. Correspondents traveling with the campaign train distributed reports by telegraph, which local newspapers frequently reproduced (Hilpert, 2015).

The evolving campaign strategy of traveling while sending messages through the press became a common practice and meant that appearances addressed more than the people physically present. Bryan himself explained in 1900 that he prepared important speeches for the press because the constituency reached through newspapers was larger than any assembled crowd. Nonetheless, the possibility of substitution between personal appearances and mass communication was still very clear. In 1912 Woodrow Wilson initially argued that the telegraph could distribute his message more effectively than crossing the country by train, but pressure to let voters see and judge him still drew him onto the campaign trail (Ellis and Dedrick, 1997). By 1920 James Cox travelled approximately 22,000 miles even as Warren Harding retained a heavily publicized front-porch strategy.

These cases show that personal appearances and mass communication were neither independent nor perfect substitutes. A candidate could use newspapers instead of traveling to communicate with

voters, but an appearance also generated newsworthy content, displayed the candidate’s personality, activated local party organizations, and changed the newspaper set that could reproduce its speech. Whichever force dominated, both depended on the number of voters a report of the appearance could be expected to reach. In turn, the number of reachable voters responded not only to the presence of telegraph wires, but also to whether newspapers could obtain rapid access to a shared news report, so that the message could be disseminated widely.

2.2 News Agencies and Access to Wire-Speed News

The first telegraph line in the United States connected Washington, D.C., and Baltimore in 1844 and was quickly adopted.⁴ By 1852 the network had reached “all major and most minor cities” in the country (Wang, 2025), in 1861 the transcontinental line was completed, and in 1866 Western Union absorbed its last serious rival to become a national monopoly operating some 2,250 offices (Schwarzlose, 1990). The telegraph made same-day national news possible for the first time and displaced the methods that had carried news before it, e.g. mail exchanges between newspapers, harbor boats meeting incoming steamers, and pony expresses relaying copy overland by horse.

The telegraph network continued to grow substantially after 1880, but by the beginning of the sample period it already reached the principal newspaper markets in the settled East, Midwest, and South (see panel (a) of Figure 2). Much subsequent growth represented densification of this footprint and expansion serving business and railroad traffic, although physical wire arrival remained a live margin in newly settled Western counties into the 1880s. Press messages accounted for only about eight per cent of Western Union’s revenue in 1887 against eighty-seven per cent for business messages, and between 1900 and 1910 growth in the telegraph industry broadly tracked population growth (Hochfelder, 2012). Western Union’s revenue per message was also broadly flat from the late 1870s onward (panel (a) of Figure A1). Physical telegraph access, therefore, was no longer the main constraint in the news distribution system.

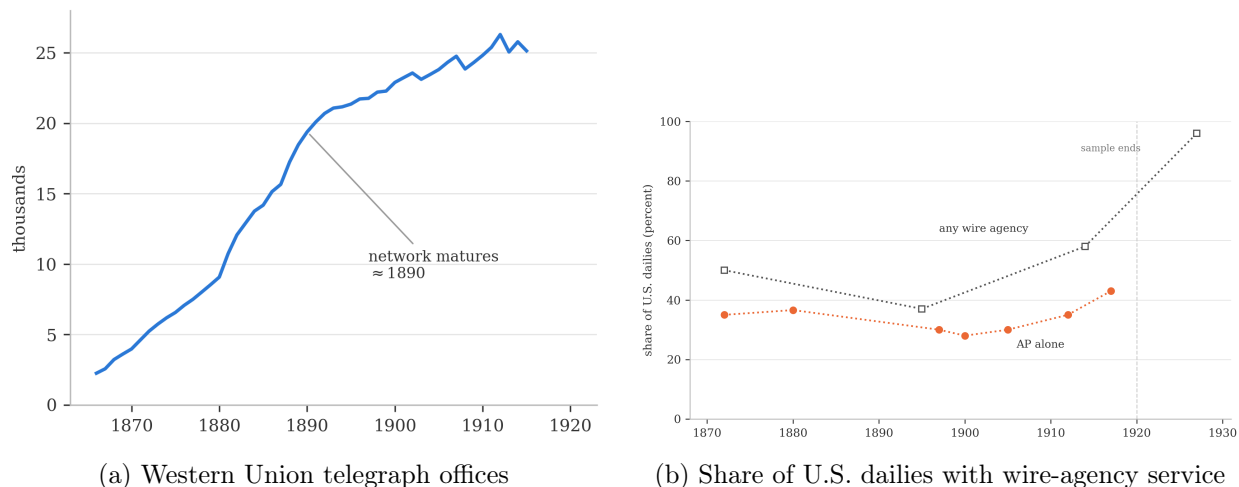
The main constraint to the distribution of national news was having to collect news across a wide geographic area or coordinate with other newspapers to share and send information. News agencies solved this coordination problem. They collected news centrally—through their own correspondents, exchange among their subscribers, or both—and distributed a single report simultaneously to many newspapers over a shared circuit. In this way, news agencies benefited from economies of scale from distributing the fixed costs of collecting and transmitting news across many newspapers, while facing near zero costs of adding another newspaper to an existing circuit.⁵

A first version of the AP formed when New York papers pooled the cost of covering the Mexican–American War (1846–48), and the arrangement was consolidated into a permanent association in 1856 (Rosewater, 1930). In 1867 the New York and Western Associated Press agreed with Western Union not to use other wires and in return Western Union agreed “not to sell news or serve other associations in said territory” (Rosewater, 1930). Each side thereby protected the other’s monopoly, and the practical effect was to close the telegraph to rival news organizations. At the beginning of

⁴The line was built by Samuel Morse and Alfred Vail, who had been experimenting with telegraphy since the 1830s. Around the world, telegraphy had been developing since the 1790s, but Morse’s system would prevail and be adopted internationally because it was far cheaper to build over long distances.

⁵Shmanske (1986) argued that news was a public good, although agency reports are more precisely understood as club goods, since they were non-rivalrous among subscribers but excludable because agencies could restrict access.

Figure 2: Telegraph infrastructure and access to news-agency service



Notes: Panel (a) reports Western Union telegraph offices (U.S. Bureau of the Census, 1960, Series R 43). The network continued to expand rapidly through approximately 1890 and grew more slowly thereafter. Panel (b) reports contemporary benchmarks of the share of U.S. dailies receiving wire-agency news service. “AP alone” denotes dailies served by the AP system (1872 and 1880) or AP members over total dailies (1897–1917); “any wire agency” adds the other agencies (American Press Association in 1872, UP from 1907, and INS from 1910). The 1872 and 1914 points are approximations, and 1914 is an upper bound because it ignores double subscriptions. Sources for panel (b): Lee (1937, pp. 509–510); Rosewater (1930, p. 157); Schwarzlose (1990, pp. 173, 178, 231); Silberstein-Loeb (2014, pp. 69, 80).

the sample, carrier access and agency membership were therefore still partly fused. However, the rise of rival telegraph companies and the shift of press associations onto leased wires around 1882, helps separate the physical infrastructure from the news agency organizations (Hochfelder, 2012).

Newspapers could buy individual telegraph service, but access to a news agency’s report was much more restricted. Panel (b) of Figure 2 shows that in 1880 only 355 of the country’s 971 dailies received the AP’s shared report (Rosewater, 1930). Together with the contractual restrictions described above, this gap indicates that access to a shared agency report, rather than access to the telegraph, was the binding constraint for access to news.⁶

During the sample period, news agencies expanded both the number of newspapers served and their own distribution layer (panels (b) and (c) of Figure A1), and offered services that differed sharply in speed and cost. The AP leased its first dedicated wire in 1875, and by 1897 its leased network ran to 26,798 miles (Schwarzlose, 1990; Lee, 1937). A leased circuit delivered the full report to every town on it simultaneously, and adding another recipient to an established circuit was inexpensive (Silberstein-Loeb, 2014). Newspapers below the full-report margin could obtain

⁶As discussed below, part of the unmet demand for access to news agencies led to the creation of other news organizations. Further suggestive evidence of this is that AP memberships were capitalized as assets: franchise bonuses ranged from \$100,000–150,000 for the New York *Sun* to \$1,000 for Devil’s Lake, North Dakota, and publishers sometimes bought entire newspapers to acquire their franchises (Rosewater, 1930). Congressional investigations initially blamed the telegraph monopoly, but by 1879 the Senate identified restricted access to the AP report as the operative constraint on newspaper entry and survival (Blondheim, 2004). On top of this, billing records provide complementary evidence that individual telegraph carriage and access to the shared agency report were distinct margins. Western Union billed the press approximately \$1.084 million in 1880, where the AP’s report accounted for \$392,800, while individual newspapers paid \$690,855 to file special dispatches of their own (and the largest single newspaper accounted for \$70,006 of the latter expenditure) (North, 1884). While newspapers could still purchase transmissions outside the AP, they still had high demand for the shared agency report.

cheaper products, including two that differed fundamentally in speed. A small daily without rapid agency service could buy news already set in type on curved stereotype plates—“boilerplate”—which the American Press Association assembled from AP copy and shipped by express (Lee, 1937). Alternatively, it could subscribe to a *pony report*, which was a condensed abstract of the full report relayed from an interior hub to smaller towns by telegraph (North, 1884, p. 108). Because it travelled by wire rather than by rail, the pony report arrived the same day.⁷

Continuous technological advances and increased demand for faster news delivery led to the appearance of cheaper news distribution alternatives and entry of new players. Around 1900, growing demand for telegraphic news during the Spanish–American War, the spread of linotypes in smaller newspaper offices, and the AP’s preference for its own pony reports jointly contributed to the decline of the express-shipped plate service (Lee, 1937). In its place, agencies increasingly sent condensed bulletins to small papers by wire or telephone. The telephone made rapid agency service feasible for smaller-town papers since it did not require a leased wire into the newsroom nor someone at the receiving end able to take Morse (Lee, 1937). Rival organizations widened access as well. Partly as a response to publishers’ exclusion from the AP, the new United Press and the International News Service were founded in 1907 and 1910, respectively, and sold their reports commercially to any newspaper able to pay (Le, 2026). Panel (b) of Figure A1 shows their membership growth.

The relevant transformation in the news distribution system during this period was therefore no longer whether a county possessed a telegraph office, but whether its newspapers had access to institutions capable of collecting and redistributing news at wire speed. Focusing on news agencies during this period also distinguishes this paper from previous studies. Wang (2025) focuses on the effect of telegraph arrival on turnout and news content between 1840 and 1852, and his diffusion mechanism is how easily small newspapers could copy and resend articles from larger newspapers with telegraph access. This paper focuses on a later period and on the role of news agencies, which operated on top of the telegraph infrastructure, and their role centralizing and redistributing news to newspapers that were already connected to the telegraph.

Just as candidates were becoming more important producers of campaign news, news agencies were broadening the network through which their appearances could be reproduced. While these parallel changes do not imply a causal relationship, contemporary accounts show that candidates understood that newspapers could enlarge the audience for an appearance. Besides the voters reached locally, campaign stops now created an additional audience in the newspaper readers reached elsewhere. Better news distribution could substitute for travel by allowing candidates to communicate without appearing in person. However, conditional on making an appearance, it could also make some locations more valuable than others if they were better connected to the news network and had a larger outside audience attached to them.

3 Data and Measurement

The empirical design requires three objects that are not readily available from existing historical sources: which counties had access to a news agency, where candidates delivered campaign speeches,

⁷Note that the word “pony” can correspond to two different things. A pony *express* was a literal horse relay before the telegraph era. A pony *report* was a cheaper, condensed wire product that the AP sent to small-town newspapers.

and which counties’ newspapers covered a speech given elsewhere. This section constructs each of them from a single corpus of historical newspaper text, described in Section 3.1. Section 3.2 infers news-agency access from the speed with which newspapers reprinted non-political articles elsewhere. Section 3.3 builds the speech and coverage measures by labeling articles that mention presidential candidates with an LLM. Section 3.4 describes the remaining inputs.

3.1 The Newspaper Corpus

Measures of news agency access, campaign stops and their coverage are derived from the Library of Congress’s Chronicling America collection of scanned historical newspapers. American Stories from Dell et al. (2023) takes over 20 million of these scanned pages, makes their text searchable, and organizes each page’s text into separate articles, yielding 438 million articles between 1774 and 1963. I use two intermediate datasets derived from that corpus. The first is the Newswire dataset (Silcock et al., 2024), which detects clusters of articles that several newspapers reprinted. I use reprinting of non-political articles to infer news-agency access. The second is a subset of political articles mentioning a presidential candidate during the campaign season, which I use to detect campaign stops, their coverage, and their content.

I focus on the ten presidential campaigns from 1880 to 1916, the window over which the two intermediate datasets are simultaneously available. The Chronicling America data used to detect campaign activities is limited beyond 1923 due to copyright restrictions. The reprint data begin in 1878 and runs through 1919 before the corpus switches to a different underlying collection. Therefore, I limit my analysis until 1916, the last campaign year for which both datasets are available. Focusing on the period since 1880 is also ideal since, as Section 2.2 argues, it corresponds to the window in which changes in the news distribution system corresponded predominantly with the operation of news agencies, rather than infrastructural changes in the telegraph.

The corpus’s main limitation is that it is a sample of the historical press. It includes digitized newspapers for approximately 22 percent of county-years for which there are records of daily newspapers. Even though we can detect speeches in a county without digitized newspapers from articles originated in other places, a county whose newspapers were never scanned cannot be observed receiving a fast reprint nor covering a speech given elsewhere. The subsections below discuss the implications and why this may lead to a measurement error that attenuates the main estimates.

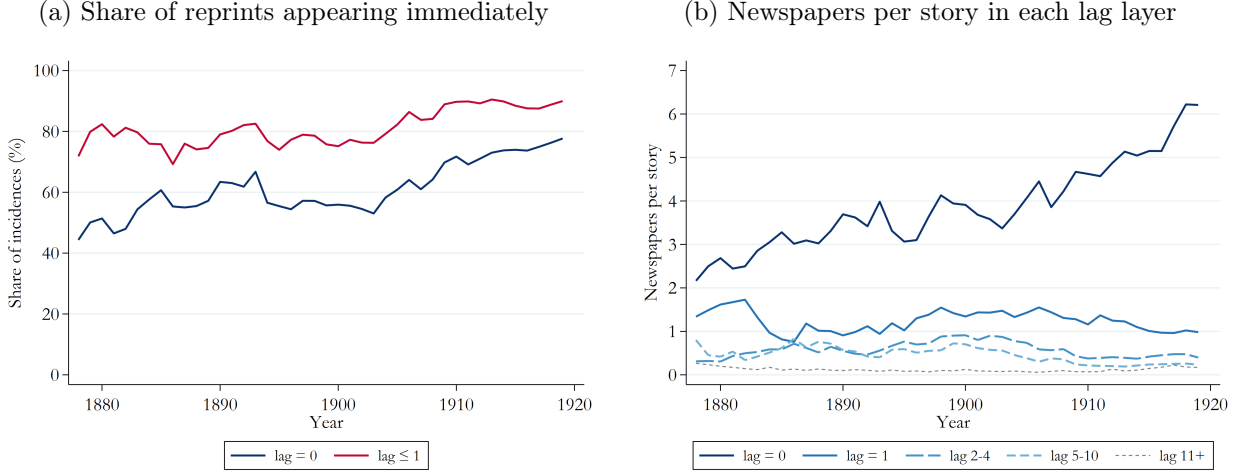
3.2 News Wires and Agencies

Given the lack of direct data on agency access, access to news agencies is inferred from the speed with which newspapers reprinted articles that also appeared on other newspapers. The data from Silcock et al. (2024) show, for each article, the date on which each newspaper printed it.

I use the time it took a newspaper to reprint an article since it first appeared in the corpus to define connection status. A county is coded as connected in year t if, at any point up to and including $t - 1$, at least one non-political article is reprinted by a newspaper headquartered in that county, within one day of that article’s first appearance anywhere in the corpus.⁸ Connection is

⁸Appendix B.3 recodes connection to require at least two, five, or ten fast reprints per newspaper-year.

Figure 3: More newspapers reprinting faster, 1878–1919



Notes: The Newswire data group articles into *clusters*: sets of articles that are reprints of the same underlying story (Silcock et al., 2024). For each article-newspaper pair (incidence), the reprint lag is the number of days between the first printing of the cluster anywhere in the corpus and the printing of that story by that newspaper. Panel (a) plots the annual share of incidences reprinted on the same day as the cluster’s first printing (lag = 0) and within one day of it (lag ≤ 1). Panel (b) decomposes the number of newspapers by their reprint lag. For each cluster, it counts the newspapers printing at each lag, and divides by the number of clusters that year.

therefore an absorbing state (i.e., a county ever observed receiving news at wire speed is treated as connected from that year onward). Only non-political articles are used to define the connection status to avoid a mechanical correlation with the political coverage outcome of interest.

I choose a threshold of one day since several news agency products had evolved over this period to arrive over a wire the same day (e.g. leased-wire, pony reports), while other methods had to be physically shipped and took days, as described in Section 2.2. In line with this, Figure 3 shows that the time between an article’s first publication and its reprinting by other newspapers fell (Panel (a)). Between 1878 and 1919, reprinting on the same day or within one day of an article’s first publication rose from 44 to 78 percent, and from 72 to 90 percent, respectively. Panel (b) shows that this was not only a change in speed, but also in how many newspapers ran a story at once. The number of newspapers reprinting an article on the day of first publication rose from 2.2 per story to 6.2, while the number of newspapers reprinting at lower speeds stagnated or declined.⁹

The fast reprints detected in the data were also not local. The median distance between a fast reprint and its dateline was 824 miles, and, by 1919, the typical fast-reprinting article was reprinted by eight newspapers from seven states.¹⁰ Overall, more newspapers were running the same story on the same day, even far away from each other. These patterns are what a shared report distributed by a news agency would produce, and not what copy passed from one newspaper to the next would.

Because connection is inferred from digitized pages, a county whose press was never scanned is coded as unconnected whether or not it was. Using available records of news agency access for a subset of newspapers, among counties that had a newspaper with membership in the Associated

⁹Figure B1 shows that the mass of reprinting concentrates on the first day an article appears in the corpus and thins sharply beyond one day (Panel (a)). These patterns survive when the counts are scaled by corpus size rather than by detected reprints, so that they are not an artifact of what is detected as a reprint (Panel (b)).

¹⁰See Appendix B.2 for more details.

Press, United Press, or International News Service, my inferred measure defines 73% as connected when the county has digitized articles that year, but only 19% when it does not.¹¹ Digitization predicts the coding about as strongly for non-members as for members, so the difference reflects which counties can be observed rather than which ones had access.

Nonetheless, this may be less damaging than it first appears. Because connection is an absorbing state, a false zero requires a county never to be observed receiving a fast reprint in any year up to t , not merely to be missed in one. Also, the measurement error occurs in only one direction, as some truly connected counties are measured as unconnected, but never the reverse. This implies that the differences between connected and unconnected counties should be attenuated rather than inflated.

3.3 Speeches: Detection and Content Classification

Studying presidential campaign speeches in the 19th and early 20th centuries is challenging due to the lack of comprehensive public records. Until recently, historians relied mainly on two alternatives to build a candidate’s tour stops. One alternative was to rely on the candidate’s biographies (e.g. [Bryan, 1896](#)) and systematically organize the information. The other option was to read through newspapers and try to recreate the campaign stops and routes, which is a time-consuming task.¹² Modern digitization and computational tools make these approaches feasible at scale.

Among the corpus’s 438 million articles, I search for mentions of presidential candidates during the typical campaigning period (July 1 to November 15) in election years. Figure 4 shows the nearly 330,000 articles mentioning at least one presidential candidate since 1800 until 1960.

I then use an LLM to label which articles contain public speeches and identify event details. In August 2026, I systematically labeled each article one at a time with GPT-5.6 terra, one of the leading models at the time, using OpenAI’s API (Application Programming Interface). I developed the prompt based on OpenAI’s official documentation and suggestions from [Korinek \(2025\)](#). I iterated over the instruction set until it correctly classified the speech status of a random sample of 100 articles from 1896 that I had previously labeled manually. The final prompt asked the LLM to think as a historian and search for public speeches among the presidential candidate’s articles. When a speech was detected, it classified its location(s), date(s), and the sentiment of the editorial coverage. Each town/State mentioned in the article is assigned to its corresponding county using data from [Berkes et al. \(2023\)](#). Additional information are described in detail in Appendix C.

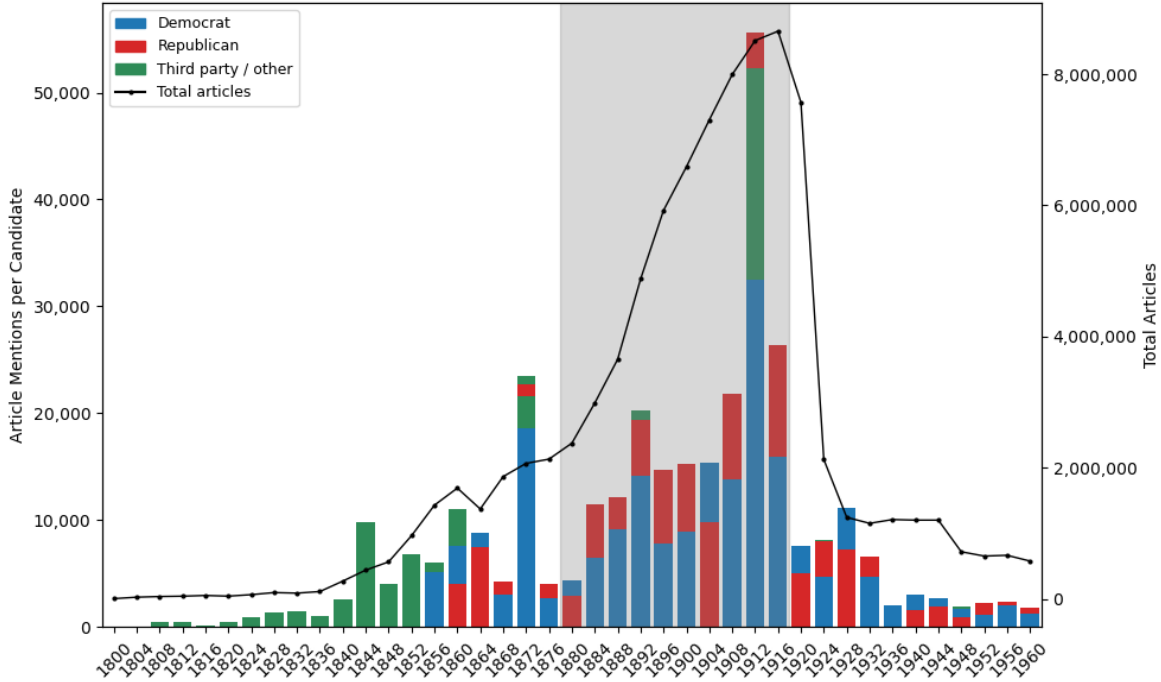
I validate the LLM’s labeling quality in two ways. First, I validate my speech location labels against Bryan’s 1896 campaign, the only available record during this period ([Sanderson, 2008](#); [Thomas et al., 2008](#)). Among the 2,828 counties in the 1896 cross-section, the LLM recovers 325 of the 380 counties with an officially recorded Bryan visit and additionally assigns a stop to 44 counties without one, implying a recall of 0.86 and a precision of 0.88 (see Table C1).¹³ These rates are high for historical work, especially given that they come after a noisy process of scanning and digitizing

¹¹A county-year without a digitized title can be coded connected because connection is an absorbing state.

¹²Two notable examples that combine both approaches are [Assouad \(2025\)](#), who reconstructs the stops of Atatürk’s propaganda visits across Turkey between 1923 and 1938 from his private diary, contemporary newspapers, and archival records. The other is [Sanderson \(2008\)](#) and [Thomas et al. \(2008\)](#), who use William J. Bryan’s own 1896-campaign memoir and articles from the *Omaha World-Herald* to build a dataset with campaign stops and the speeches delivered.

¹³The LLM also correctly classifies 92% of the long speech extracts recorded by historians (Appendix C.3.4).

Figure 4: Newspaper Articles and Candidate Mentions in American Stories



Notes: Bars show the number of articles mentioning each presidential candidate in each election year, with colors denoting the candidate's party: blue for Democrats, red for Republicans, and green for third-party and other candidates (left axis). The black line shows the total number of articles in the “American Stories” corpus in each election year (right axis). The shaded area marks the years inside the 1880–1916 study period.

old newspapers and then organizing at the article level before labeling.¹⁴

Importantly, I show that at least some of the misclassifications are due to campaign stops missed by the historical records that the LLM helps recover. I find that at least 19 out of 44 counties initially flagged as false positives are genuine campaign stops.¹⁵ Furthermore, while some of these detection errors are not fully random, their structure is likely either absorbed by the design or works against the results. Figure C2 shows that false positives rise with the volume of digitized text, which is mostly constant within counties and absorbed by county fixed effects. False negatives are more common in counties without a connection and, if anything, in counties with *higher* media reach. This means that missed speeches tend to attenuate the estimated location response rather than generate it. Appendix C.3 provides additional details and tests.

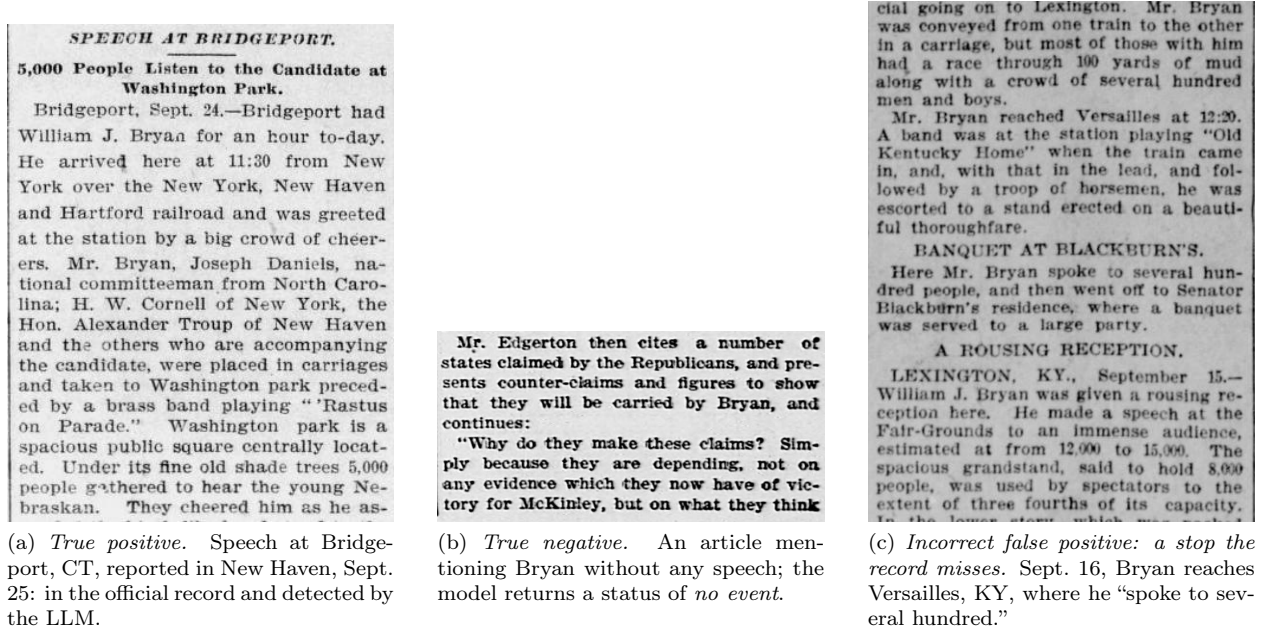
Second, I also compare the number of counties visited by each candidate in my data and the aggregate data from Pino and Salisbury (2025, Table 1, p. 57). These authors follow a similar strategy and build their panel of campaign stops by searching for candidate mentions between 1872–1908 in an online newspaper repository. Our measures have a high correlation of 0.98, though my method tends to classify fewer counties as visited (Figure C3).

Overall, between the ten campaigns from 1880 to 1916, the LLM classifies 8,705 (7.4%) of the

¹⁴Silcock et al. (2024) report similar rates for their classifiers on the same newspaper collection (see Appendix C.3).

¹⁵Correcting the benchmark for the documented misses raises precision to 0.93. This is likely because an LLM can read thousands of articles without any “labeler-fatigue”, and thus detect speeches that the historical record misses. This is in line with the findings of Gilardi et al. (2023) that LLMs could outperform human labelers in some tasks.

Figure 5: Examples of Agreement and Disagreement with the Official Record



Notes: Each panel is the scan behind one county-level verdict against the official record of Bryan's 1896 stops. Panels (a) and (b) are correct verdicts. Panel (c) counts as a false positive because the historians' reconstruction missed it, thus showing the usefulness of this method. Figure C4 carries six further scans.

118,147 filtered articles as reporting a speech delivered by the candidate. Figure 5 gives an insight of this process and shows the newspaper page behind the classification decision for three different counties and compares each against the official record of Bryan's 1896 stops.

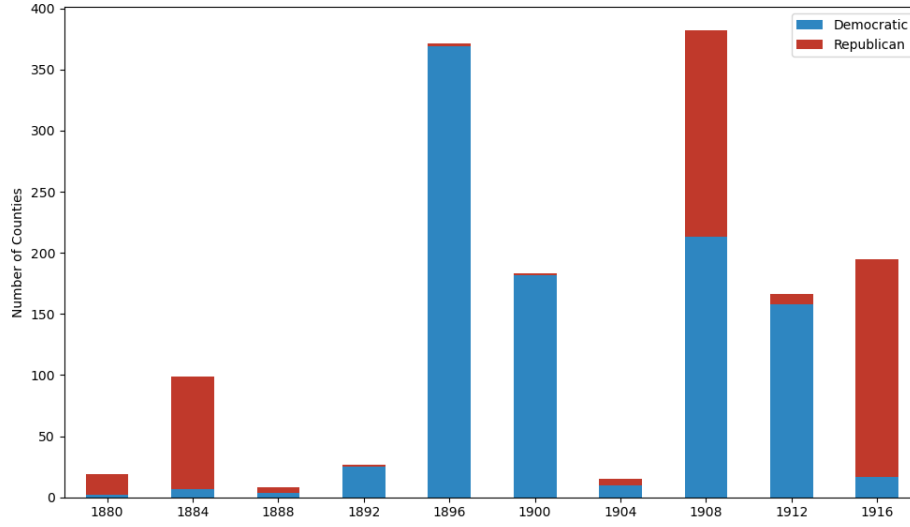
Besides knowing where speeches occurred, I can identify where these events were covered using information on each newspaper's headquarters from [Library of Congress \(2026\)](#). This allows me to build directed pairs of where a speech occurred (the origin of the message) and where a newspaper covered it (the destination of the message).¹⁶

Figure 6 shows the number of counties visited by each party's presidential candidate in every election year. It largely reproduces historical accounts of each campaign's strategy. Before 1892 most candidates visited few counties, and Bryan's 1896 whistle-stop campaign marked a sharp break, with roughly 370 counties visited. However, campaigns would not become the norm yet, especially for incumbents, so 1904 had very few campaign speeches as Roosevelt ran for president while in office. In general, parties didn't have split opinions on campaigning, and would generally support or reject it depending on the previous' candidates results ([Ellis and Dedrick, 1997](#)).

Figure 7 illustrates the geography of speeches and their newspaper coverage across all campaigns, while Figure C6 in the Appendix focuses on Bryan's 1896 campaign. Dots mark the counties where the LLM labels detected a speech and concentrate in the Midwest and the Northeast. Amber marks the counties whose newspapers covered a speech. Coverage reached over 500 counties, many of them on the frontier at the time, and even reached 340 counties that no candidate visited.

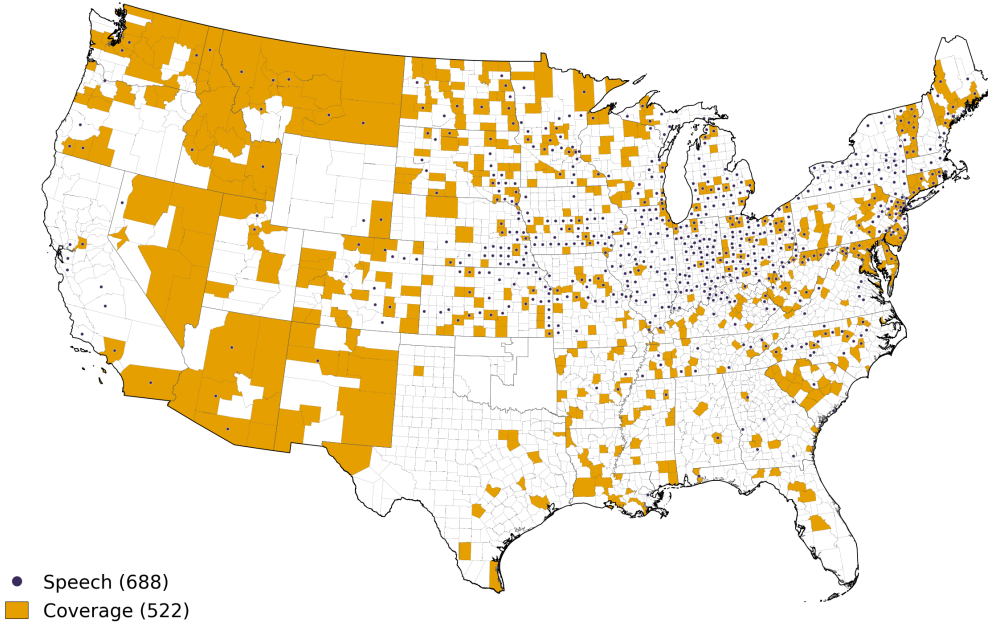
¹⁶The headquarters county is likely the relevant unit of coverage, as even today the median newspaper sells more than 80 percent of its copies in its headquarters county ([Gentzkow and Shapiro, 2010](#); [Gentzkow et al., 2011](#)).

Figure 6: Counties Visited per Party and Election Year



Notes: Bars show the number of counties in which the LLM labels detect at least one public speech by each party's presidential candidate in that election year.

Figure 7: Speeches and Coverage by County, All Candidates 1880–1916



Notes: This figure shows the counties where the LLM detects at least one public speech by a major-party presidential candidate in any election year between 1880 and 1916 (dots) and the counties where newspapers circulated at least one article covering one of these speeches (amber). The two overlap in 182 counties. A dot on an amber county marks a speech and some coverage over the period, not necessarily in the same year. The map uses the 1890 county boundaries, the frame used throughout the analysis. Counties created after 1890 (16 with speeches and 60 with coverage) are not included in this map.

Beyond detecting that a speech occurred, I also detect the content of the speeches from newspaper articles. At the time, given that there was no radio nor way for people to fully recollect

what candidates said, it was common for journalists to copy the exact words at of political speeches. A second labeling pass extracts content from speech articles and classifies that content along different schemes: the topic addressed (21 categories), the audience the appeal is directed at (four categories), the geographic jurisdiction it concerns (four categories, from the specific locality to federal or national policy), and the values reflected in the content of the speech. I also label the statements of each party’s presidential platform 1880–1916 (Woolley and Peters, 2026) to compare alignment between individual speeches and party platforms. Appendix C.5 documents the content prompt, the full category vocabularies, and their validation.

3.4 Other Data Sources

Newspaper industry. Data on daily newspaper counts come from Gentzkow et al. (2011).

Electoral data. Presidential turnout and party vote shares at the county level come from Gentzkow et al. (2011). I complement this with state Electoral College apportionment for 1872–1920, assembled from the National Archives, the American Presidency Project (Woolley and Peters, 2026), and Dave Leip’s Atlas of U.S. Presidential Elections (Leip, 2026).

Census and IPUMS. County demographic and economic characteristics come from the decennial censuses. I obtain county-level aggregates for 1870–1920 from Haines and ICPSR (2010) and IPUMS USA (Ruggles et al., 2025). From these data I linearly interpolate between census years and construct the county-level controls used throughout the analysis: log population, the share of males aged 21 and older, log manufacturing output per capita, and urbanization shares. The Haines data additionally provide the detailed agricultural and manufacturing series from which I build the county economic-interest vectors used in the speech-alignment analysis.

Geographic and transportation data. Data on historical railroads and market access come from Atack (2016) and Hornbeck and Rotemberg (2024). I also use Berkes et al. (2023) to geolocate the historical political events by their town and state names.

4 Empirical Strategy

In this section, I first show how to estimate the effect of joint news-agency access of a sender-receiver pair on the flow of campaign news between them. Then, I explain how I use these estimates to build a measure of the additional outside audience that news-agency access made reachable from a potential speech location. Finally, I use this measure to study how candidates responded to changes in the reachable audience due to expansions elsewhere in the news-agency network.

4.1 News Agencies and Coverage of Political Events

Until the development of the telegraph, the exchange of information resembled the trade of physical goods as it needed to be physically delivered, meaning it could travel as fast as the most rapid

means of transportation could carry it. Therefore, distance had also been one of the main barriers to the flow of information. Once the telegraph was adopted, it allowed for near-instantaneous transmission of information. However, a local newspaper that wanted to cover national news still had to engage in costly coordination with other newspapers across the country to share news. News agencies solved this coordination problem by centralizing the recollection and distribution of news across their subscribing newspapers. This subsection studies how the removal of one of the supply barriers to the flow of information affected the coverage of distant events of national interest, such as presidential candidate speeches.

In line with the literature on the trade of goods, I model the flow of political news between a sending county (i.e. where the event takes place) and a receiving county (i.e. where a newspaper covers it) as a gravity equation, where the flow of news is a function of the “trading costs” between the two counties. Equation 1 specifies the dyadic regression that captures whether a presidential candidate’s speech given in an *origin* county was more likely to be covered in a *destination* county once *both* counties had gained access to a news agency.¹⁷ Standard errors are clustered at the unordered county pair level.

$$\begin{aligned} \text{Coverage}_{o,d,t} = & \psi_1 \text{Connect}_{o,t} \cdot \text{Connect}_{d,t} + \psi_2 \text{Connect}_{o,t} \cdot \text{Connect}_{d,t} \cdot \log(\text{Distance})_{o,d} \\ & + \gamma_{o,t} + \delta_{d,t} + \phi_{o,d} + \epsilon_{o,d,t} \end{aligned} \quad (1)$$

The outcome variable $\text{Coverage}_{o,d,t}$ is an indicator equal to one if a speech given in county o was covered by at least one newspaper headquartered in county d in campaign year t , and zero otherwise. Because a county whose newspapers connect to a news agency is different than other counties, e.g. they might be growing and have higher demand for news and political events, I include origin-year ($\gamma_{o,t}$) and destination-year ($\delta_{d,t}$) fixed effects. Because some counties might also be more similar and have persistently higher demand for news from each other, e.g. counties in the same state or with similar demographics, I also include origin–destination pair fixed effects ($\phi_{o,d}$). The term ψ_1 then captures the effect of sending news from an origin to a destination county, once they have both gained access to a news agency. The term ψ_2 captures how this effect varies with the linear distance between the two counties, since the reduction in the costs of transmitting news should matter more for long-distance coverage.

Identification of the effect of news-agency access on coverage of political events thus comes from within-pair variation over time, after accounting for all time-varying shocks at the speech and newspaper locations. A causal interpretation requires that the timing of when both counties have connected is not driven by time-varying pair-specific shocks that also affect coverage between the two counties.¹⁸ Given that subscribing to a news agency was costly, as explained in Section 2.2, it seems unlikely that counties would have connected to the network to exchange political news exclusively from one county, perhaps with the exception of state capitals and a handful of cities. However, this could still be violated if a shock specific to a pair increased both their propensity to exchange political news and make each county more likely to connect to the news agency, such as the

¹⁷I estimate this equation among the set of origin counties with at least one campaign stop and destination counties with at least one newspaper in the sample, which are the counties that could have potentially “sent” and “received” news about a speech. I drop county pairs where the origin and destination are the same county, since a county’s own newspapers would cover its own events regardless of news-agency access.

¹⁸Because the connection status of a county is an absorbing state, the interaction of the connection variable of both counties turns on once the second county connects.

construction of a railroad between the two counties. To account for these threats, I build alternative specifications allowing different pair types to have their own time-trends, excluding capital and big-city counties, and excluding pairs that never connect. I also present diagnostics with an event-study and placebo tests for whether leads of the connection status predict coverage. However, these are limited since coverage of presidential campaigns is observed every four years and there are only 10 campaign years in the data.¹⁹

4.2 From News Coverage to Media Reach

A campaign appearance can reach voters in two ways. The candidate addresses a local audience directly, while newspapers can report the same appearance to readers elsewhere. Let $V_{j,t}$ denote the number of eligible voters in county j . For a speech delivered in county o , $p_{o,d,t}V_{d,t}$ is the expected outside audience reached in destination county d , where $p_{o,d,t}$ is the probability that newspapers there report the appearance. The potential audience attached to the event can then be written as

$$A_{o,t} = V_{o,t} + \underbrace{\sum_{d \neq o} p_{o,d,t} V_{d,t}}_{\text{outside media audience}} \quad .$$

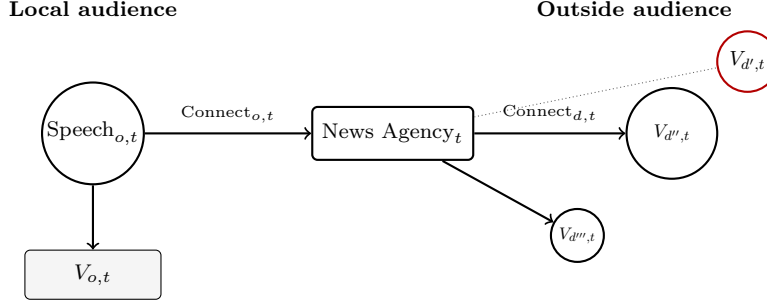
This accounting separates the local voter population from the expected outside audience. By increasing coverage probabilities across destinations, news-agency development can raise the potential audience attached to speaking in county o even if its local voter population is unchanged. Figure 8 illustrates this idea. On the left side, before a location has news-agency access, the only audience is the local electorate. However, as soon it gets connected to a news agency, a speech there can now reach many potential counties d' that are at different distances and have different voter populations, as long as those counties also have news-agency access.

To build a measure of the potential audience, I take the estimates from equation 1 and predict the additional probability that a speech in county o would reach county d during year t due to changes in news-agency access. One should not include the fixed effects in the prediction because they would add some of the local factors from the potential speech location o that correlate with political unobservables.²⁰ I then multiply this probability by the number of eligible voters (i.e. men aged 21 and older) in the destination county to obtain the expected number of *additional* voters who could receive news coverage. Summing this term across all other counties gives the total number of additional outside voters who could receive news of the political event, shown in equation 2.

¹⁹While connection is observed annually, presidential campaigns are observed every four years, which constrains the usual event-study diagnostics that rely on sharp timing of treatment. With only 10 campaign years in the data, the time window available for event studies is also limited. Placebo tests of leads of the connection status might be informative, but a one-year lead for one campaign is a three-year lag for the next campaign's treatment status. Furthermore, because connection status is inferred from the first *observed* fast reprint, any delay between the start of service and the first reprint we detect leads us to assign connection later than it actually occurred.

²⁰Hence why Media Reach is the *additional* outside audience that can be reached due to news-agency access, rather than the total outside audience.

Figure 8: Local and Media Audiences of a Campaign Appearance



Notes: $\text{Speech}_{o,t}$ is a campaign appearance in county o , $V_{o,t}$ is its local voter population, and $V_{d,t}$ is the voter population in any of the outside destination (e.g. d' , d'' , d'''). Solid arrows show active paths to voters: $\text{Connect}_{o,t}$ is the speech county's own news-agency connection, and $\text{Connect}_{d,t}$ denotes the corresponding connection for a destination county. Destination-circle sizes represent voter populations, and line lengths represent distance from the speech county. The dotted segment marks an unconnected destination, drawn in red, which therefore does not contribute to Media Reach. Media Reach excludes local audience $V_{o,t}$ and sums the predicted additional coverage probability attributable to news-agency access, multiplied by $V_{d,t}$, across outside destinations.

$$\begin{aligned}
 \text{MediaReach}_{o,t} &= \sum_{d \neq o} \widehat{\text{Coverage}}_{o,d,t} \cdot \text{EligibleVoters}_{d,t} \\
 &= \text{Connect}_{o,t} \sum_{d \neq o} \text{Connect}_{d,t} \cdot \text{EligibleVoters}_{d,t} \cdot (\hat{\psi}_1 + \hat{\psi}_2 \log(\text{Distance})_{o,d}) \\
 &= \text{Connect}_{o,t} \cdot \text{OutsideAudience}_{o,t}
 \end{aligned} \tag{2}$$

Media Reach is the product of a county's own connection status and its potential outside audience. The first term in equation 2 is an extensive margin—whether a local event can be “sent out”—and, correlates with local conditions, such as population growth, that also attract campaign activity. The second term is the additional voters a speech could reach in *other* counties, and varies with the timing of when those counties connected and the size of their electorates. In the main analysis, I study how media reach impacts campaign strategy controlling for each individual term.

Because Media Reach sums predicted coverage over every possible destination elsewhere, it varies with changes anywhere in the news network, even when nothing changes in county o itself. This aggregation follows the logic of “Market Access” in the trade literature, which summarize both the direct effect of a location's own link and the indirect effects of changes elsewhere in the network, by taking the average trading costs of reaching other markets, weighted by those markets' sizes (Donaldson and Hornbeck, 2016). In the case of Media Reach, the estimated coverage coefficients, $\hat{\psi}_1 + \hat{\psi}_2 \log(\text{Distance})_{o,d}$, play the role of trade-costs, and the eligible voters the role of market size.

However, there are three important differences with the market access measure. First, the weights are estimated from the coverage regression in equation 1 rather than derived from a structural model, so Media Reach is a reduced-form access measure that captures the two channels identified in the regression, news-agency access and physical distance.²¹ Second, whereas market access responds to changes anywhere in the network for every location, Media Reach responds to those changes only once a county is connected to a news agency (i.e. when a county is unconnected,

²¹Other dyadic features beyond distance that may drive news coverage, e.g. ethnic or political similarity between two counties, could be incorporated by adding them to equation 1 and carrying their fitted terms in equation 2.

Connect_{*o,t*} = 0, the measure is exactly zero). Our regressions therefore compare how Media Reach affects campaign strategy relative to the reachable audience of unconnected counties. Third, the two networks have different shapes. In Donaldson and Hornbeck (2016), counties are connected to one another through least-cost routes over the physical network, so building a segment between any two counties can change bilateral trade costs—and hence market access—everywhere. Here counties are not connected to one another directly but through a shared intermediary, so bilateral coverage depends only on whether both endpoints have access to any news agency and the distance between them. A third county’s connection therefore leaves the coverage unchanged between a pair of counties and enters Media Reach only by becoming an additional reachable county. This makes the source of every change in the measure easy to trace, but with a simpler structure of indirect effects than in the trade network.²²

4.3 Media Reach and Campaign Strategy

While in some electoral systems some places are more important than others from an electoral perspective, like the electoral college in the U.S., candidates’ objective function overall is to maximize their votes. How then should candidates respond to changes in the potential audience of a speech? Candidates could respond by changing where they speak. On the one hand, if reports of speeches can persuade voters and increase their votes, then all else equal, greater Media Reach should make a county more attractive because a speech delivered there could reach more newspaper readers elsewhere. On the other hand, better news distribution could also substitute for time-consuming travel if candidates could reach voters instead through published letters or other campaign documents.

Candidates could also respond by changing what they said. On the one hand, an audience-composition mechanism suggests that candidates might have been more likely to give a speech that was less aligned with the local audience and more with a national audience if they thought that the local topics might not be as relevant outside. For example, in 1900 a presidential candidate speaking in Scranton, an anthracite-mining center in Pennsylvania, might have had strong incentives to discuss miners’ wages and working conditions when addressing the local audience. However, if the speech was also likely to be reproduced by newspapers elsewhere, the candidate might have devoted relatively more attention to nationally salient issues, such as imperialism or the Philippines.²³ On the other hand, a newsworthiness mechanism suggests that candidates could have also tried to make their speeches more unique and tailored to a local audience if they were concerned that voters elsewhere might have already been exposed to similar speech content before.

To study campaign responses to media expansion, I estimate a two-way fixed effects regression in which the treatment of interest, media reach, is the interaction of a county’s own connection to a news agency and the audience it can reach through the network. The design is a difference-in-differences with a continuous dose, where the dose is the outside audience, in the sense of Callaway et al. (2024). It compares the effect of connecting across counties facing larger and smaller outside audiences or, equivalently, the response to growth in the outside audience between connected and

²²Because we don’t have a continuous cost measure of accessing news agencies, we ignore that a county getting access to a news agency might make it easier for its neighbors to access national news too.

²³This example comes from a newspaper in Scranton that reported a miners movement towards a strike. In the same page, it also reported a speech elsewhere by William J. Bryan, a presidential candidate in 1900, about imperialism in the Philippines. See <https://www.loc.gov/resource/sn84026355/1900-09-07/ed-1/?st=text&> and <https://www.loc.gov/resource/sn84026355/1900-09-08/ed-1/?sp=1&st=text&> (retrieved on August 12, 2026).

unconnected counties.

$$Y_{o,t} = \beta \text{MediaReach}_{o,t} + \alpha_1 \text{Connect}_{o,t} + \alpha_2 \text{OutsideAudience}_{o,t} + \theta \mathbf{X}_{o,t} + \gamma_o + \delta_t + \varepsilon_{o,t} \quad (3)$$

The outcome $Y_{o,t}$ is the political response of interest in potential speech county o , such as an indicator for whether a campaign visit occurred or the content composition of the reported speech.²⁴ I standardize $\text{OutsideAudience}_{o,t}$, so that $\text{MediaReach}_{o,t}$ equals the standardized outside audience for connected counties and zero otherwise. The regression is at the county-year level rather than the dyad level because in the construction of Media Reach, destination counties d have been summed out. The fixed effects are γ_o , for the potential speech county, and δ_t , for the campaign year. The controls in $\mathbf{X}_{o,t}$ cover the county’s own media, economic, demographic, political and transportation characteristics.²⁵ Because the outside audience is constructed from the estimates of equation 1, both it and media reach are generated regressors. I cluster standard errors by county. Accounting for the uncertainty in the coverage estimates does not affect inference.²⁶

Taking the visit indicator as outcome, α_1 is the change in probability of a campaign stop when a county connects to a news agency, evaluated at the average outside audience. The term α_2 is the change in probability per standard deviation of outside audience among unconnected counties. Since Media Reach is the product of the two components, β measures how much more the probability of a visit rises with the outside audience when the county is connected than when it is not.

With county fixed effects, β is identified from within-county changes in media reach, which arise when a county connects and when the network expands elsewhere, at different rates in different regions. Because β compares responses across counties facing different outside audiences, identification requires the strong parallel trends assumption of Callaway et al. (2024): absent the media channel, counties facing larger and smaller outside audiences, connected or not, would have followed the same path. This is more plausible here than in settings where units select their own dose, such as localities setting minimum wages, since a county’s outside audience is determined by the connection decisions of newspapers elsewhere and by distance rather than by anything county o does, though the geography behind a large outside audience may still shape how campaigns respond.

The identifying assumption can be violated in two ways, both stemming from the fact that a county’s outside audience is largest near populous, connected regions. First, a regional shock—a boom, a gold rush, a new railroad—could raise a connected county o ’s political salience and, by expanding nearby markets, lead nearby newspapers to connect, raising o ’s outside audience at the same time.²⁷ Second, campaigns might respond more to any change in such places, for instance because they lie on the rail routes that tours follow, regardless of how many readers the coverage

²⁴Content composition can be the share of content devoted to different topics, targeted at different audiences, or aligned with the county’s characteristics rather than the national average. Appendix C.5 describes the content labels.

²⁵The set of controls includes number of dailies, log population, manufacturing output per capita, share of men aged 21 and over, railroad access and length, log market access, and lagged turnout and lagged vote closeness.

²⁶Both regressors are built from the standardized outside audience, so they depend on the estimates of equation 1 only through the distance-decay ratio $\hat{\psi}_2/\hat{\psi}_1 = -0.099$ (with log distance not demeaned), whose 95% confidence interval is $[-0.118, 0.035]$. When I rebuild both regressors at every value in this interval and re-estimate equation 3, the coefficient in column (2) of Table 2 stays between 2.0 and 2.4. First-step uncertainty raises the standard errors in that table by less than 7% if the two estimation steps are independent, and by at most 39% if perfectly correlated.

²⁷The concern is not that newspapers elsewhere subscribed in order to obtain news from o . A news-agency subscription, like newspaper entry, was a fixed-cost decision driven primarily by a newspaper’s own market size and income (Gentzkow et al., 2011), so subscriptions elsewhere are unlikely to have responded to political events in o .

Table 1: Coverage of Presidential Speeches

Dep. Var.: 1[Coverage > 0] x 100	(1)	(2)	(3)	(4)
Both endpoints wired	1.22*** (0.07)	1.44*** (0.08)	1.45*** (0.17)	1.51*** (0.18)
Distance (logs), demeaned		-0.09*** (0.02)		
Both endpoints wired x Distance (logs)		-0.97*** (0.11)		-0.46* (0.26)
Same state	0.54*** (0.05)	0.33*** (0.05)		
Origin x Year FE	X	X	X	X
Destination x Year FE	X	X	X	X
Origin x Destination FE			X	X
Observations	1,037,498	1,037,498	692,312	692,312
R-squared	0.152	0.153	0.442	0.442

Notes: Distance is the straight line distance between the two counties' centroids. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by speech county x coverage county pairs.

reaches. The controls for railroad and market access absorb the level of these differences. To address the trends, I re-estimate the model using reach computed only from counties more than 100 miles from o , which removes both the nearby newspapers a regional shock would draw onto the wire and the dense regions behind the second concern. I also recompute reach with each county's eligible voters fixed at its 1880 level, so that reach varies only through changes in access to news agencies elsewhere rather than through population growth, which the same regional shocks would drive.

5 Results

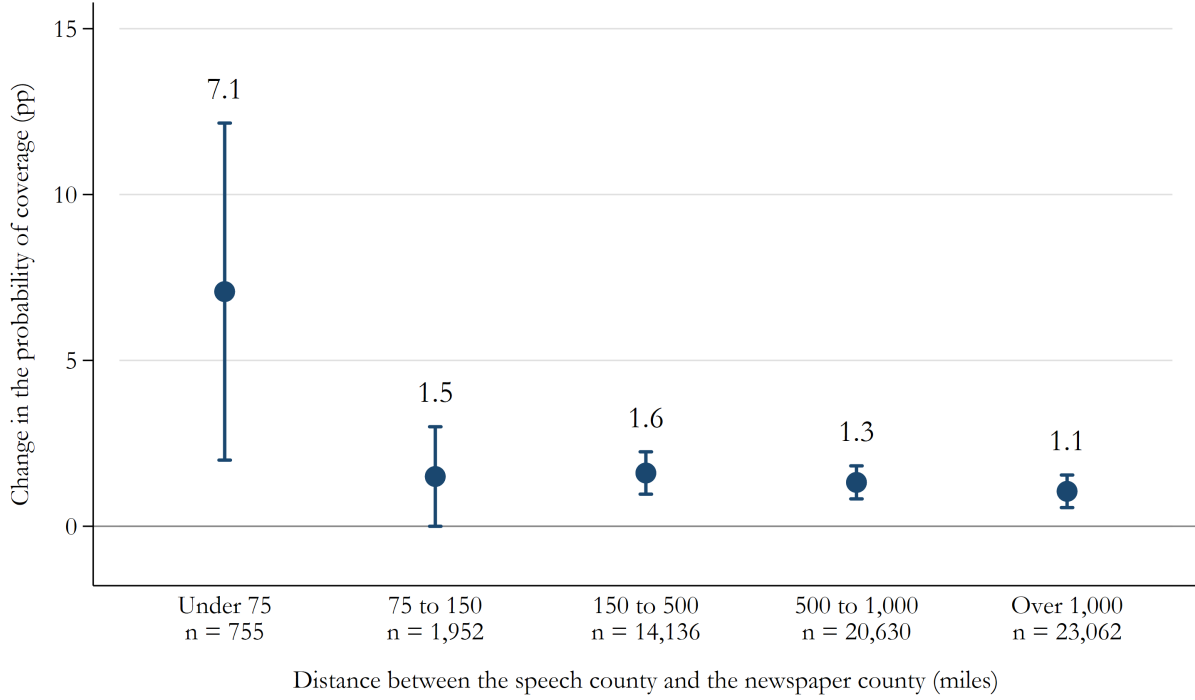
5.1 How News Agencies Increased Coverage of Events Across the Country

Table 1 presents the results of estimating equation 1 by OLS, where columns (1) and (2) estimate the model in the cross-section, while columns (3) and (4) estimate the model in the panel with county pair fixed effects. The first key result is that, as expected, once newspapers in both counties get connected to a news agency, the probability of coverage increases by 1.5 percentage points at the average log distance in the preferred specification, from a baseline of 0.4 percent among county pairs that were not both connected.

The second key result is that news agencies increased coverage between all counties, but especially for nearby ones. The interaction coefficient in column (4) implies that the connection effect declines by 0.5 percentage points per log point of distance, or roughly 0.3 percentage points each time the distance between the two counties doubles. The effect nonetheless remains positive across the continental U.S. once both endpoints are connected.²⁸ For example, once both counties were connected, a speech that occurred in Washington, DC was 1.5 percentage points more likely to be covered in Chicago (600 miles away), 1.1 percentage points in Denver (1,500 miles), and 0.9 percentage points in Los Angeles (2,300 miles).

²⁸Log distance is centered at its mean, around 570 miles. The estimates in column (4) turn negative only beyond 15,000 miles, farther than any two counties in the continental U.S. are apart.

Figure 9: Effect of News-Agency Connections by Distance Band



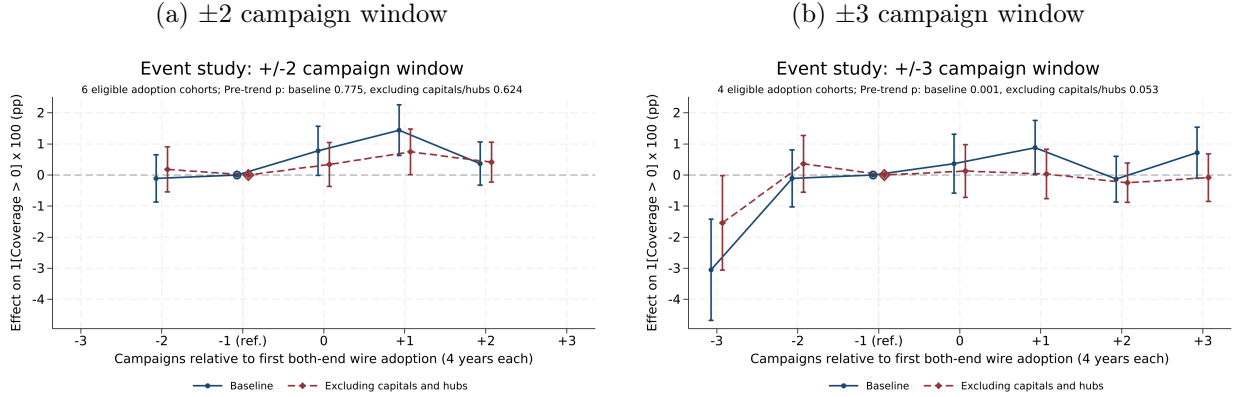
Notes: Estimated effect of the speech and coverage counties both being connected to a news agency on the probability that the coverage county covers the speech, separately for five bands of distance between the two counties. Connected-pair status varies over time within pairs, so I estimate one effect per band. The regression absorbs speech-county $\times$ year, coverage-county $\times$ year, and county-pair fixed effects. Label “n” is the number of connected pair-years in the band. Vertical bars are 95% confidence intervals from standard errors clustered by county pair.

While it might not come as a surprise that newspapers covered more events outside their county once they subscribed to a news agency, one might have expected distance to have mattered less upon connection, since newspapers could now cheaply cover events further away (i.e. a positive or null coefficient on the distance interaction term would mean distance’s deterrent effect is reduced). To better understand this, I estimate a modified version of eq. 1 that replaces the distance interaction with five distance bands, allowing the effect of both endpoints being connected to vary non-linearly with distance. Figure 9 shows that coverage increased the most for county pairs less than 75 miles apart and by a smaller, roughly constant amount at every distance beyond 75 miles. I also explore whether the effect of being connected and its interaction with distance varied over time, by interacting the main terms in equation (1) with campaign-year dummies. While the probability of coverage increased over time as counties got connected, the distance friction term did not attenuate (see Figures D1 & D2).²⁹ Although news agencies removed important supply barriers for the flow of information, demand for political news remained local.

These main findings are robust to a series of alternative specifications. The estimated coefficients from column (4) in Table 1 are unchanged if I account for year-specific effects of different characteristics, such as counties being in the same state or either county being a capital of that

²⁹There is also no evidence that non-linear effects varied over time (e.g. to test whether coverage first increased locally and later expanded to more regional and national level coverage). See Figure D4.

Figure 10: Event Studies Around First Joint Connection: Baseline vs. Excluding Capitals/Hubs



Notes: Event studies around the first campaign in which a county pair is both-connected, with event time in campaigns (four years each) and never-treated pairs as controls. Each panel overlays two separately estimated series: the baseline and the specification excluding capital and hub counties from either side of the pair, corresponding to columns of Table D6. Panel (a) shows the ± 2 window and includes six eligible adoption cohorts, with joint pre-trend p-values of 0.775 (baseline) and 0.624 (excluding capitals and hubs). Panel (b) shows the ± 3 window and includes four cohorts, with p-values of 0.001 and 0.053. Vertical bars are 95% confidence intervals from standard errors clustered by county pair.

state, 1880-population deciles of each county in the pair (Table D1) or if I drop the sample of pairs that were never simultaneously connected (Table D2). Even when I drop county pairs that include a state capital or any large city, which are those that might arguably drive demand for coverage from smaller counties, the coefficient on both endpoints being connected remains positive, despite falling by nearly half (from 1.5 to 0.8 percentage points), while the distance interaction is only slightly flatter (Table D3). These results are also robust to defining a newspaper as news-agency connected if it had at least two, five, or ten fast reprints per year (Figure D5).

Furthermore, I show that the timing of connection cannot be fully explained by a larger demand for coverage, using leads and lags of the connection status and event studies at the four-year cadence of presidential campaigns, subject to the horizon limits discussed in Section 4.1. Figure 10 shows that, even though there is some evidence of rising coverage in the campaigns leading up to connection, these effects disappear once I exclude county pairs that include a state capital or hub city (Tables D5 and D7 in Appendix D.4 report the corresponding leads-and-lags and event-study estimates). These results suggest that demand for news of the capitals and hub cities drove part of the timing of connection, but that anticipation is confined to those pairs: among ordinary county pairs the pre-trends are flat, and Appendix D.4 shows that the small remaining lead behaves like late dating of the connection measure—which is inferred from the first observed fast reprint—rather than like pair-directed selection.

Together, these results suggest that, while news agencies decreased the costs for newspapers to cover political events across the entire country, most increases in coverage were for nearby local counties (or large cities), and this trend persisted during the period of study.

5.2 Impact of Media Reach on Political Behavior

The estimates in column (4) of Table 1 are the coefficients $\hat{\psi}_1$ and $\hat{\psi}_2$ that enter the construction of Media Reach in equation 2.³⁰ This subsection asks whether candidates directed their appearances toward the counties where that audience was larger, and whether they also adjusted the content of the speeches they delivered there.

Location response Table 2 reports results from estimating equation 3. Media reach has a large and statistically significant impact on the probability that a county receives a campaign stop. Column (1) estimates the model only controlling for own access and outside audience terms, while column (2) includes controls for media, economic, demographic, political and transportation characteristics. In this preferred specification, a one-standard-deviation increase in the outside audience among connected counties raises the probability of a speech by around 2.3 percentage points relative to unconnected counties. This is roughly 47 percent of the mean probability of a speech among unconnected counties of 4.8 percent (the mean for connected counties is 10.9 percent). Note that one standard deviation of outside audience is roughly 44,000 additional voters reachable elsewhere, or the eligible voting population in nearly 6 average-sized counties. In other words, increasing the reachable audience by 10,000 voters, slightly more than the voters in one average-sized county, raises the probability of a speech by 0.5 percentage points or 11% of the unconnected mean.

Columns (3) through (6) address concerns of endogeneity by re-computing the outside audience term in equation 2 while turning off different sources of variation one at a time. To account for geographically correlated shocks, column (3) exploits only destination counties beyond 100 miles, so that the variation comes from the timing of distant connections alone. This yields a slightly smaller but statistically indistinguishable estimate from the preferred baseline specification with controls in column (2). Holding destination electorates at their 1880 levels in column (4), so that estimates aren't driven by changes in destination population, also yields a similar estimate. Counting all adults aged 21 and over as the eligible voters, rather than only men, has no effect on the estimate (column (5)), suggesting that mismeasurement or shocks correlated with women's suffrage are unlikely to drive the results. Finally, dropping state capitals and large urban centers (New York, Chicago, and D.C.) as destinations, leaves the estimate unchanged in column (6), so the response is not only explained by the pull of the great cities' audiences.

Media reach captures the effect of reaching a larger audience between counties with and without news-agency access. Yet, a speech given in an unconnected county would be harder to send through the network. Therefore, among counties without access to news agencies we might expect that the size of the outside audience should not predict where candidates went. Table E1 includes the coefficients for the outside audience among unconnected counties and the local news-agency access terms omitted in Table 2. Across most specifications and robustness checks, the coefficients for the outside audience among unconnected counties are small and statistically indistinguishable from zero.

In Appendix E.2 I show that candidates did not move toward counties whose reachable audience was about to grow. If anything, among counties already connected, next year's growth in the

³⁰See footnote 26 for how I account for the uncertainty in these estimates.

Table 2: Estimated Impact of Media Reach on Political Visits

Dep. Var.: $1[\text{Speech} > 0] \times 100$	(1) Baseline	(2) Baseline + Controls	(3) 100-Mile Buffer	(4) Fixed 1880 Population	(5) All Aged 21+	(6) No Capital/Hub Destinations
Media Reach	2.73*** (0.59)	2.26*** (0.57)	2.06*** (0.56)	2.15*** (0.60)	2.24*** (0.57)	2.18*** (0.56)
Controls	None	All	All	All	All	All
County FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Observations	23,048	23,048	23,048	23,048	23,048	23,048
R-squared	0.315	0.320	0.321	0.320	0.320	0.320

Notes: The dependent variable is an indicator for at least one speech in the county-year, times 100. Each column estimates equation 3 and reports β , the coefficient on media reach. The connection indicator and the outside audience are included in every regression (Table E1 reports them). The reachable audience elsewhere is the summation term in equation 2, defined for unconnected counties as well and standardized over all counties, so each coefficient displayed is the effect of a one standard deviation larger reachable audience among connected counties. Columns (2) through (6) add controls for the number of dailies published in the county, railroad indicator, railroad length, market access (log), population (log), manufacturing output per capita, own-county share of men 21+, lagged turnout (%) and lagged vote closeness (%). Columns (3) through (6) recalculate the OutsideAudience term using different specifications. Column (3) only includes counties beyond 100 miles of a county, column (4) holds counties' male 21+ population levels fixed at 1880 levels, column (5) counts all adults, rather than only men, aged 21 and over, column (6) excludes state capitals and New York City, Chicago, and D.C.. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by county. Accounting for first-step uncertainty in the outside audience raises them by less than 7%, or by at most 39% if the two estimation steps were perfectly correlated (footnote 26).

reachable audience predicts slightly fewer speeches, so it's likely that politicians' response followed rather than preceded the network that existed at the time of the campaign (Table E2).

Taken together, these results show that candidates allocated appearances toward counties from which the news network could carry their speeches to a larger outside audience. The response runs through the interaction of a county's own connection with the audience reachable elsewhere, as counties with large outside audiences but no capacity to send out the message were not more likely to receive a speech. The estimate is robust to fixing destination populations, redefining the audience, discarding nearby counties, and removing the capital and hub audiences entirely (Appendix E) or using different connection thresholds (Figure E1). Beyond substituting for personal campaigning, mass communication raised the value of appearing in well-connected places relative to poorly connected ones.

Content response Candidates could have responded to a larger outside audience not only by choosing better-connected places, but also by changing what they said once there. To test this second margin, I estimate equation 3 on the county-years with at least one speech and usable extracted content, with the full control set of Table 2. I replace the visit indicator with the composition of reported speech content (the share of content devoted to each topic, the shares directed at each audience target, and the shares concerning each policy jurisdiction). This sample is much smaller than the one Table 2—436 county-years rather than roughly 23,000—and the typical county-year contributes only one or two labeled speech reports, so these estimates are considerably less precise. I also use state rather than county fixed effects because few counties contribute content in more than one campaign.

The two mechanisms described in Section 4 make opposite predictions for these shares. If candidates redirected their message toward the national audience their speeches could reach, content

Table 3: Media Reach and the Content of Speeches

	Mean	Media reach	(s.e.)	[RW p]
<i>Panel A: Speech topics</i> (family: 21 topic shares)				
Party record / opponent	27.6	5.54	(3.37)	[0.968]
Money & currency	11.9	-0.29	(2.25)	[1.000]
Governance & constitution	8.6	-0.17	(2.11)	[1.000]
Foreign affairs & war	8.2	1.08	(2.70)	[1.000]
Trusts & regulation	6.4	-0.83	(1.72)	[1.000]
Labor & industry	7.1	-3.22	(2.51)	[0.996]
Tariff & trade	5.6	-0.88	(2.37)	[1.000]
<i>Panel B: Audience target (place)</i> (family: 4 place shares)				
Not place-bound	82.5	-5.14*	(2.94)	[0.249]
State / region-bound	8.2	-0.70	(2.53)	[0.956]
Local place-bound	4.4	5.69***	(1.90)	[0.034]
<i>Panel C: Policy jurisdiction</i> (family: 4 jurisdiction shares)				
Federal / national	61.5	-0.95	(4.63)	[0.984]
Non-policy	29.1	0.71	(3.83)	[0.984]
Foreign / international	7.9	0.28	(2.66)	[0.984]

Notes: Each row reports β from a separate estimate of equation 3 for the named content share (multiplied by 100), estimated on the 436 county-years with at least one detected speech and usable extracted content, 1880–1916. “Mean” is the sample mean of the share. All regressions include state and year fixed effects, the outside audience, and the control set of columns (2)–(6) of Table 2. Media reach and the outside audience are standardized as in Table 2. Standard errors, in parentheses, are clustered by county. The bracketed column reports Romano–Wolf stepdown p -values, computed by cluster bootstrap (1,000 replications) within the family of shares named in each panel header; each panel displays a subset of its family (seven of the 21 topic shares, three of the four audience-target shares, and three of the four jurisdiction shares), while the adjustment is computed over the full family. Residual and unclassifiable shares are excluded from the families. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, based on unadjusted p -values.

in high-reach counties should shift away from locally bound themes and toward nationally framed ones. If instead candidates valued distinctive, newsworthy appearances, greater reach could push content toward the local and particular. Table 3 reports each share’s coefficient on media reach (second column) together with Romano–Wolf stepdown p -values that adjust for multiple testing within each classification scheme.

The overall answer is that content responded little, if at all. No topic share moves with media reach in any of the twenty-one tested categories. The smallest unadjusted p -value in the topic family is 0.10. The nationalization prediction finds no support—the shares of content that was not explicitly tied to any specific location (Panel B) and concerning federal or national policy (Panel C) are, if anything, negative. The only share that moves significantly runs in the opposite direction, as the share of content where the candidate explicitly refers to a specific county rises by 5.7 percentage points per standard deviation of media reach, from a mean of only 4.4 percent (Romano–Wolf p -values are 0.03 and 0.58 when adjusted across the 4 audience groups and all the 36 regressions, respectively). However, the design has limited power, as the standard errors imply that effects smaller than roughly 5 to 13 percentage points per standard deviation of reach—a sizable fraction of most shares’ means—cannot be ruled out. I read these findings as at most suggestive evidence that reach favored locally distinctive material—the direction the newsworthiness mechanism predicts.

Rather than shifting attention across topics, media reach could instead have changed which reference points candidates’ messages followed and how standardized they were across stops. To examine this, I move from county-year shares to the individual newspaper reports of a speech, and score each report against four reference points. The first asks whether content drifted toward what the country was already talking about: whether the report’s main topic was one of the year’s domi-

Table 4: Media Reach, National References, and Message Discipline

Dep. Var. (x100)	National issues	National salience	Own platform	Opp. platform	Own – opp.	Own-stump share
Media reach	1.19 (3.54)	0.98 (1.11)	3.09** (1.41)	-2.36** (1.08)	5.45*** (1.23)	2.39* (1.22)
State FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Mean dep. var.	28.8	7.2	15.9	11.5	4.4	19.3
Observations	2,410	1,174	1,331	1,331	1,331	2,323
R-squared	0.210	0.182	0.415	0.356	0.414	0.135

Notes: Each column is a separate regression at the level of individual newspaper reports that directly attribute content to a candidate’s speech, 1880–1916. The dependent variables, each multiplied by 100, are: (1) an indicator that the report’s main topic is on the year’s pre-specified list of dominant national issues; (2) the share of the year’s national newspaper conversation devoted to the report’s topic, measured from topic-term frequencies in *Chronicling America*; (3) the share of the report’s content matching the candidate’s own party platform; (4) the share matching the opposing platforms; (5) the difference between (3) and (4); and (6) the leave-out share of the candidate’s other reported stops in the same campaign discussing the report’s topic. All columns include additive state and year fixed effects, the outside audience, and the control set of columns (2)–(6) of Table 2: the county’s own news-agency connection, number of dailies, railroad indicator and length, log market access, log population, manufacturing output per capita, share of men aged 21 and over, lagged turnout, and lagged vote closeness. Media reach and the outside audience are standardized as in Table 2. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by county.

nant national issues, and how large a share of the year’s national newspaper conversation that topic occupied. The next asks whether it followed the candidate’s party script or the opponent’s instead: how much of the report’s content matched the candidate’s own party platform—the statement of positions a party adopts at its convention, of which each numbered position is a *plank*—and how much matched the opposing platforms. The last asks how uniform the stump message was: the share of the candidate’s *other* reported stops in the same campaign that discussed the same topic.³¹

Table 4 reports β for each of these outcomes, estimated at the report level with the same specification as Table 3. Columns (1) and (2) show no pull toward generic national reference points: reports from high-reach counties were no more likely to cover the year’s dominant campaign issues, nor the topics most salient in the national press. Columns (3) through (5) show that content did follow one national reference—the candidate’s own party platform. The share of content matching the candidate’s own platform is 3.1 percentage points higher per standard deviation of media reach, the share matching the opponents’ platforms is 2.4 percentage points lower, and the gap between the two widens by roughly 5 percentage points. Column (6) shows the same discipline within the candidate’s own campaign: where reach was greater, the share of the candidate’s other stops discussing the same topics is 2.4 percentage points higher, though this estimate is significant only at the ten percent level. These results suggest that media reach could have helped build party discipline and coherence across the candidate’s stops, without changing the topics they discussed.

Two further content margins were tested. The first concerns values rather than topics, by exploring the community invoked in a speech (e.g. the nation or a specific group) and the warrant it offers for support (e.g. party, identity, stake). On the county-years with usable direct content, community references shift with media reach in the direction the audience-composition mechanism

³¹The national-issue lists were fixed from standard campaign histories before estimation; the national-conversation measure counts topic-matched newspaper pages in *Chronicling America*; the platform vectors label the 681 planks of all 22 party platforms of 1880–1916 with the same topic labeler as the speech content (Woolley and Peters, 2026); and generic campaign material is excluded from the national-issue and platform references, which is why sample sizes differ across columns. Appendix C.5 details the construction of all four reference points.

predicts, jointly significant at the ten percent level. National framing rises and class- or group-based framing falls, while the warrant composition does not move (Table F1). The second focuses on whether the content of a speech matches the economic characteristics and interests of the county it was delivered in. I find no evidence that media reach changed this alignment, but the comparison rests on 261 county-campaigns, so it should not be read as an informative null.³²

An important caveat applies to these estimates. They are descriptive associations about the content that was extracted, which depends on characteristics of the newspapers that chose to cover them. While most articles covering speeches come from newspapers outside the State where the speech occurred, and so editorial preferences might not bias these estimates, more work is needed to understand how the press filtered and framed the candidates’ words.

With these caveats, the results here suggest that candidates neither reallocated attention across topics nor drifted toward the nationally salient issues of the moment. Instead, where the press could carry their words further, they repeated their party’s national script more uniformly and framed their appeals in more national, less class-bound terms.

6 Conclusion

Researchers have long recognized that media can influence those exposed to it, and have documented how politicians respond to the media environment by timing their actions, adjusting what they say, and shifting their policy decisions to persuade voters. This paper shows that media also creates incentives for politicians to adjust *where* they speak. Using historical newspaper articles, I create a dataset of presidential campaign speeches and their media coverage in the United States between 1880 and 1916. I show that the expansion of news agencies, which allowed for the rapid dissemination of news, increased coverage of speeches differentially across the country. In turn, candidates responded by directing their appearances toward the counties from which their speeches could reach a larger outside audience. The content of what they said changed less: they did not reallocate attention across topics but repeated the party’s national script more uniformly—a pattern measured on a far smaller sample and best read as suggestive.

Given that candidates strategically adjusted their campaign stops to reach larger audiences, what does this mean for measuring the effects of speeches on political outcomes? A long literature has found mixed results on the effects of rallies on turnout and vote choice (Shaw, 1999; Hill et al., 2010; Wood, 2016; Kalla and Broockman, 2018; Devine, 2024; Snyder and Yousaf, 2020; Pino and Salisbury, 2025; Buggle and Vlachos, 2023; Heersink and Peterson, 2017; Assouad, 2025). The findings in this paper suggest that one reason for these conflicting results may be that the effects are measured on the wrong population. If newspaper coverage carried exposure to a larger outside audience, and candidates strategically acted upon this, then analyses limited to the local population may miss an important part of the effect of campaign speeches on political outcomes.

The findings here are also relevant for understanding the effects of media on political behavior more broadly, and they raise questions about the longer-run consequences of a national news market. The standardization of the stump message that accompanied greater reach hints at a possible mechanism behind the nationalization of American campaigns: when the same words could travel

³²More details and results are available in Appendix F.

everywhere, candidates increasingly spoke to the nation rather than only to the crowd in front of them. Whether this homogenization of discourse helped consolidate national parties, whether split-ticket voting across levels of government declined as campaigns nationalized, and whether voters came to see themselves as citizens of the nation rather than of their states—following, and caring about, news from places far from their own—are questions the speech and coverage data assembled here could help answer. Finally, as radio, television, and the internet succeeded the wire report, each new technology again changed which audiences a physical appearance could reach and at what cost. How these later transformations reshaped politicians’ incentives over where—and whether—to appear in person is an important question for future research.

References

- Agrawal, A. and Goldfarb, A. (2008). Restructuring research: Communication costs and the democratization of university innovation. *American Economic Review*, 98(4):1578–1590.
- Allen, T. (2014). Information frictions in trade. *Econometrica*, 82(6):2041–2083.
- Althaus, S. L., Nardulli, P. F., and Shaw, D. R. (2002). Candidate Appearances in Presidential Elections, 1972–2000. *Political Communication*, 19(1):49–72.
- Associated Press (1918). “*M.E.S.: His Book, a Tribute and a Souvenir of the Twenty-Five Years, 1893–1918, of the Service of Melville E. Stone as General Manager of the Associated Press*.” Harper & Brothers, New York. William McKinley’s November 5, 1896 letter to Melville E. Stone appears on p. 50.
- Assouad, L. (2025). Leadership and nation-building. Working paper, London School of Economics and Political Science. Previously circulated as “Charismatic Leaders and Nation-Building”.
- Atack, J. (2016). Historical Geographic Information Systems (GIS) Database of U.S. Railroads, 1826–1911. Vanderbilt University.
- Bartels, L. M. (1985). Resource allocation in a presidential campaign. *The Journal of Politics*, 47(3):928–936.
- Berkes, E., Karger, E., and Nencka, P. (2023). The Census Place Project: A Method for Geolocating Unstructured Place Names. *Explorations in Economic History*, 87:101477.
- Blondheim, M. (2004). Rehearsal for media regulation: Congress versus the telegraph-news monopoly, 1866–1900. *Federal Communications Law Journal*, 56(2):299–328.
- Brams, S. J. and Davis, M. D. (1974). The 3/2’s rule in presidential campaigning. *American Political Science Review*, 68(1):113–134.
- Bryan, W. J. (1896). *The First Battle: A Story of the Campaign of 1896*. W. B. Conkey Company, Chicago.
- Buggle, J. C. and Vlachos, S. (2023). Populist Persuasion in Electoral Campaigns: Evidence from Bryan’s Unique Whistle-Stop Tour. *The Economic Journal*, 133(649):493–515.

- Callaway, B., Goodman-Bacon, A., and Sant’Anna, P. H. C. (2024). Difference-in-differences with a continuous treatment. NBER Working Paper 32117, National Bureau of Economic Research.
- Chen, L. J. and Reeves, A. (2011). Turning out the base or appealing to the periphery? An analysis of county-level candidate appearances in the 2008 presidential election. *American Politics Research*, 39(3):534–556.
- Dell, M., Carlson, J., Bryan, T., Silcock, E., Arora, A., Shen, Z., D’Amico-Wong, L., Le, Q., Querubin, P., and Heldring, L. (2023). American Stories: A Large-Scale Structured Text Dataset of Historical U.S. Newspapers. In *Advances in Neural Information Processing Systems*, volume 36.
- DellaVigna, S. and Gentzkow, M. (2010). Persuasion: Empirical Evidence. *Annual Review of Economics*, 2(Volume 2, 2010):643–669.
- Devine, C. J. (2024). *I’m Here to Ask for Your Vote: How Presidential Campaign Visits Influence Voters*. Columbia University Press, New York.
- Djourelouva, M. and Durante, R. (2022). Media attention and strategic timing in politics: Evidence from U.S. presidential executive orders. *American Journal of Political Science*, 66(4):813–834.
- Donaldson, D. and Hornbeck, R. (2016). Railroads and american economic growth: A “market access” approach. *The Quarterly Journal of Economics*, 131(2):799–858.
- Durante, R. and Zhuravskaya, E. (2018). Attack when the world is not watching? U.S. news and the Israeli–Palestinian conflict. *Journal of Political Economy*, 126(3):1085–1133.
- Ellis, R. J. and Dedrick, M. (1997). The Presidential Candidate, Then and Now. *Perspectives on Political Science*, 26(4):208–216.
- García Jimeno, C. and Yildirim, P. (2025). Persuasion and dissuasion in political campaigns: Political communication and media coverage in senate races. Technical report, Working Paper.
- García-Jimeno, C., Iglesias, A., and Yildirim, P. (2022). Information Networks and Collective Action: Evidence from the Women’s Temperance Crusade. *American Economic Review*, 112(1):41–80.
- Gaspar, J. and Glaeser, E. L. (1998). Information technology and the future of cities. *Journal of Urban Economics*, 43(1):136–156.
- Gentzkow, M., Petek, N., Shapiro, J. M., and Sinkinson, M. (2015). Do newspapers serve the state? incumbent party influence on the us press, 1869–1928. *Journal of the European Economic Association*, 13(1):29–61.
- Gentzkow, M. and Shapiro, J. M. (2010). What Drives Media Slant? Evidence from U.S. Daily Newspapers. *Econometrica*, 78(1):35–71.
- Gentzkow, M., Shapiro, J. M., and Sinkinson, M. (2011). The Effect of Newspaper Entry and Exit on Electoral Politics. *American Economic Review*, 101(7):2980–3018.
- Gilardi, F., Alizadeh, M., and Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30):e2305016120.

- Haines, M. R. and ICPSR (2010). Historical, Demographic, Economic, and Social Data: The United States, 1790–2002. ICPSR 02896-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research (ICPSR) [distributor], 2010-05-21.
- Harpine, W. D. (2000). Playing to the press in McKinley’s front porch campaign: The early weeks of a nineteenth-century pseudo-event. *Rhetoric Society Quarterly*, 30(3):73–90.
- Heersink, B. and Peterson, B. D. (2017). Truman defeats Dewey: The effect of campaign visits on election outcomes. *Electoral Studies*, 49:49–64.
- Hill, J. S., Rodriguez, E., and Wooden, A. E. (2010). Stump speeches and road trips: The impact of state campaign appearances in presidential elections. *PS: Political Science & Politics*, 43(2):243–254.
- Hilpert, J. M. (2015). *American Cyclone: Theodore Roosevelt and His 1900 Whistle-Stop Campaign*. University Press of Mississippi, Jackson.
- Hochfelder, D. (2012). *The Telegraph in America, 1832–1920*. Johns Hopkins University Press, Baltimore.
- Hornbeck, R. and Rotemberg, M. (2024). Growth Off the Rails: Aggregate Productivity Growth in Distorted Economies. *Journal of Political Economy*, 132(11):3547–3602.
- Jensen, R. (2007). The digital provide: Information (technology), market performance, and welfare in the south indian fisheries sector. *The quarterly journal of economics*, 122(3):879–924.
- Kalla, J. L. and Broockman, D. E. (2018). The minimal persuasive effects of campaign contact in general elections: Evidence from 49 field experiments. *American Political Science Review*, 112(1):148–166.
- Korinek, A. (2025). AI Agents for Economic Research: Mid-2025 Update to “Generative AI for Economic Research: Use Cases and Implications for Economists”. *Journal of Economic Literature*, 61(4). August 2025 update.
- Le, Q. (2026). Network competition and exclusive contracts: Evidence from news agencies. Working paper, Harvard Business School (R&R, American Economic Review).
- Le Pennec, C. (2024). Strategic campaign communication: Evidence from 30,000 candidate manifestos. *The Economic Journal*, 134(658):785–810.
- Lee, A. M. (1937). *The Daily Newspaper in America: The Evolution of a Social Instrument*. Macmillan, New York.
- Leip, D. (2026). Dave Leip’s Atlas of U.S. Presidential Elections.
- Library of Congress (2026). Chronicling America: Historic American Newspapers. National Digital Newspaper Program, Library of Congress and National Endowment for the Humanities. Newspaper directory and digitized pages.

- North, S. N. D. (1884). *History and Present Condition of the Newspaper and Periodical Press of the United States*, volume 8 of *Tenth Census of the United States, 1880*. U.S. Government Printing Office, Washington, DC.
- Petrova, M. (2011). Newspapers and parties: How advertising revenues created an independent press. *American Political Science Review*, 105(4):790–808.
- Pino, F. and Salisbury, L. (2025). The roots of the modern american presidential campaign. NBER Working Paper 34447, National Bureau of Economic Research.
- Prat, A. and Strömberg, D. (2013). The political economy of mass media. *Advances in economics and econometrics*, 2:135.
- Rosewater, V. (1930). *History of Coöperative News-Gathering in the United States*. D. Appleton, New York.
- Ruggles, S., Flood, S., Sobek, M., Backman, D., Cooper, G., Rivera Drew, J. A., Richards, S., Rogers, R., Schroeder, J., and Williams, K. (2025). IPUMS USA: Version 16.0 [dataset]. Minneapolis, MN: IPUMS.
- Sanderson, N. (2008). William Jennings Bryan and the Railroad. Digital project, part of *Railroads and the Making of Modern America*, University of Nebraska–Lincoln.
- Schwarzlose, R. A. (1990). *The Nation’s Newsbrokers. Volume 2: The Rush to Institution, from 1865 to 1920*. Northwestern University Press, Evanston, IL.
- Shaw, D. R. (1999). A Study of Presidential Campaign Event Effects from 1952 to 1992. *The Journal of Politics*, 61(2):387–422.
- Shmanske, S. (1986). News as a public good: Cooperative ownership, price commitments, and the success of the associated press. *Business History Review*, 60(1):55–80.
- Silberstein-Loeb, J. (2014). *The International Distribution of News: The Associated Press, Press Association, and Reuters, 1848–1947*. Cambridge University Press, Cambridge.
- Silcock, E., Arora, A., D’Amico-Wong, L., and Dell, M. (2024). Newswire: A Large-Scale Structured Database of a Century of Historical News. In *Advances in Neural Information Processing Systems*, volume 37.
- Snyder, James M., J. and Yousaf, H. (2020). Making Rallies Great Again: The Effects of Presidential Campaign Rallies on Voter Behavior, 2008-2016. Working Paper 28043.
- Snyder, Jr., J. M. and Strömberg, D. (2010). Press coverage and political accountability. *Journal of Political Economy*, 118(2):355–408.
- Steinwender, C. (2018). Real Effects of Information Frictions: When the States and the Kingdom Became United. *American Economic Review*, 108(3):657–696.
- Storper, M. and Venables, A. J. (2004). Buzz: Face-to-face contact and the urban economy. *Journal of Economic Geography*, 4(4):351–370.

- Strömberg, D. (2004a). Mass media competition, political competition, and public policy. *The Review of Economic Studies*, 71(1):265–284.
- Strömberg, D. (2004b). Radio’s impact on public spending. *The Quarterly Journal of Economics*, 119(1):189–221.
- Stromberg, D. (2008). How the Electoral College Influences Campaigns and Policy: The Probability of Being Florida. *American Economic Review*, 98(3):769–807.
- Strömberg, D. (2015). Media and politics. *Annual Review of Economics*, 7:173–205.
- Thomas, William G., I., Healey, R., and Cottingham, I. (2008). Railroads and the Making of Modern America. Digital project, University of Nebraska–Lincoln.
- U.S. Bureau of the Census (1960). *Historical Statistics of the United States, Colonial Times to 1957*. U.S. Government Printing Office, Washington, DC.
- U.S. Bureau of the Census (1975). *Historical Statistics of the United States, Colonial Times to 1970*. U.S. Government Printing Office, Washington, DC.
- Wang, T. (2025). The electric telegraph, news coverage, and political participation. *The Journal of Economic History*, 85(1):1–32.
- Wood, T. (2016). What the heck are we doing in Ottumwa, anyway? Presidential candidate visits and their political consequence. *The Annals of the American Academy of Political and Social Science*, 667(1):110–125.
- Woolley, J. T. and Peters, G. (2026). The American Presidency Project. Santa Barbara, CA: University of California.

Appendix

A	Historical Series on the Telegraph and News Agencies	36
B	Measuring News-Agency Connections	37
B.1	The Trend toward Same-Day Reprinting	37
B.2	Fast Reprints as Agency Copy	38
B.3	Stricter Connection Thresholds	40
C	Measuring Speeches, Coverage, and Content	42
C.1	Article Frame, Prompt, and API Implementation	42
C.2	From Labeled Events to the County Measure	46
C.3	Validation against the Historical Record	47
C.4	Maps	55
C.5	Content Extraction, Topic Labels, and Values Labels	56
D	Robustness and Timing of the Coverage Effect	58
D.1	The Effect by Campaign Year and the Stumping-Campaign Sample	58
D.2	Pair-Level Trend Controls and Capital/Hub Exclusions	62
D.3	Stricter Connection Thresholds	64
D.4	Identification Tests: Leads, Lags, and Event Studies	65
E	Robustness and Identification of the Media-Reach Response	71
E.1	The Full Decomposition	71
E.2	Anticipation of Network Growth	72
E.3	Stricter Connection Thresholds	73
F	Additional Content Results	75
G	Robustness to the Labeling Model: The GPT-4o Labels	76
G.1	The Prompt and the Run	76
G.2	Differences in the Resulting Measure	76
G.3	Results under the GPT-4o Labels	78

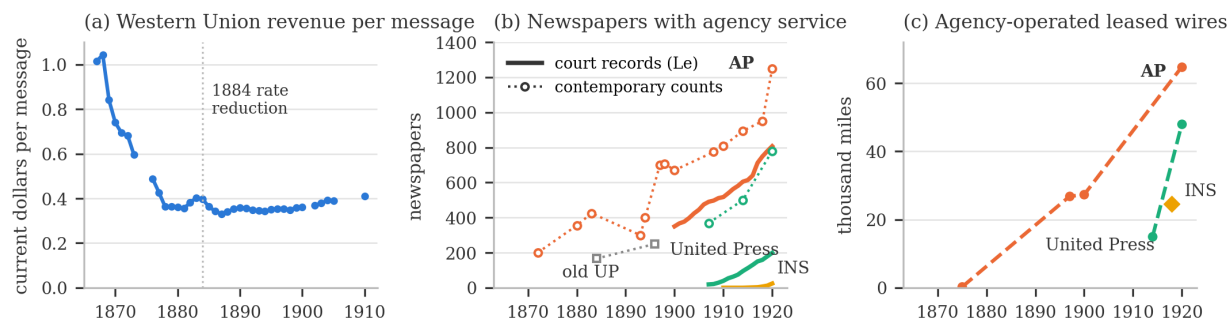
A Historical Series on the Telegraph and News Agencies

This appendix collects the three historical series behind Section 2.2. Panel (a) of Figure A1 asks whether the telegraph itself became cheaper over the sample period. Western Union’s average revenue per message fell by two thirds between 1868 and 1878, before the sample begins, and then stayed between 33 and 41 cents through 1910. The general rate reduction of 1884 lowered it by about seven cents, and by 1890 it was back at 36 cents. In real terms the price rose by about a sixth between 1880 and 1910, because consumer prices fell by a seventh between 1880 and 1896 and were still slightly below their 1880 level in 1910 (U.S. Bureau of the Census, 1975, Series E 135). A cheaper wire therefore cannot explain the spread of same-day news documented in Section 3.2.

Panel (b) shows how many newspapers had that access. The contemporary counts rise from about 200 dailies receiving the Associated Press report in 1872 to 1,250 members in 1920. The series changes definition in 1893: until 1883 it counts the dailies that received the report through the association and its regional auxiliaries, and from 1893 the members of the reorganized Associated Press of Illinois. The fall between 1883 and 1893 is this change of definition, and the rise between 1894 and 1897 includes the papers of the old United Press, which joined when it failed. Membership court records understates how many papers printed Associated Press material before 1900, because they only list papers that were still in operation in 1941.

Panel (c) shows the distribution layer the agencies built for that report. Leased press wires grew from a single circuit in 1875 to 65,000 miles for the Associated Press alone in 1920, and the United Press and the International News Service added networks of their own after 1907. Taken together, the price of transmission was stable while access to a shared report and the wires that delivered it expanded, which is the sense in which Section 2.2 calls the variation over the sample organizational rather than infrastructural.

Figure A1: Telegraph prices and the expansion of news-agency service, 1867–1920



Notes: This figure shows the price of telegraph transmission and the growth of news-agency service. Panel (a) divides Western Union’s annual revenues by messages handled, in current dollars (U.S. Bureau of the Census, 1960, Series R 45 and R 47); the series averages all traffic, most of it commercial, gaps are years not reported or not legible in the source table, and the dotted line marks the general rate reduction of 1884. Panel (b) plots contemporary counts of newspapers with agency service as hollow markers and, as solid lines, the memberships that Le (2026) reconstructs from the join dates in the *AP v. United States* (1945) trial rosters, which list only papers alive in 1941 and so understate early levels. Panel (c) plots leased press-wire mileage operated by the agencies. Sources for panels (b) and (c): Lee (1937, pp. 509–539); Rosewater (1930, pp. 157, 245–246, 253); Schwarzlose (1990). A stable price beside a growing membership and wire network is consistent with the expansion of same-day news being organizational.

B Measuring News-Agency Connections

This appendix provides additional evidence on the reprint data behind Figure 3 and on the connection measure that I build from them. I use four terms throughout. A *cluster* is a set of articles that Silcock et al. (2024) identify as reprints of the same underlying story. I refer to a cluster as a story. By construction, it contains at least four newspapers. An *incidence* is one newspaper’s printing of one story. The *reprint lag* of an incidence is the number of days between the first printing of the story anywhere in the data and that newspaper’s own printing of it. A lag of zero means that the newspaper ran the story on the day it first appeared. The *dateline* of a story is the place the story was reported from. All figures use Chronicling America newspapers only and therefore stop in 1919.

The appendix has three parts. Appendix B.1 asks whether the shift toward same-day reprinting reflects a real change in the news distribution system or a change in what the data detect as a reprint. Appendix B.2 asks whether a reprint within one day reflects access to a news agency or a newspaper copying a story from another newspaper. Appendix B.3 defines the stricter connection thresholds that I use as robustness checks in the rest of the paper.

B.1 The Trend toward Same-Day Reprinting

Panel (a) of Figure B1 shows the full distribution of reprint lags behind panel (a) of Figure 3, by decade. Reprinting concentrates on the day of first printing, and the same-day share rises from 55 percent in the 1880s to 74 percent in the 1910s while every slower lag loses share. The one-day window that defines connection therefore covers most reprints at the start of the sample period and nearly all of them by the end, so the connection measure does not depend on a narrow cutoff.

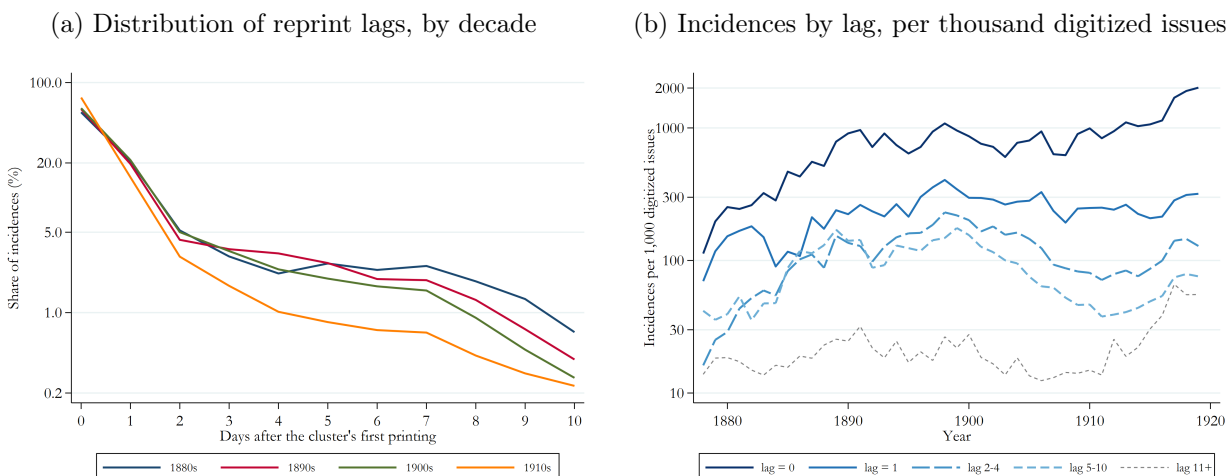
The timing of the increase matters. In the annual series of panel (a) of Figure 3, the same-day share does not rise in the 1880s, when the telegraph network was still expanding, and it is flat or slightly declining between 1893 and 1903. It rises steeply after 1903, the years in which the agencies replaced express-shipped plates with pony reports sent by wire and telephone and in which the United Press and the International News Service entered the market. The rise therefore follows the reorganization of news distribution described in Section 2.2 and not the construction of the wires.

Both panels of Figure 3 use only the articles that the data detect as reprints. A newspaper that received a story late and rewrote it in its own words appears in neither, so if rewriting became more common over time, both panels would change even if the news distribution system did not. Panel (b) of Figure B1 addresses this concern by dividing the same counts by the number of digitized Chronicling America issues, a denominator that does not depend on whether any reprint was detected. Same-day incidences per thousand issues nearly triple between 1885–1889 and 1915–1919 and never decline, incidences at lags of two to four days are flat, and only those at five to ten days fall. If the trend came from late copies going undetected, slow reprints would fall while fast reprints stayed constant, which is the opposite of what the figure shows. The same denominator answers the concern for panel (b) of Figure 3, where a story enters the data only if four newspapers reprinted it. The number of stories per thousand issues rises by half over the period, so the data gain stories rather than lose small ones.

Scaling by issues also decomposes the increase. Same-day incidences per issue are the product

of stories per issue, which rise by 53 percent, and the newspapers that each story reaches on its first day, which rise by 83 percent, so the amount of same-day agency news in a typical issue roughly tripled.³³

Figure B1: The trend toward same-day reprinting, 1878–1919



Notes: This figure shows the reprint lags of *Chronicling America* incidences, 1878–1919, on logarithmic vertical scales. Panel (a) plots the share of incidences at each reprint lag by decade of publication; lags above 10 days are omitted from the plot but included in the denominator, and the partial decade 1878–1879 is omitted (2.68 million incidences, 1880–1919). Panel (b) plots incidences at each lag per thousand digitized issues. The denominator is built from Library of Congress title metadata, spreading each digitized title's issue count evenly over the years in which it has at least one digitized issue and summing across the 3,267 titles with an issue calendar (of 4,606 digitized titles); it captures the entry and exit of titles but not year-to-year variation within a title. Numerator and denominator come from different sources, so the levels are not meaningful and only the trends and their ratios are. A rising fast layer beside flat or falling slow layers is consistent with growth in agency distribution rather than in the detection of reprints.

B.2 Fast Reprints as Agency Copy

Section 3.2 argues that a reprint within one day reflects a shared agency report and not a story copied from another newspaper. This section presents the evidence for that claim. The evidence in Appendix B.1 cannot address this concern, because newspapers that simply became faster at copying one another would produce the same patterns in panel (b) of Figure B1. I use two features of the data that distinguish an agency report from copying between newspapers.

Distance from the dateline. In this period, every way of redistributing news without a wire was physical. Newspapers exchanged copies by mail, and syndicates shipped stereotype plates and readyprint by express. All of these took days to arrive (Lee, 1937). A newspaper that printed a story within a day of its first appearance, several hundred miles from its dateline, therefore did not receive it by rail. Panel (a) of Figure B2 shows the distribution of the distance between the dateline and the reprinting newspaper for each group of reprint lags. I measure distance between

³³Reprints at lags of two to four days and of five to ten days increase through the 1890s together with the fast reprints and peak in 1898–1899. They then fall by 44 and 57 percent, respectively, by 1919. These are the years in which the agencies withdrew from express-shipped stereotype plates and readyprint in favor of pony reports sent by wire and telephone.

county centroids, so a reprint in the dateline’s own county has a distance of zero. Fast reprints are not local. Among incidences with a lag of one day or less, the median distance is 824 miles, 70 percent are more than 500 miles away, and 1.0 percent are in the dateline’s own county. Fast reprints are also only slightly more local than slow ones. The share within 250 miles is 15 percent at a lag of zero and 8 percent at lags of two to four days, and the rest of the two distributions is very similar. Finally, reprints did not become more local as fast reprinting grew. The share of fast reprints beyond 500 miles rises from 60 percent in the 1880s to 72 percent in the 1910s.³⁴

Simultaneity. A newspaper that copies a story from another newspaper does so one newspaper at a time. In contrast, an agency sends a single report to many newspapers at once, as described in Section 2.2. If fast reprints come from an agency, many newspapers in many states should print the same story within a day of one another. Panel (b) of Figure 3 provides the first piece of evidence. Pooling all years, the average story is printed by 4.26 newspapers on the day it first appears and by another 1.21 newspapers on the next day. In comparison, 0.59 newspapers print it at lags of two to four days, 0.42 at five to ten days, and 0.11 after that. In total, 5.47 newspapers print the average story within a day and 1.12 print it later, a ratio of about five to one.³⁵ Panel (b) of Figure B2 provides the second piece of evidence. For each story, I define the same-day group as the newspapers that printed it with a lag of zero or one day. I then count the newspapers in the group and the states where they are located. In 1878, the median fast reprint belongs to a same-day group of four newspapers in three states. By 1919, it belongs to a group of eight newspapers in seven states. Over the same period, the mean rises from 4.0 to 10.7 newspapers and from 3.6 to 8.7 states. Pooling all years, 94.4 percent of fast incidences belong to a group that spans three or more states, 61.2 percent to a group that spans five or more, and only 1.1 percent to a group within a single state. The largest increase, both here and in panel (b) of Figure 3, occurs between 1896 and 1898, when the mean group grows from 5.2 to 7.4 newspapers. In these years the old United Press failed, the modern Associated Press was organized in New York, and the Spanish–American War increased the demand for telegraphic news.

These two features do not rule out every channel other than a formal news agency. A newspaper could buy telegraph service and receive a special dispatch of its own, and the billing records discussed in Section 2.2 show that many did. However, a special dispatch is original copy. It would not appear word for word in seven other newspapers in six states on the same day unless those newspapers had bought a shared report. A shared report is what I refer to as news-agency access, whether or not the seller was a formal association. Distance and simultaneity do rule out the channel that would

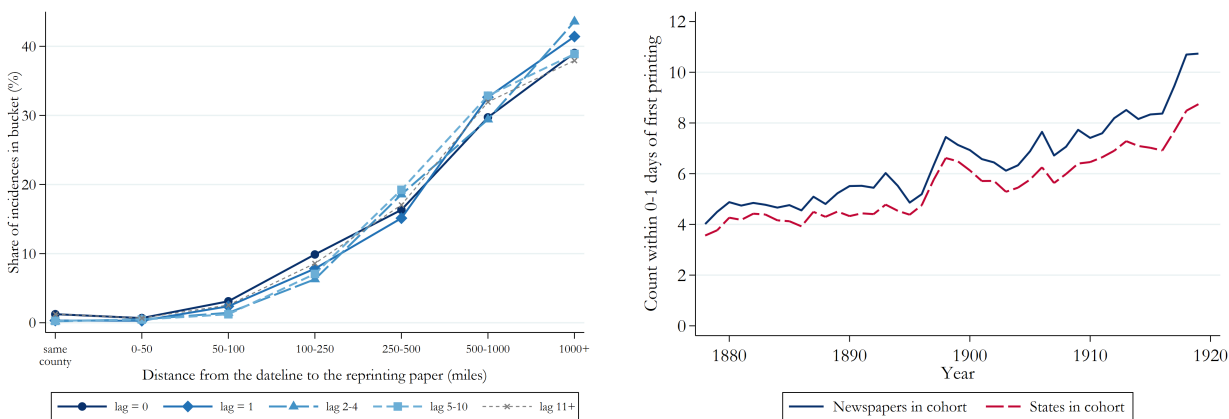
³⁴I can measure distance only for stories whose dateline is a county in the United States, which are 61.5 percent of the stories in *Chronicling America*. Almost all of the remaining stories have foreign datelines, such as London, Paris, Berlin, or St. Petersburg. Including them would shift the distribution toward longer distances. In addition, [Silcock et al. \(2024\)](#) already remove locally reproduced content with their classifier, so the small share of reprints in the dateline’s own county is not an independent test of local copying. The informative part of the figure is the rest of the distribution.

³⁵Comparing the total size of a cluster across lag groups would not be informative, because larger clusters are mechanically more likely to contain at least one late reprint. This probability rises from 45 percent among clusters of four or five newspapers to 88 percent among clusters of twenty-one to fifty and to 100 percent above fifty. As a result, slow incidences come disproportionately from large clusters (16.5 newspapers on average at lags of eleven days or more, against 8.6 at a lag of zero). The two measures that I use do not have this problem. Every cluster contributes to every lag group in panel (b) of Figure 3, and the same-day group in panel (b) of Figure B2 is a characteristic of the cluster as a whole.

be most problematic for the measure, which is newspapers redistributing one another’s copy.

Figure B2: Distance and simultaneity of fast reprints, 1878–1919

- (a) Distance from the dateline to the reprinting newspaper (b) The same-day group of newspapers and the states it spans



Notes: This figure shows where fast reprints come from and how many newspapers print them at once. Panel (a) plots the share of incidences at each distance from the dateline to the reprinting newspaper, by reprint-lag bucket. Distance is the great-circle distance between 1890 county centroids, in miles, so a reprint within the dateline’s own county is zero and falls in the “same county” bin. The sample is the 1.5 million incidences (55.5 percent of all *Chronicling America* incidences) whose cluster has a dateline resolved to a United States county and whose newspaper has a coverage county in the distance table; the excluded clusters are overwhelmingly foreign-datelined. Newspapers with more than one Library of Congress coverage county (4.0 percent of titles) are scored at their minimum distance, which biases the panel toward the short-distance bins. Panel (b) plots, for the 2.24 million incidences at lags of zero or one day, the mean number of newspapers in the cluster’s same-day group, the newspapers whose own lag is zero or one day, and the mean number of distinct states where they are located, weighted by incidence so that each series describes the typical fast reprint. The count is higher than the 5.47 same-day and next-day newspapers per story of Figure 3 because a fast reprint is more likely to belong to a cluster with many fast reprints. Each newspaper is assigned one home county. Fast reprints that are hundreds of miles from their dateline and shared at once by newspapers in many states are consistent with a report distributed by an agency rather than copied between newspapers.

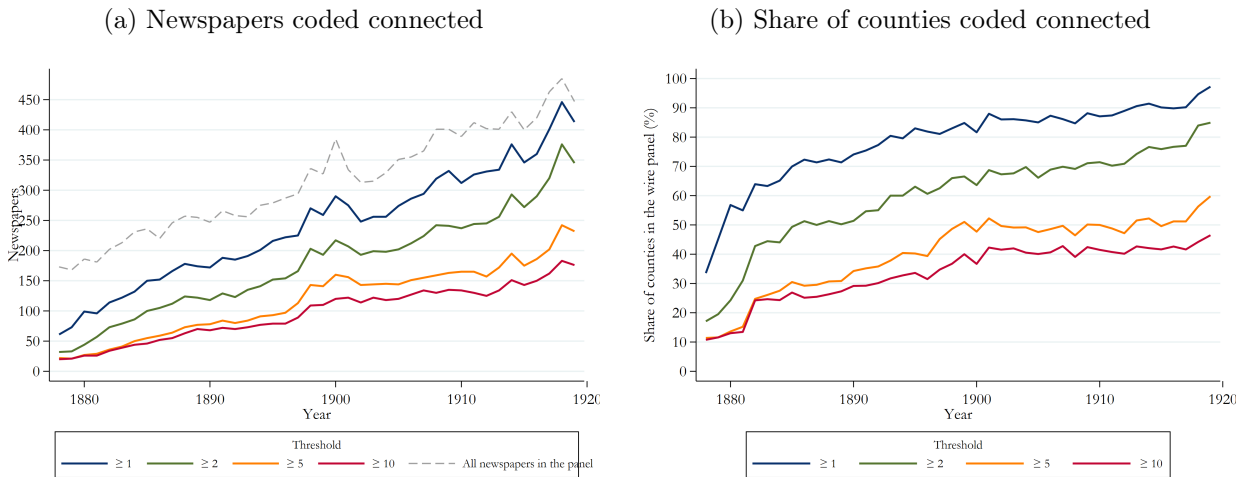
B.3 Stricter Connection Thresholds

A single fast reprint is thin evidence of agency access. It could be one lucky word-for-word match or one article that a newspaper happened to buy, and the definition of Section 3.2 codes a county as connected from the first such reprint onward. In the newspaper-year panel behind the measure, a newspaper with a fast reprint in one year often has none the next. Newspapers that lapse had a median of one qualifying reprint in the year before the lapse, against 29 for those that continue, and newspapers whose best year reaches fifty reprints have one in 92 percent of the digitized years of their span, against 72 percent for those whose best year reaches two to four. The absorbing coding turns a single such reprint into permanent connection, so the marginal connections are where the measure is weakest. In this section I use stricter definitions to test whether the results depend on them. A newspaper counts as connected only after it prints at least two, five, or ten fast non-political articles in a year. The rule that aggregates newspapers to counties, the absorbing state, and all later steps are unchanged.

Figure B3 shows that the thresholds change the measure substantially. By 1916, roughly 90 percent of the counties observed in the reprint data are connected under the baseline definition,

compared with 76 percent at two articles, 51 percent at five, and 43 percent at ten. Because a county crosses each threshold in a different year, the stricter definitions also change the dates of connection. I re-estimate the two main specifications of the paper under each definition. Appendix D.3 reports the coverage regression (Figure D5), and Appendix E.3 reports the media reach regression (Figure E1).

Figure B3: Connection under stricter reprint thresholds, 1878–1919



Notes: Under threshold k , a newspaper is coded connected in year t if it printed at least k non-political wire articles with a reprint lag of at most one day in t or any earlier year, and a county is connected if any of its newspapers is. $k = 1$ is the paper's baseline definition. Panel (a) counts newspapers so coded, against all newspapers appearing in the Newswire data that year (dashed); panel (b) reports the share of counties with at least one Newswire newspaper that are coded connected. Chronicling America era only (1878–1919).

C Measuring Speeches, Coverage, and Content

This appendix documents how the speech, coverage, and content measures of Section 3.3 are built and how well they perform. Appendix C.1 reproduces the labeling instructions and describes the run, and Appendix C.2 defines a speech and assigns each one to a county. Appendix C.3 validates the labels against the official record of Bryan’s 1896 campaign, against the counts of Pino and Salisbury (2025), through an article-level review of every disagreement, and on transcribed speeches. Appendix C.4 maps that campaign, and Appendix C.5 documents the content labels behind the results of Section 5.2.

C.1 Article Frame, Prompt, and API Implementation

I labeled all articles in August 2026 with OpenAI’s `gpt-5.6-terra` model through the Batch API, using the prompt reproduced below. The sample consists of the articles that mention a presidential candidate and were published between July 1 and November 15 of an election year. It contains 123,348 article–candidate observations over the eleven election years from 1880 to 1920. Of these, 118,147 correspond to the ten campaigns from 1880 to 1916 that I use in the analysis. I also labeled 16,291 observations from 1872 and 1876 in order to compare my measure with the data of Pino and Salisbury (2025) over their full period. In total, I labeled 139,639 observations across thirteen election years. In Appendix G I show results under GPT-4o.

Using the API is currently the only feasible way to label hundreds of thousands of articles with an LLM. It also ensures that every article is labeled under the same instructions and conditions. Each article is submitted as a separate request, so the answer to one article cannot influence the answer to the next, and the quality of the answers does not deteriorate over a long session.

Each request has two parts. The first part is a fixed set of labeling instructions, which is identical across requests. The second part contains the line **Candidate:** `<full name>`, followed by the publication date and the text of the article. I set the model’s reasoning effort to `low`, which was the level that produced the most accurate answers in a small pilot. The model must answer in a fixed format (a strict JSON schema), so it cannot omit a field or return an unexpected one. Following Korinek (2025), the instructions ask the model to explain its reasoning *before* giving each answer (the `speech_why` and `sentiment_why` fields below). The instructions end with five examples about a fictional candidate, “John Ward,” so that no example refers to a real campaign. The examples distinguish a speech by the candidate from the four cases that are most easily confused with it. These are a rally addressed by a surrogate, a speech that has been announced but has not yet happened, a past speech quoted as background, and a meeting held in the candidate’s honor at which he does not speak.

The model classifies each speech event in an article along two dimensions. The first is who spoke, which can be the candidate himself, a surrogate speaking on his behalf, or another speaker. The second is the status of the event, which can be that it occurred, that it is scheduled for the future, or that the article refers back to a speech given earlier. The model returns one entry for each combination of speaker, place, and date. Each entry includes the town, the state, the date, and the crowd size when the article reports them, and the model returns an entry even when the location is unknown. The instructions tell the model to report the status and the dates as stated in

the article, and not to judge whether an event falls inside the campaign season. I apply these filters afterwards, as described in Appendix C.2. The model also classifies the sentiment of the article toward the candidate.

Labeling the eleven election years from 1880 to 1920 cost \$222.92, or \$0.0018 per article, and the two additional years cost \$31.46.

I reproduce the instructions below. For readability, I present the definitions, the output fields, and the examples as tables. The exact text submitted to the model will be available in a public repository.

The instructions open as follows:

You are a historian compiling coverage of U.S. presidential campaign speeches, public addresses, and rallies from historical newspaper articles. You will receive the name of a presidential candidate and the OCR text of one newspaper article. Classify how the article covers public political speaking by or on behalf of that candidate, and extract the speech events it mentions. Output a JSON object following the schema you were given.

The instructions then define the two classifications.

Speaker type	Definition
candidate	The named candidate personally. Match the name in any form the article uses—full name, surname, title plus surname (“Mr. Ward”, “Gov. Ward”), or an unambiguous description (“the Democratic nominee”). OCR errors may garble the name; use context to decide.
surrogate	Someone speaking for the candidate or his campaign—a running mate, party leaders, campaign orators.
other	Any other speaker, including opposing candidates and their supporters.

Speech status	Definition
occurred	The article reports the speech, address, or rally as delivered.
scheduled	The speech is announced, expected, or planned but has not yet happened at the time of writing (“will speak”, “is expected to address”, “has an appointment to speak”).
historical_reference	The article refers back to a past speech as background, quotation source, or reminiscence—it does not report the speech as news.
none	The article describes no public speech, address, or rally.

The instructions then list four decision rules, which I reproduce verbatim:

- 1. Attribution is the critical judgment. Report speaker “candidate” only if the text supports that the candidate personally delivered or will deliver the speech. A rally, ratification meeting, or celebration where OTHERS speak about the candidate or on his behalf is NOT a candidate speech, even when the candidate is named, quoted from another occasion, honored, or nominated there. The candidate merely being present while others speak does not make him the speaker.*

2. A quotation, letter, written statement, or interview is not a speech unless the article ties it to a public speaking event.
3. Do not judge whether a speech falls inside the campaign season; report the status and any dates as stated, and let downstream analysis filter by date.
4. If the OCR is too garbled to support a judgment, use `speech_status` “none” and explain in `speech_why`.

The model must then fill in the following fields for every article:

Variable	Explanation
<code>speech_why</code>	One or two sentences naming who speaks (or why nobody does) and justifying the status.
<code>speech_status</code>	<code>occurred</code> , <code>scheduled</code> , <code>historical_reference</code> , or <code>none</code> .
<code>speaker</code>	<code>candidate</code> , <code>surrogate</code> , <code>other</code> , or <code>none</code> ; <code>none</code> if and only if <code>speech_status</code> is <code>none</code> .
<code>events</code>	List of the distinct speech references in the article, by any speaker with any status—one entry per combination of speaker type, place, and date. Each entry carries <code>speaker</code> , <code>status</code> , <code>town</code> , <code>state</code> (2-letter abbreviation, when the article states it or it is unambiguous), <code>date</code> (Y/M/D as stated or clearly inferable), and <code>crowd</code> (integer, or null if not stated). Empty strings are used for text fields the article does not support; the model is told not to guess.
<code>sentiment_why</code>	One-sentence justification of the article’s sentiment towards the candidate.
<code>sentiment</code>	Sentiment of the article specifically towards the candidate (positive, negative, neutral).

The article-level status and speaker must equal the highest-priority entry in the list of events. The candidate has priority over a surrogate, and a surrogate over other speakers. For the same speaker, `occurred` has priority over `scheduled`, and `scheduled` over `historical_reference`. The answer format cannot enforce this rule, so I check it afterwards. The two disagree in 170 of the 123,348 observations from 1880 to 1920 (0.138 percent). These cases do not affect the analysis, because I build all measures from the list of events.

The instructions end with five examples, each consisting of a short passage and the expected answer. A dash (—) indicates a field that the article does not support.

Example 1. Context: “Senator Poe addressed five thousand at the court house square in Peoria last night, urging the election of John Ward.”

Variable	Response
<code>speech_why</code>	A surrogate, Senator Poe, gave the speech urging Ward’s election; Ward himself did not speak.
<code>speech_status</code>	<code>occurred</code>
<code>speaker</code>	<code>surrogate</code>
<code>events</code>	<code>surrogate</code> ; <code>occurred</code> ; Peoria, IL; —; 5000
<code>sentiment_why</code>	The article reports a rally for Ward without praising or criticizing him.
<code>sentiment</code>	<code>neutral</code>

Example 2. Context: “Hon. John Ward will speak at the fair grounds Saturday afternoon.”

Variable	Response
speech_why	Ward’s own speech is announced but has not yet taken place.
speech_status	scheduled
speaker	candidate
events	candidate; scheduled; — ; — ;—
sentiment_why	A bare announcement with no evaluation of Ward.
sentiment	neutral

Example 3. Context: “In his celebrated Springfield address of six years ago, John Ward declared for sound money.”

Variable	Response
speech_why	The article recalls a past Ward speech as background rather than reporting it as news.
speech_status	historical_reference
speaker	candidate
events	candidate; historical_reference; Springfield; — ;—
sentiment_why	Calling the address ‘celebrated’ is favorable to Ward.
sentiment	positive

Example 4. Context: “John Ward spoke to 3,000 at Trenton yesterday and closed the day with a rally at Newark, where the Hon. Amos Reed followed him.”

Variable	Response
speech_why	Ward personally delivered speeches at Trenton and Newark, reported as delivered.
speech_status	occurred
speaker	candidate
events	candidate; occurred; Trenton, NJ; — ; 3000 candidate; occurred; Newark, NJ; — ;— surrogate; occurred; Newark, NJ; — ;—
sentiment_why	The itinerary is reported without evaluation.
sentiment	neutral

Example 5. Context: “The Democrats of Marion county ratified the nomination of John Ward last night; Judge Palmer and Gen. Rusk addressed the meeting, and a letter from Ward was read.”

Variable	Response
<code>speech_why</code>	A ratification meeting where surrogates spoke and Ward’s letter was read aloud; Ward himself did not speak.
<code>speech_status</code>	occurred
<code>speaker</code>	surrogate
<code>events</code>	surrogate; occurred; — ; — ;—
<code>sentiment_why</code>	The meeting celebrates Ward’s nomination.
<code>sentiment</code>	positive

C.2 From Labeled Events to the County Measure

Definition of a speech. The model classifies every event in an article, so I can define a speech in a more or less restrictive way. Throughout the paper, an article reports a speech if the speaker is the candidate and the event occurred. I exclude speeches by surrogates, events that are only announced, and references to a speech given earlier. This definition classifies 9,503 of the 123,348 article–candidate observations from 1880 to 1920 as speeches (7.7 percent). Including announced events would raise this number to 14,363 (11.6 percent), and also including references to past speeches by any speaker would raise it to 19,484 (15.8 percent). Appendix G compares this measure with the one that I obtain from an earlier labeling model.

Town-to-county assignment. The model returns a town and, in most cases, a state. I assign each speech to a county, which is the level of the other data sources. To do so, I use the Census Place Project (Berkes et al., 2023), which matches historical places to the counties that contained them, with one file per census decade from 1790 to 1940. I look up each speech in the decade closest to its year. For example, I look up a speech from 1896 in the 1900 file and a speech from 1912 in the 1910 file. There is no file for 1890, because the records of that census were destroyed in a fire, so I use the closer of 1880 and 1900.

Most places are easy to assign. If the town and the state appear in the file of that decade and correspond to exactly one county, I assign the speech to that county. Three quarters of all distinct place names are assigned in this way.

The remaining cases are of two types. First, some names appear in more than one county in their own decade. In most cases this happens because the file includes both towns and civil *townships*, and many townships share a name with a town elsewhere in the same state. For example, Hammond, Indiana, is a city, but Indiana also has townships named Hammond. In these cases I use the town. When the town entries for a name correspond to a single county, I assign that county and ignore the townships. Second, some names do not appear in their own decade at all. The coverage of the files is uneven, and large places such as Indianapolis and Kansas City are missing from some decades. In these cases I look for the town in the adjacent decade and then in more distant ones, again using only town entries. A match from a distant decade is weaker evidence, so I give it a lower confidence score, and I search distant decades only for names that are missing from their own.

I complement the Census Place Project with two lists that I built by hand. The first is a short list of the candidates’ home towns. It covers the front-porch campaigns, in which reporters often name the house or the town without a state. The second is a list of 200 places that I researched

one at a time. It includes county seats that stand for a county named without a town, historical names of cities, places that are missing from the Census Place Project, and specific event locations. Each entry has its own confidence score and, where relevant, the candidate and the years to which it applies. I classify 181 of the 200 entries as high confidence. Four of the entries come from the review of Bryan’s 1896 campaign in Appendix C.3, which identified towns that the files did not include or included only as townships in the wrong county.

When a name matches several counties and none of these rules applies, I drop the speech and do not assign it to the most populous county. Overall, I assign a county to 8,520 of the 10,815 distinct combinations of town, state, and year (79 percent). The share is much higher across article mentions. Of the 12,254 mentions of a speech with a location, 449 cannot be assigned, so the assignment rate is 96 percent. The names that I cannot assign are rare. They are mostly misspellings created when the text was digitized, very small places, and one-time errors, while the names that appear in many articles are assigned.

A larger loss occurs before this step. Among the articles that report a delivered speech by the candidate, 80 percent end up with a county. Most of the remaining articles report a speech without saying where it happened (for example, “he stumped Nebraska last week”). In these cases the model correctly returns an event without a town. This explains why the share is low in 1888 and 1904, at 37 and 66 percent, under both labeling models. Newspaper coverage in these two campaigns is the thinnest and the least specific about places.

C.3 Validation against the Historical Record

I validate the labels in four ways. First, I compare them with an independent record of one campaign. Second, I compare them with the data that Pino and Salisbury (2025) collected for the same campaigns. Third, I review every county in which the labels and the record of that campaign disagree. Fourth, I run the prompt on transcribed speeches, which do not depend on the digitized newspapers.

C.3.1 Comparison with the Official Record of Bryan’s 1896 Campaign

The only campaign in the sample period with an independent record of campaign stops is Bryan’s 1896 tour. The *Railroads and the Making of Modern America* project reconstructed it from Bryan’s own memoir and the *Omaha World-Herald* (Sanderson, 2008; Thomas et al., 2008).³⁶ The first column of Table C1 compares my county-level labels with this record for the 2,828 counties in 1896. Precision is 0.88 and recall is 0.86, and the F1 score, their harmonic mean, is 0.87. These rates are high relative to other work on the same newspaper collection. Silcock et al. (2024) report an F1 score of 0.85 for their politics topic classifier and an accuracy of 94.9 percent for their procedure to assign locations. Their figures are not directly comparable to mine, because each of them measures a single step. My rates combine three steps, which are detecting a speech, extracting its location, and assigning it to a county. LLMs have also been shown to match or outperform human annotators in political text-annotation tasks (Gilardi et al., 2023).

³⁶Data available at https://railroads.unl.edu/student_projects/nsanderson/WJB_RRpage.html (Last accessed on August 13, 2026).

The official record was assembled from one newspaper and the candidate’s memoir, so it can miss stops that other newspapers reported. I therefore review every county in which the labels and the record disagree (Appendix C.3.3), and I add to the record the counties that the review confirms as Bryan stops. Of the forty-seven counties that the labels classify as visited and the record does not, nineteen are confirmed as stops that the record misses.³⁷ The second column of Table C1 compares the labels with the corrected record. Precision increases to 0.93 and the F1 score to 0.90. Every false positive was reviewed, so these figures are not lower bounds.

Table C1: Label Quality for Bryan’s 1896 Campaign, Official and Corrected Records

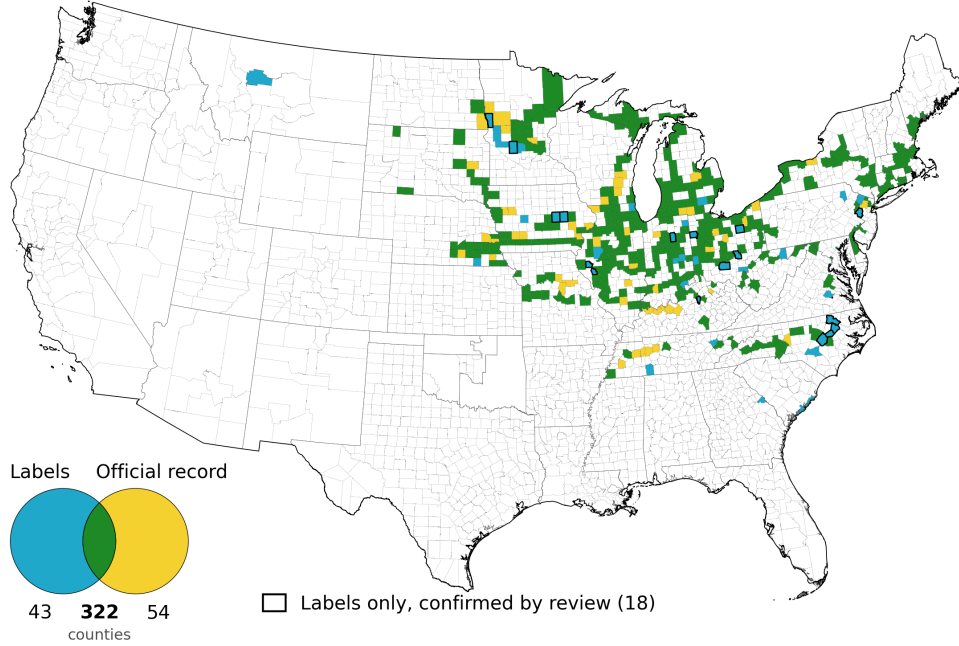
	Official record	Audit-corrected
Counties with a recorded visit	380	399
True positives (TP)	325	344
False negatives (FN)	55	55
False positives (FP)	44	25
True negatives (TN)	2404	2404
Recall	0.86	0.86
Precision	0.88	0.93
F1 score	0.87	0.90
Accuracy	0.96	0.97

Notes: County-level comparison of the speech labels with the official record of Bryan’s 1896 campaign stops (first column) and with the record corrected by the review of Appendix C.3.3, which adds the nineteen counties confirmed as Bryan stops (second column), over the 2,828 counties of the 1896 cross-section. Recall is $TP/(TP+FN)$, precision is $TP/(TP+FP)$, and the F1 score is their harmonic mean; accuracy, $(TP+TN)/N$, is inflated by the many never-visited counties and is reported only for completeness. One of the nineteen counties is no longer detected by the current labels, so it enters the corrected comparison as a false negative. Every false positive was reviewed, so corrected precision is not a lower bound.

Figure C1 maps the comparison. Most of the counties in either source are in both. The counties in the record only include two strings of stops in Kentucky and Tennessee.

³⁷I conducted the review on an earlier version of the county assignment, which had forty-seven false positives and sixty-one false negatives. Four findings of the review led to corrections in the town-to-county assignment (Washington Court House and Shelby Junction, Ohio; Waverly, Minnesota; and Carroll, Missouri). After these corrections and a revision of the geocoding of the official record, which now contains 380 counties, the tables report forty-four false positives and fifty-five false negatives. The counts in the text refer to the counties that were reviewed.

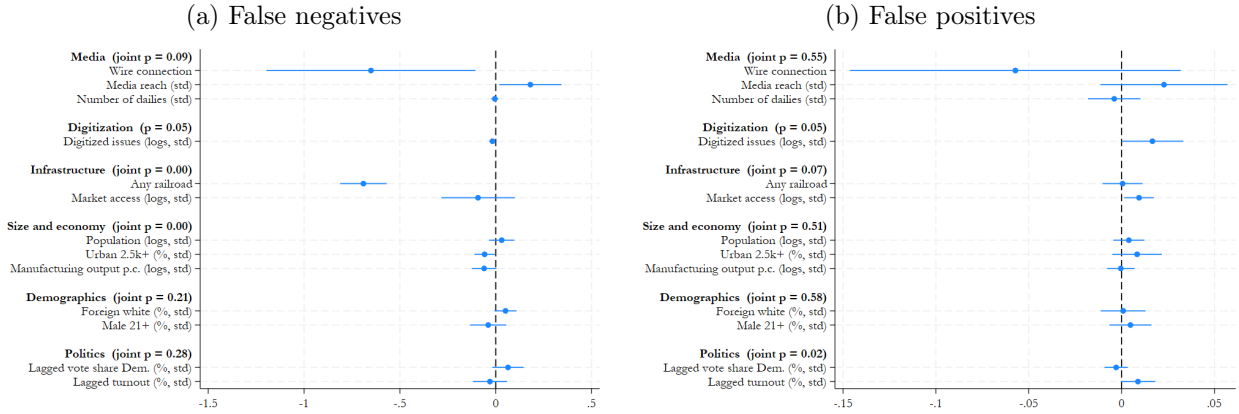
Figure C1: Speech Labels and the Official Record by County, Bryan 1896



Notes: This figure compares the counties where the LLM labels detect at least one public speech by Bryan during the 1896 campaign with the official record of his stops (Sanderson, 2008; Thomas et al., 2008). Green counties are in both (322), cyan counties are in the labels only (43), and yellow counties are in the record only (54). Black outlines mark the 18 labels-only counties that the review of Appendix C.3.3 confirms as Bryan stops the record misses. Four counties of the labels and four of the record have no place on the 1890 county boundaries used here. One of them is Emporia, Virginia, the nineteenth confirmed stop.

Finally, I study which counties are misclassified. Figure C2 shows regressions of an indicator for disagreement with the official record on county characteristics, separately for the counties the record lists as visited, where the error is a missed stop, and for the counties it does not, where the error is a flagged stop. Misclassification is mainly related to how visible a county is in the digitized newspapers. False negatives are more common in counties without a railroad, with few digitized local newspapers, and without a news-agency connection. False positives increase with the amount of digitized text. These characteristics are either controls in the main regressions or, in the case of digitization, almost constant within counties and absorbed by the county fixed effects. The news-agency connection is a control in the main regressions and is not the coefficient of interest. Media reach also predicts false negatives. If anything, missed speeches are more common in counties with higher media reach, so this measurement error should attenuate the estimated effect of media reach on speech locations. Part of this correlation may come from the record itself. The review below finds that several of the recorded stops that the labels miss were appearances from the rear platform of the train at which Bryan did not speak. These stops were likely more common along the main railroad lines, where media reach is highest. Lagged electoral outcomes do not predict false negatives. False positives rise with lagged turnout, by about one percentage point per standard deviation, so I control for lagged turnout in the main regressions. Pooling the two outcomes, or scoring them against the corrected record, gives the same picture.

Figure C2: Predictors of False Negatives and False Positives, Bryan 1896



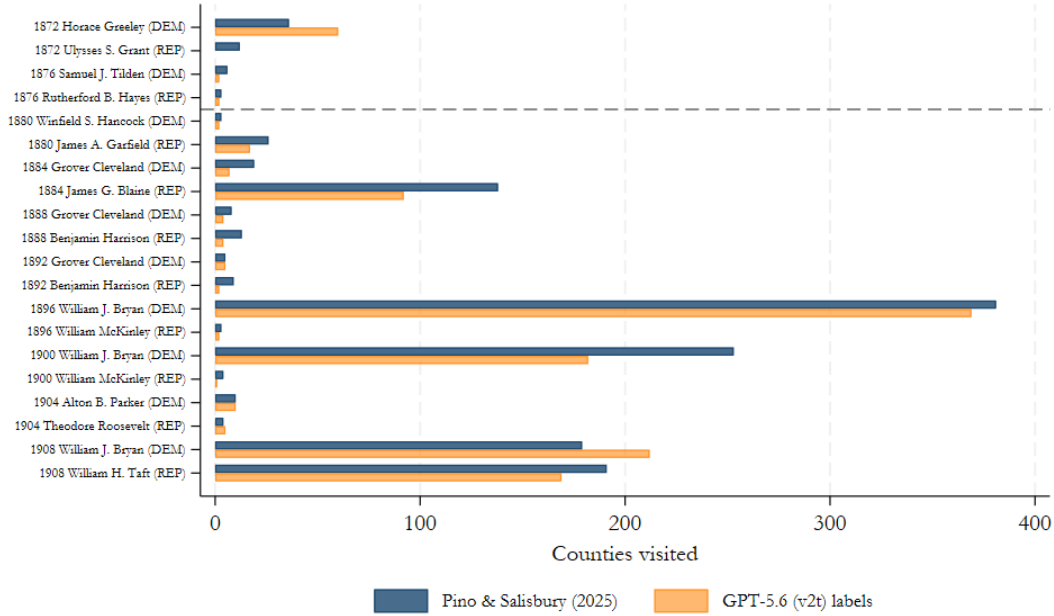
Notes: Each panel plots the coefficients of a linear probability regression of an indicator for disagreement with the official record of Bryan's 1896 stops on the county observables listed on the vertical axis, grouped by variable family; family headings report the p-value of a joint significance test of the family. Panel (a) restricts the sample to counties with a recorded visit, and the outcome is the label missing it. Panel (b) restricts the sample to counties without a recorded visit, and the outcome is the label flagging one. To limit collinearity, the specification keeps one measure per construct; continuous regressors are standardized over the 1896 sample, while wire connection and any railroad are indicators. Horizontal lines are 95% confidence intervals from heteroskedasticity-robust standard errors. Pooling the two samples into one regression, or scoring disagreement against the corrected record of Table C1, gives the same picture. Missed stops concentrate where a county is hard to see in the digitized press, which is consistent with the attenuation argument of Section 3.3.

C.3.2 Comparison with Pino and Salisbury (2025)

Pino and Salisbury (2025) reconstruct presidential campaign trips from newspaper clippings between 1872 and 1908. Figure C3 compares the number of counties visited by each candidate in each year in my data with the counts in their Table 1 (page 57). My labels start in 1872, so the comparison covers their full period, with ten election years and twenty major-party candidates. The correlation between the two measures is 0.98 in levels and 0.93 in logs. For the period 1880–1908, the correlations are 0.98 and 0.97. My measure classifies *fewer* counties as visited than theirs for fifteen of the twenty candidates, the same number for two, and more for three. For Bryan in 1896, they report 381 counties and I find 369.

For Greeley in 1872, I find 60 counties and they report 36. Greeley was the editor of the *New-York Tribune* and appears very often in the press, which makes his speeches the hardest to identify. For Grant in that same year, I find no counties while they report twelve. Grant did not campaign in 1872, and none of the 519 articles that mention him in my sample reports a speech that he delivered.

Figure C3: Counties Visited per Candidate: Comparison with Pino and Salisbury (2025)



Notes: Navy bars are transcribed from Table 1 of [Pino and Salisbury \(2025\)](#); orange bars count the counties with a labeled speech for the corresponding candidate and year. Third-party candidates are excluded. The window is 1872–1908, the full span of their table. The dashed line separates the two pre-1880 election years, which are labeled but lie outside the 1880–1916 window used elsewhere in the paper. The correlation between the two series is 0.980 in levels and 0.935 in logs, against 0.968 and 0.870 under the superseded GPT-4o labels.

C.3.3 Review of the Disagreements

The official record is incomplete, so treating it as the truth understates the quality of the labels. I therefore reviewed, at the article level, every county in which the labels and the record disagree. For each of the sixty-two counties reviewed, I assembled the full text of every article behind the label, with its newspaper and publication date, the location, date, and reasoning returned by the model, a link to the scanned page in *Chronicling America*, and Bryan’s documented itinerary within three days of the event. A second LLM from a different developer (Claude, by Anthropic) reviewed each county separately and without access to the other results. The reviewer followed strict rules. A speech requires that Bryan himself addressed an audience in person in that county, so quotes, letters, surrogate speakers, and meetings about the candidate do not count. The reviewer also had to compare the extracted town name with towns of the same name in other states. For another forty-six counties, I use the conclusions of an earlier review of the GPT-4o labels, in the cases where the conclusion depends on the digitized collection and not on a particular article.

Every conclusion was then challenged in a second review, which tried to overturn it. The second review also covered the conclusions that favor the labels, such as the conclusion that the labels were right to exclude a county, and it rejected nine of the twenty-one conclusions of this type. In five of the nine cases it accepted the facts but classified the article as an announcement, in two cases it could not reach a conclusion, and in two cases it argued that the labels may have missed a speech. Challenging only the conclusions that go against the labels would have overstated their quality.

False positives. Of the forty-seven counties that the labels classify as visited and the record does not, the first review concludes that twenty-seven are stops that the record misses, and nineteen of these survive the second review. Another fifteen are real speeches assigned to the wrong county, and five are not speeches by Bryan. Therefore, 40 percent of the apparent false positives are stops that are missing from the record. Most of the remaining errors are speeches assigned to the wrong county. The main remaining source of error is therefore the town-to-county assignment of Appendix C.2, and not the prompt. In three of the eight cases that the second review rejected, the record does contain the stop but codes it incorrectly. Waukegan, Illinois, is entered as a town in Wisconsin, and Sullivan, Bethany, and Newcastle are listed without a county. In these cases the speech took place in the county that the labels identify, and only the claim that the record misses the stop is wrong. I do not add these three counties to the corrected record in Table C1, so the corrected precision is conservative.

False negatives. Of the sixty-one counties that are in the record and that the labels do not classify as visited, thirty-two have no article about the stop in the digitized newspapers. These cases reflect the coverage of the digitized collection and not the labeling. For the remaining twenty-nine, the review asked whether the evidence shows that Bryan himself spoke in the county in 1896. The answer is yes for eight counties, no for eleven, and undetermined for ten. Roughly a third of the apparent false negatives are therefore missed speeches. The rest are stops at which Bryan did not speak, such as a handshake from the rear platform or a train passing through, or stops for which the surviving newspapers only carry the announcement. Of the eight missed speeches, four were errors in the town-to-county assignment, which I have since corrected.

The review could also conclude that the speaker was someone other than the candidate. It did not reach this conclusion in any of the twenty-nine counties. Surrogate speakers are common in the data, and they account for 13.5 percent of the articles that the GPT-4o labels classify as speeches and my definition excludes (Table G1). However, they do not explain the false negatives.

Examples. Figure 5 in the main text shows the newspaper page behind the label of three counties. The first is a stop on which the record and the labels agree. The second is a mention that the labels correctly exclude. The third is a county that the labels classify as visited and that the review confirms as a stop missing from the record. Figure C4 shows six more pages. The first row shows three correct labels on difficult pages, which are a stop recovered from a long list of towns and two events that are reported but are not a delivered speech by the candidate. The second row shows one example of each type of error, which are a surrogate mistaken for the candidate, a real speech assigned to the wrong county, and a stop whose town the model did not extract.

Figure C4: Further Examples of Agreement and Disagreement with the Official Record

Brass bands, young women in costumes of white and silver, marching clubs and a liberal display of Bryan badges all had their part in the demonstrations. . . . Twenty-one speeches, including three at Ottumwa and four here to-night, constituted the day's record. The others were made at Mount Pleasant, Fairfield, Batavia, Albia, Chariton, Osceola, Creston, Corning, Villisca, Red Oak, Hastings, Malvern, Glenwood and Pacific Junction. Mr. Bryan's second address at Ottumwa was the most important of the day. It was devoted to summing up the campaign and caused the crowd to become wild with enthusiasm.

(a) *True positive, from a run-on list.* The *Morning News* (Savannah), Nov. 1, lists Bryan's twenty-one Iowa speeches of Oct. 31. The OCR fuses Albia to its neighbors; it is extracted anyway and Monroe County, IA is flagged.

Judge Doster has spoken at Atchison, Topeka and elsewhere. The other candidates on the State ticket have not yet engaged actively in the campaign. David Overmyer has held rousing meetings at Emporia, Topeka, Independence, Neosho Falls, Burlington, Chanute and Iola. Senator Martin also has made a few speeches. Webb McNeill, E. C. Little and R. W. Turner, representing the silver Republican element, are on the stump most of the time. The biggest meeting of the campaign will be in Topeka, Saturday, October 3. William J. Bryan will be the speaker. All the railroads will run excursion trains and

(b) *True negative: an announcement.* The *People's Voice* (Wellington, Kan.), Sept. 17, on a Topeka meeting three weeks away at which "William J. Bryan will be the speaker." The model classifies the event as *scheduled*, so Shawnee County is not flagged.

perity. "We have down here a giant of our own whom we all delight to honor—a gentleman whom the Times and others of our friends have recently labeled 'anarchist.' It is such anarchists as he who have been doing so much in California, Wisconsin, Michigan, Indiana and Illinois to make William J. Bryan next president of these United States. "It gives me great pleasure, and it is an honor of which I am proud, to present to this large and enthusiastic audience as the orator of the evening one who needs no formal introduction, surrounded by his neighbors in the house of his friends. This, ladies and gentlemen, is the Little Giant of California, our own and only "Steve," United States Senator Stephen M. White." "LITTLE GIANT OF THE PACIFIC." And then the audience broke forth in one mighty acclaim. The applause was spontaneous, earnest and enthusiastic. When it ceased Senator White began a discourse which lasted for several hours. Once again in this, his home city, the "Little Giant of the Pacific," as Mr. Snyder aptly termed him, by the force of his logic, the power of his oratory, swayed thousands. In that plain, unanswerable, yet beautifully phrased language of his, the senator explained why William J. Bryan should be the next president of the United States, and why it was the duty of every loyal American citizen to vote for him.

(c) *True negative: a surrogate speaker.* The *Los Angeles Herald*, Nov. 1, on a rally for Bryan where Senator White was "the orator of the evening." The model classifies the speaker as a *surrogate*, so Los Angeles County is not flagged.

Simply because the election is due here in 30 days, and if it hadn't been for that election they would have had a bond issue before that time.

Why, my friends, if the nomination of William Jennings Bryan upon a free coinage platform will have the effect of pouring thirty or forty millions of gold into the treasury in 30 days, what would his triumphant election do in this country? (Loud and enthusiastic applause.) They tell us that it is over production that causes low prices. Why, my friends, they tell us that the fall in the price in wool in this country is caused by over production, and the loss of the tariff of McKinley; and my friends who are engaged in the wool industry who have lost so much money upon it in the past few years, I want to ask you if you ever stopped to think whether the man who is raising horses has not lost as much money by over-production of horses, and your neighbor who owned a house and lot three or four years ago worth \$5,000 and today couldn't sell it for \$1,500, ask him if he has not lost \$3,500 by over-production of houses and lots. (Applause.) And the merchant who

(d) *False positive: a surrogate read as the candidate.* The *Great Falls Weekly Tribune*, Oct. 2, prints a silver speech whose orator calls Bryan "he" throughout. Bryan never entered Montana; Cascade County is flagged in error.

not speak in their absence, besides he expected to have quite a large audience to talk to in New York, and he wanted to save some of his voice for that meeting. Then the crowd filed past him in usual good order, and all who cared to secure a grasp of his hand. I want to shake, said one man in a loud voice. "Well," said Mr. Bryan, as he reached for him, "don't shake me on election day."

Hurrah for the next President of the United States, came a voice from the crowd.

"That is sufficiently indefinite," added Mr. Bryan, to permit all to join in three cheers were given. As the train drew away from the station, someone shouted, "Put Old Allen County down for 5,000 plurality for Bryan."

Stops and short addresses were made by Mr. Bryan at Washington, Dixon, Monroeville, Conway, Vanwert, Middleport, Delphos, Lima, Lafayette, Ada, Crestline and Orrville.

(e) *False positive: a real speech, the wrong county.* The *Wheeling Register*, Aug. 11. Between Van Wert and Delphos the stop is Middle Point, which the digitized text renders as "Middleport", a town 200 miles away. The speech took place, but it is assigned to Meigs County.

Bryan at Mascoutah, Ill., had been advertised for 7:45, but the train reached there at 7:20, and Mr. Bryan made a few remarks to a small crowd of undemonstrative people. They cheered as the train drew out.

Mount Vernon, Ill., Sept. 14.—When the Bryan special train reached Nashville, Ill., the democratic candidate found several hundred people there. They cheered him when he came on the car platform and applauded his remarks about the money question. Ashley, the next stop, had a large and enthusiastic crowd. Mr. Bryan again spoke.

Mount Vernon had sent a committee ahead to meet Mr. Bryan, and when he reached here at 9:20 o'clock, they formed in line and escorted him to a stand erected in front of the state supreme court building, a hundred yards from the railway station.

It was found that Mount Vernon had prepared to give Mr. Bryan a rousing reception. A Bryan and Sewall club, composed of young men and young women, the former in white duck uniforms and the girls in white muslin, had formed a double line extending from the station toward the court house, and through this Mr. Bryan passed, while the crowd cheered loudly. It was a very good sized crowd, numbering several thousand people, and they had plenty of enthusiasm. A brass band marched ahead of Mr.

(f) *False negative: a town never extracted.* The *Morning News* (Savannah), Sept. 15, with a Mount Vernon, Ill., dateline, has Bryan escorted to a stand at the state supreme court building, which is the venue in the official record. The model extracts Nashville and Ashley from the same paragraph but not Mount Vernon, so Jefferson County is missed.

C.3.4 Detection on Transcribed Speeches

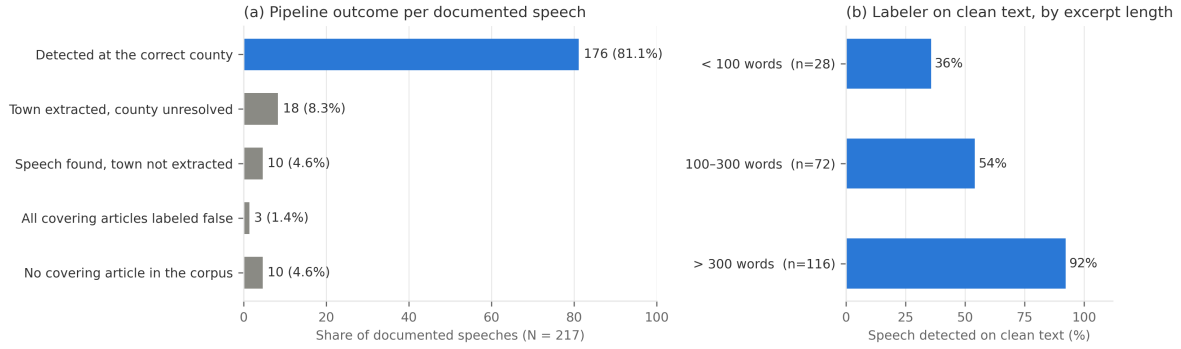
All the previous checks depend on the digitized newspapers, so they combine the quality of the labeling with the coverage of the collection. To separate the two, I use the 217 speeches from Bryan’s 1896 campaign that the *Railroads* project transcribed (Sanderson, 2008; Thomas et al., 2008). None of the newspapers that the project cites as sources is digitized in Chronicling America, including the *Omaha World-Herald*, so I cannot follow any of these documents through its own source.

First, I ask whether my procedure finds these 217 speeches in other newspapers. Panel (a) of Figure C5 shows the results. I detect 176 of the 217 speeches and assign them to the correct county. Another 18 are detected but assigned to a different county, and 10 are detected without a town. For 3 speeches, all the articles about the stop are labeled as not reporting a speech, and for 10 speeches there is no article in the collection within three days of the visit. Most of the coverage of the tour comes from newspapers elsewhere, since only 21 percent of Bryan’s recorded stops are in counties with any digitized issue from 1896. Recall is high because many newspapers covered each stop. The main limitations are therefore digitization and the short articles about small stops, and not the model’s ability to recognize a speech.

Second, I submit the 217 transcripts themselves to the labeling prompt. All of them are speeches and none has errors from digitization, so this exercise isolates the detection step. It is also the one exercise in which the labels perform worse than the earlier GPT-4o labels (Appendix G). The prompt classifies 72.4 percent of the transcripts as delivered speeches by the candidate, and the difference with the GPT-4o prompt depends on the length of the text. Both prompts detect 35.7 percent of the excerpts under 100 words. Between 100 and 300 words, the current prompt detects 54.2 percent and the GPT-4o prompt 83.3 percent. Above 300 words, they detect 92.2 and 100 percent. The definition of a speech does not explain the difference. Including announced events and references to past speeches leaves the detection rate at 72.4 percent, so the model finds no speech event at all in the sixty remaining texts.

An important caveat is that these texts are *transcripts* of speeches and not newspaper articles that report a speech. The prompt asks the model to describe how an article *covers* public speaking and to return events with a speaker, a status, a town, and a date. A transcript has no dateline and no reporting context, so the prompt was not designed for this type of text. The difference is largest for texts of intermediate length, which have some narrative but not much, and this pattern is consistent with that explanation.

Figure C5: Tracing the 217 Documented Speeches of Bryan’s 1896 Campaign

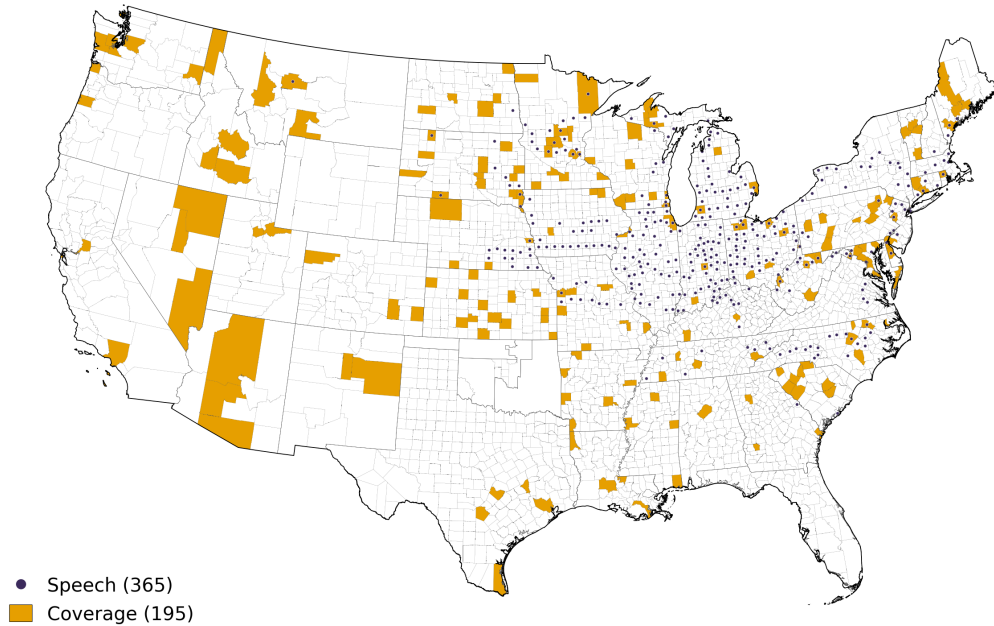


Notes: Panel (a) classifies each of the 217 speech documents hosted by the *Railroads* project (Sanderson, 2008; Thomas et al., 2008) by its outcome in my procedure. Of these, 176 are detected and assigned to the correct county, 18 are extracted but assigned elsewhere, 10 are detected without the town extracted, 3 have covering articles that are all labeled as non-speeches, and 10 have no covering article in the collection within three days of the visit. Panel (b) shows the share of the same 217 transcribed texts that the labeling prompt classifies as a delivered speech by the candidate, by length of the excerpt.

C.4 Maps

Figure C6 maps the counties in which the labels place a speech by Bryan during the 1896 campaign and the counties whose newspapers covered one.

Figure C6: Speeches and Coverage by County, Bryan 1896



Notes: This figure shows the counties where the LLM labels detect at least one public speech by Bryan during the 1896 campaign (dots, 365 counties) and the counties where newspapers circulated at least one article covering one of his speeches (amber, 204 counties). The two overlap in 43 counties. A further 4 speech counties and 11 coverage counties were created after 1890 and so have no place on this map.

C.5 Content Extraction, Topic Labels, and Values Labels

The content measures of Section 5 come from a second round of labeling, which I apply to the articles that report a delivered speech. It has two steps. The first step extracts the passages of an article that contain the content of the speech. The second step labels each passage without access to the rest of the article. I separate the two steps so that the label of a passage does not depend on where in the page it appears.

Extraction. In the first step, I use OpenAI’s `gpt-5.6-luna` model with reasoning effort `low` through the Batch API. The model returns the opening and closing words of each passage, and I then cut the passage from the digitized text at those words. Every passage is therefore a verbatim part of the original page, and the model cannot paraphrase it. I classify passages into three types. *Direct* passages are the candidate’s own words, quoted or printed at length. *Characterization* passages are the newspaper’s account of what the speech argued. The remaining passages are statements in the candidate’s voice that have no visible link to a public speech. The extraction finds 30,217 passages, of which 102 (0.3 percent) cannot be matched back to the digitized text. I group them into 18,078 units, one per article and type, of which 6,890 are direct. I only use direct units in the results of the paper. Characterization units describe the same speeches in the newspaper’s voice and not in the candidate’s, so I keep them separate.

Topic and scope labels. In the second step, the model only sees the extracted unit. It first decides whether the passage is sufficient to tell what the speech was about, which is the case for 86 percent of the units. It then assigns one primary topic out of 24 and indicates any other topic on which the passage has substantial content. The topics cover the policy agenda of the period, including tariff and trade, money and currency, banking, agriculture, labor and industry, trusts and regulation, taxation, foreign affairs and empire, immigration, race, prohibition, and internal improvements. Three topics have no policy content of their own, and I treat them as generic campaign material. These are the party record and attacks on the opponent, constitutional and governance questions, and ceremonial courtesy toward the audience. The model also labels three other dimensions of each passage, each as a separate question. The first is how specific to a place the content is (a locality, a state, a region, or no specific place). The second is the group that it addresses (the general public, an economic sector, a demographic or civic group, or a party). The third is the level of government of the policy (local, state, federal or national, foreign, or no policy). Table 3 uses the first and the third of these dimensions, and Table 4 uses the primary topic. I again use `gpt-5.6-luna` with reasoning effort `low` through the Batch API, at a cost of \$6.

From units to county-year shares. I assign each unit to the county where the speech was delivered, using the town-to-county assignment of Appendix C.2. Some articles report speeches in several counties, and a passage from such an article cannot be attributed to one of those counties. I therefore only use passages from articles that report a speech in exactly one county. The share of a label in a county-year is the percentage of its usable direct units with that label. Using shares removes the mechanical relationship between the number of units and the number of newspapers that covered the speech. The shares are based on roughly 2,500 units in 457 county-years. The

median county-year has two units, and 41 percent of them have only one. The precision of the content estimates is limited by this small number of units per county-year more than by the number of counties. Three topics (prohibition, Indian affairs, and territorial questions) appear in too few counties for inference with clustered standard errors, so I exclude them. For this reason, Table 3 adjusts for 21 topic shares and not 24.

Validation. On a set of 39 units that I labeled by hand, the model agrees with my label on whether the unit is usable in 95 percent of the cases and on the primary topic in 92 percent. An additional check uses reprints. When different newspapers reprint the same report, the two copies receive the same primary topic 86 percent of the time.

Values labels. A second prompt labels each unit according to the type of appeal that it makes and the community that it refers to as “we”. The type of appeal describes whether the argument is based on material interest, grievance, character, civic principle, or the party’s record. The community ranges from the locality to the nation. I use the same model and settings (`gpt-5.6-luna` with reasoning effort `low`), at a cost of \$5. For the analysis, I group the labels into six county-year shares along two dimensions. The *community* dimension distinguishes passages that refer to the nation in domestic terms, to the nation in relation to the world, and to a class or group. The *warrant* dimension distinguishes appeals based on material stakes, on identity or principle, and on the party’s record, and its three shares sum to one. I build the values labels for five campaigns (1896, 1900, 1908, 1912, and 1916) for which there were enough extracted speech passages. The sample is therefore smaller than in Table 3. It contains 365 county-years with at least one usable direct passage. Appendix F reports the estimates.

Reference-point outcomes. I build the four reference points of Table 4 as follows. I fixed the list of national issues from standard campaign histories before estimation. The issues are the tariff (1880, 1884, and 1888), the tariff and silver (1892), free silver (1896), imperialism and silver (1900), the trusts (1904), the trusts and banking (1908), the trusts and the tariff (1912), and the war and labor (1916). The national-conversation measure counts, for each topic and year, the newspaper pages in *Chronicling America* that match the search terms of the topic, as a share of all the pages that match any topic in that year. The measure behaves as expected, with peaks for money and currency in 1896 and for foreign affairs in 1900 and 1916. Only about eighteen of the twenty-four topics have search terms that are specific enough, so this outcome is defined for fewer reports than the others. For the platform measures, I use all 22 party platforms from 1880 to 1916 (the major-party platforms, the 1892 Populist platform, and the 1912 Progressive platform). I split them into 681 planks and label each plank with the same topic prompt as the speech content (Woolley and Peters, 2026). I also labeled the 1920 platforms (163 planks), but they are outside the estimation period and do not enter any estimate. The opponent measure averages the platforms of the rival parties in the same year, and the measure of uniformity across stops excludes the report’s own stop. I exclude generic campaign material (praise of the audience, discussion of the parties’ records, and ceremonial content) from the national-issue and platform measures. These outcomes are therefore defined only for reports with policy content, which explains why the number of observations differs across the columns of Table 4.

D Robustness and Timing of the Coverage Effect

The identifying assumption of equation (1) is about timing. The campaign in which both counties of a pair become connected must not coincide with shocks that are specific to that pair and that also affect their exchange of news. Section 4.1 argues that a county is likely to connect in order to exchange news with one specific place only when that place is a state capital or one of a few large cities, because a news-agency subscription was too costly to maintain for the news of any one ordinary county. Throughout this appendix, I refer to New York, Kings, Cook, and the District of Columbia as hub counties.

This appendix has four parts. Appendix D.1 lets the effect vary by campaign year and restricts the sample to the campaigns in which candidates toured. Appendix D.2 shows that the estimates are robust to pair-level trend controls, to restricting the sample to pairs that eventually connect, and to excluding capital and hub counties. Appendix D.3 re-estimates the coverage regression under stricter definitions of connection. Appendix D.4 presents leads and lags of the connection status and event studies around the first campaign in which a pair is connected. The anticipation that these tests reveal is concentrated in the pairs that include a capital or a hub county.

D.1 The Effect by Campaign Year and the Stumping-Campaign Sample

Four election years have very few detected speeches. In 1880, 1888, 1892, and 1904, speeches originate in fewer than thirty counties, compared with more than ninety in each of the other six campaigns. The first three are years of front-porch campaigns, and in 1904 neither Roosevelt nor Parker campaigned actively. The small number of speeches therefore reflects how candidates behaved and not a problem with the labels.

Since the paper asks whether candidates chose the location of their appearances to generate coverage, the campaigns in which candidates made appearances are the most relevant sample. In this section I re-estimate the coverage results of Section 5 on the six campaigns with at least fifty speech-origin counties. This threshold falls between the two groups of years, and it depends only on campaign activity and not on any estimate.

The restriction changes the sample very little, because the excluded years have few observations in the pair-level data. They contain five percent of the observations, and in column (4) of Table 1 the connection effect and the distance interaction both move by less than half a standard error.

I do not use the restricted sample as the main specification because connection is an absorbing state. With county-pair fixed effects, the effect is identified only from pairs that are observed both before and after they become connected. Excluding early years could turn pairs that were connected by 1896 into pairs that are always connected, which do not contribute to identification. Keeping 1884 in the sample prevents this problem. Of the 10,384 county pairs that change connection status over the ten campaigns, 95 percent still change status within the restricted sample. The restricted estimates are therefore identified from almost the same pairs as the full sample.

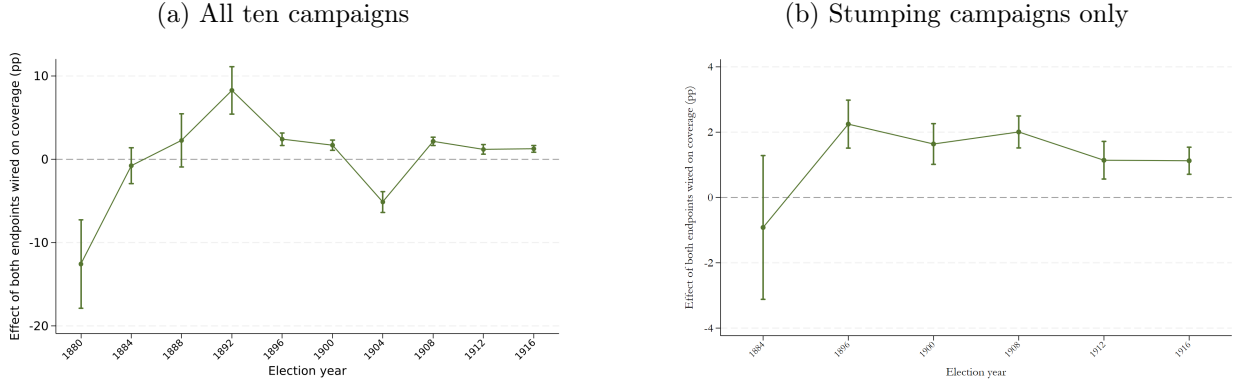
Equation (1) assumes that the coefficients are constant over time. Figures D1 to D3 come from

a single regression that interacts all the terms of equation (1) with campaign-year indicators,

$$\begin{aligned} \text{Coverage}_{s,d,t} = & \sum_{\tau} \psi_{1,\tau} \text{Connect}_{o,t} \cdot \text{Connect}_{d,t} \cdot \mathbf{1}[t = \tau] \\ & + \sum_{\tau} \psi_{2,\tau} \text{Connect}_{o,t} \cdot \text{Connect}_{d,t} \cdot \log(\text{Distance})_{s,c} \cdot \mathbf{1}[t = \tau] \\ & + \sum_{\tau} \lambda_{\tau} \log(\text{Distance})_{s,c} \cdot \mathbf{1}[t = \tau] + \gamma_{o,t} + \delta_{d,t} + \phi_{s,c} + \epsilon_{s,d,t} \end{aligned} \quad (4)$$

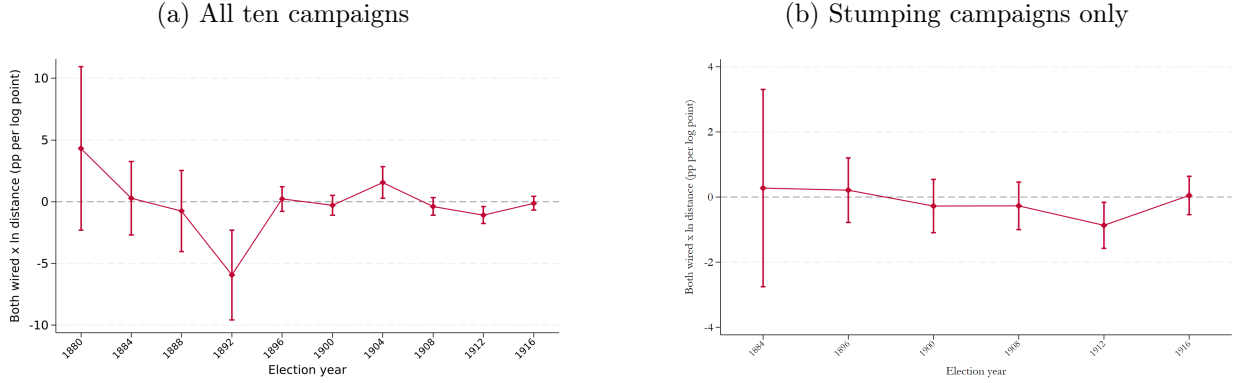
where $\psi_{1,\tau}$ and $\psi_{2,\tau}$ are the counterparts of ψ_1 and ψ_2 in column (4) of Table 1 for year τ . Log distance is demeaned, so $\psi_{1,\tau}$ is the connection effect at the average log distance, and λ_{τ} is the distance slope among county pairs that are not both connected. Each figure shows the estimates for all ten campaigns in panel (a) and for the six stumping campaigns in panel (b).

Figure D1: Connection effect by election year, at the average log distance



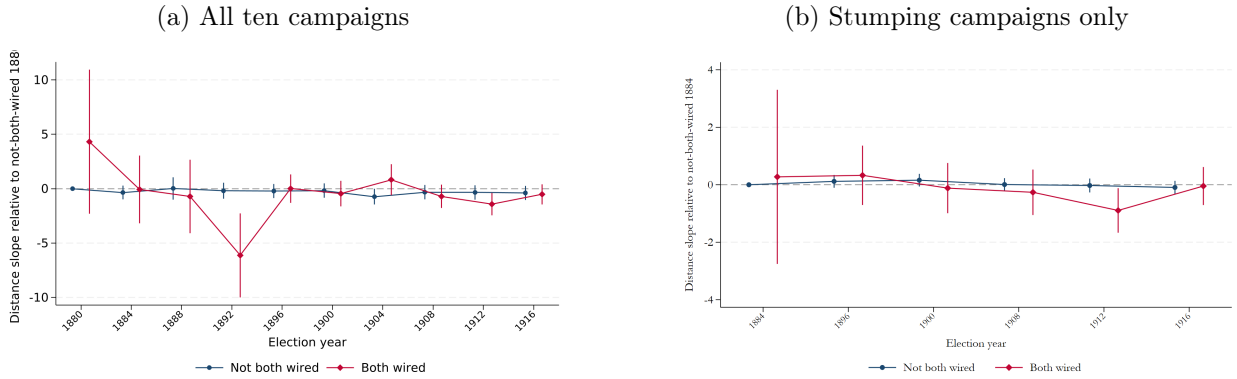
Notes: This figure shows $\psi_{1,\tau}$ from equation (4), the effect of both counties being connected to a news agency on the probability that the speech is covered, at the average log distance, in each election year. Panel (a) estimates it on all ten campaigns from 1880 to 1916; panel (b) on the six in which candidates stumped, dropping 1880, 1888, 1892 and 1904. Log distance is centered on the full ten-year panel in both panels, so they are read at the same distance and are comparable with Table 1. The outcome is multiplied by 100, so coefficients are in percentage points. Vertical lines are 95% confidence intervals from standard errors clustered by county pair. The estimates for 1880 and 1892 rest on fewer than thirty speech-origin counties and set the scale of panel (a); the narrow band of the later campaigns is consistent with a connection effect that is present in every campaign with many speeches.

Figure D2: Distance interaction by election year



Notes: This figure shows $\psi_{2,\tau}$ from equation (4), the additional distance slope of both-connected county pairs in each election year, in percentage points per log point of distance. It is a level rather than a change from a reference year, because equation (4) contains no uninteracted connection-by-distance term and so omits no year from that set. Panel (a) uses all ten campaigns, panel (b) the six stumping campaigns; the year-specific distance terms λ_τ are based at 1880 in panel (a) and 1884 in panel (b), which does not affect $\psi_{2,\tau}$. Vertical lines are 95% confidence intervals from standard errors clustered by county pair. A series that stays at or below zero is consistent with a distance friction that did not attenuate as more counties connected.

Figure D3: Distance slope by connection status and election year

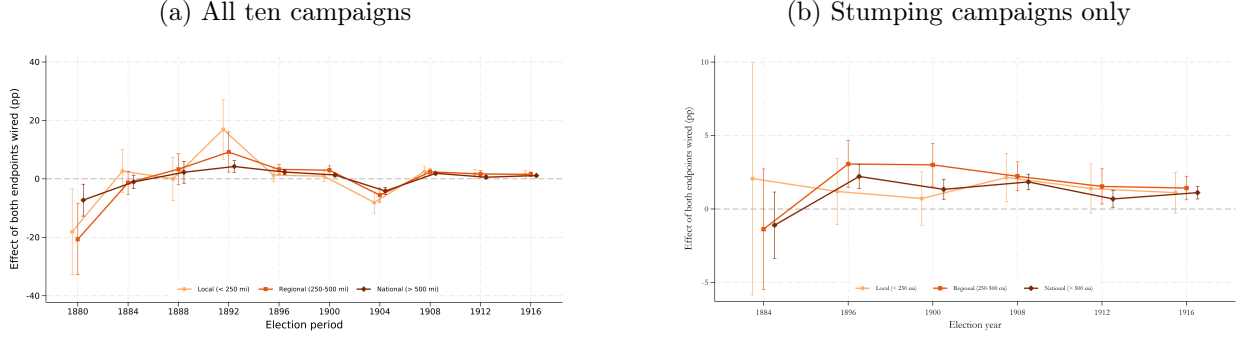


Notes: This figure shows the two components of the interaction in Figure D2. The navy series is λ_τ from equation (4), the distance slope among county pairs that are not both connected; the red series is $\lambda_\tau + \psi_{2,\tau}$, the slope among pairs that are both connected. Both are measured relative to the slope of not-both-connected pairs in the first campaign shown, 1880 in panel (a) and 1884 in panel (b), and the vertical distance between the two series in a given year is the coefficient in Figure D2. Confidence intervals are shown for each series separately and not for their difference. Vertical lines are 95% confidence intervals from standard errors clustered by county pair. A flat navy series and a negative red series are consistent with a distance gradient that exists only where a connection creates coverage (see the text).

Equation (4) imposes a log-linear effect of distance in every year. Figure D4 relaxes this assumption. It replaces the continuous distance terms with indicators for three distance bands, which are local (under 250 miles), regional (250 to 500 miles), and national (over 500 miles), each interacted with campaign year:

$$\begin{aligned}
 \text{Coverage}_{s,d,t} = & \alpha_t^L \text{Local}_{s,c} + \alpha_t^R \text{Regional}_{s,c} + \alpha_t^N \text{National}_{s,c} \\
 & + \psi_t^L \text{Connect}_{o,t} \cdot \text{Connect}_{d,t} \cdot \text{Local}_{s,c} + \psi_t^R \text{Connect}_{o,t} \cdot \text{Connect}_{d,t} \cdot \text{Regional}_{s,c} \\
 & + \psi_t^N \text{Connect}_{o,t} \cdot \text{Connect}_{d,t} \cdot \text{National}_{s,c} + \gamma_{o,t} + \delta_{d,t} + \phi_{s,c} + \epsilon_{s,d,t}
 \end{aligned} \tag{5}$$

Figure D4: Connection effect by distance band and election year



Notes: This figure shows ψ_t^L , ψ_t^R and ψ_t^N from equation (5), the effect of both counties being connected on the probability that the speech is covered, relative to not-both-connected pairs in the same distance band and the same election year. Panel (a) uses all ten campaigns, panel (b) the six stumping campaigns; the thin campaigns of 1880 and 1892 set the scale of panel (a). Vertical lines are 95% confidence intervals from standard errors clustered by county pair. Three series that move together are consistent with a connection effect whose distance profile did not change over the period.

The effect and the distance interaction by year. Every campaign that appears in both panels of Figures D1 to D4 has essentially the same coefficient, since the restriction removes only five percent of the observations. The panels differ in scale because the estimates for 1880 and 1892 rest on fewer than thirty speech-origin counties, so panel (b) is the readable one. In panel (b) of Figure D1, the connection effect is imprecise in 1884, peaks in 1896, and declines gradually through 1916, and it is positive in every campaign with many speeches. Figure D2 shows that the distance interaction is negative or close to zero in the same campaigns and never turns positive. Figure D4 shows the same pattern without the log-linear restriction. The three bands move together from campaign to campaign and their confidence intervals overlap in every campaign with many speeches, so there is no evidence that coverage first rose locally and later spread to regional and national pairs. The coverage that connection created therefore remained concentrated among nearby pairs throughout the period.

Distance slope by connection status. Figure D3 shows the distance slope separately for county pairs that are both connected and for pairs that are not. Among pairs that are not both connected, the slope λ_τ is small in every campaign, within half a percentage point per log point of its 1880 value. Among connected pairs, the slope $\lambda_\tau + \psi_{2,\tau}$ is negative in every campaign with many speeches, between -0.6 and -1.5 percentage points per log point. The distance gradient in coverage therefore comes from connected pairs.

This result does not imply that distance was irrelevant for newspapers without a connection. Pairs that are not both connected have almost no coverage at any distance. The probability that a newspaper in one county covers a speech in another county is 0.35 percent among these pairs, pooling 1880–1916. A slope of -1 percentage point per log point, as among connected pairs in 1908, implies a fall of about 2.3 percentage points between a pair 250 miles apart and a pair 2,500 miles apart, more than six times the coverage rate of unconnected pairs. A slope of that size is not possible for a probability that cannot fall below zero. The near-flat slope among unconnected pairs is therefore mechanical. The informative result is that a news-agency connection creates coverage, and that the coverage it creates falls with distance.

D.2 Pair-Level Trend Controls and Capital/Hub Exclusions

The speech-county-by-year and coverage-county-by-year fixed effects in equation (1) absorb *any* characteristic of a single county interacted with time. For example, they absorb an indicator for state capitals interacted with year, an indicator for New York interacted with year, or population above any threshold interacted with year. For the same reason, an indicator for *either* county being a capital, interacted with year, would be identified only from the pairs in which both counties are capitals. The fixed effects do not absorb trends that are specific to a type of pair. This is the concern raised in Section 4.1, in which a county connects in order to exchange news with one particular place, such as its state capital or a hub county.

Table D1 therefore re-estimates column (4) of Table 1 with pair-level trend controls, first one group at a time and then all together. Column (2) adds trends for state-capital pairs. These are indicators for the origin being the capital of the destination's state, for the destination being the capital of the origin's state, and for both counties being capitals, each interacted with campaign-year indicators. Column (3) adds a trend for pairs in the same state, since the pair fixed effects only absorb the level. Column (4) adds trends for hub pairs. These are an indicator for both counties being hubs and each side's hub status interacted with log distance, all interacted with year. Column (5) replaces these lists with a systematic version, which is a set of fixed effects for the origin's 1880 population decile interacted with the destination's 1880 population decile and with year. These allow pairs of counties of any given sizes to follow their own trend in coverage. Column (6) allows the effect of distance on coverage to change by year, and column (7) includes all the controls at once.

Table D1: Coverage of Presidential Speeches: Pair-Level Trend Controls

Dep. Var.: 1[Coverage > 0] x 100	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Both endpoints wired	1.51*** (0.18)	1.50*** (0.18)	1.49*** (0.18)	1.52*** (0.18)	1.51*** (0.18)	1.50*** (0.18)	1.50*** (0.18)
Both endpoints wired x Distance (logs)	-0.46* (0.26)	-0.45* (0.26)	-0.43* (0.26)	-0.48** (0.25)	-0.45* (0.26)	-0.42 (0.26)	-0.43* (0.25)
Origin x Year FE	X	X	X	X	X	X	X
Destination x Year FE	X	X	X	X	X	X	X
Origin x Destination FE	X	X	X	X	X	X	X
State-capital pair x Year		X	X				X
Same state x Year			X				X
Hub pair x Year				X			X
1880 size-decile pair x Year FE					X		X
Distance (logs) x Year						X	X
Observations	692,312	692,312	692,312	692,312	692,308	692,312	692,308
R-squared	0.442	0.443	0.443	0.443	0.443	0.443	0.444

Notes: All columns are the panel specification of column (4) of Table 1. State-capital pair trends interact indicators for the origin being the capital of the destination's state, the reverse, and both endpoints being capitals with campaign-year dummies; capitals are mapped to the counties of the 1890 frame (Denver in Arapahoe, Richmond in Henrico; the mid-window moves in Arizona, Louisiana, and West Virginia are dated; Nevada's and Oklahoma's capital counties do not exist in the frame and carry no flag). Hub pairs interact both-hub status and each side's hub status times log distance with year, where hubs are New York, Kings, Cook, and the District of Columbia. Size-decile pairs are fixed effects for origin 1880-population decile x destination 1880-population decile x year, with deciles over unique counties; counties without a positive 1880 population use their earliest positive election year.

* p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors clustered by speech county x coverage county pairs.

The estimates do not change. The connection effect stays within a fifth of a standard error of the baseline in every column, including column (7) with all the controls at once, and the distance interaction is equally stable. Observable trends by type of pair therefore do not explain the results. Appendix D.4 returns to this point.

Table D2 changes the comparison group. Pairs that never become both-connected may differ from pairs that do, and with an absorbing treatment they supply much of the counterfactual trend. If that difference in type drove the estimate, restricting the sample to pairs that eventually connect would move it. The table re-estimates column (4) of Table 1 on pairs that are both-connected at some point in the window, so identification comes only from *when* they connected, with not-yet-connected pair-years as the controls. The connection effect is statistically unchanged on twelve percent of the observations, with a somewhat steeper distance decay. The pairs that never connect therefore do not drive the estimate.

Table D2: Coverage of Presidential Speeches: Ever-Adopter Pairs Only

Dep. Var.: 1[Coverage > 0] x 100	(1) All pairs	(2) Ever both-wired pairs only
Both endpoints wired	1.51*** (0.18)	1.69*** (0.42)
Both endpoints wired x Distance (logs)	-0.46* (0.26)	-0.49 (0.31)
Origin x Year FE	X	X
Destination x Year FE	X	X
Origin x Destination FE	X	X
Observations	692,312	86,087
R-squared	0.442	0.445

Notes: Column (1) repeats column (4) of Table 1. Column (2) keeps only pairs that are ever both-connected inside the 1880–1916 window; pairs already connected in 1880 are retained but carry no identifying variation under the pair fixed effects.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by speech county x coverage county pairs.

Table D3 returns to the capital and hub counties and drops them from the sample, instead of controlling for their trends as Table D1 does. Excluding the 39 counties that are ever a state capital or a hub during the sample period as speech origins lowers the connection effect from 1.74 to 1.26 percentage points, and excluding them on either side of the pair lowers it to 1.05. Both estimates are precise, and the distance interaction is slightly flatter. The difference reflects heterogeneity. Connections that involve capitals and hubs have a larger effect on coverage, which is natural if their news is more newsworthy. However, the result does not depend on them. Among ordinary county pairs, the effect is still three times the coverage rate of 0.35 percent among pairs that are not both connected. Together with Table D1, these results limit the concern raised in Section 4.1. The few places for which a pair-specific motive to connect is plausible increase the estimate, but they do not drive it.

Table D3: Coverage of Presidential Speeches: Excluding Capitals and Hubs

Dep. Var.: 1[Coverage > 0] x 100	(1) Baseline	(2) Excl. capital/hub origins	(3) Excl. capital/hub either side
Both endpoints wired	1.51*** (0.18)	1.01*** (0.18)	0.81*** (0.17)
Both endpoints wired x Distance (logs)	-0.46* (0.26)	-0.38 (0.26)	-0.43** (0.22)
Origin x Year FE	X	X	X
Destination x Year FE	X	X	X
Origin x Destination FE	X	X	X
Observations	692,312	599,945	557,687
R-squared	0.442	0.454	0.435

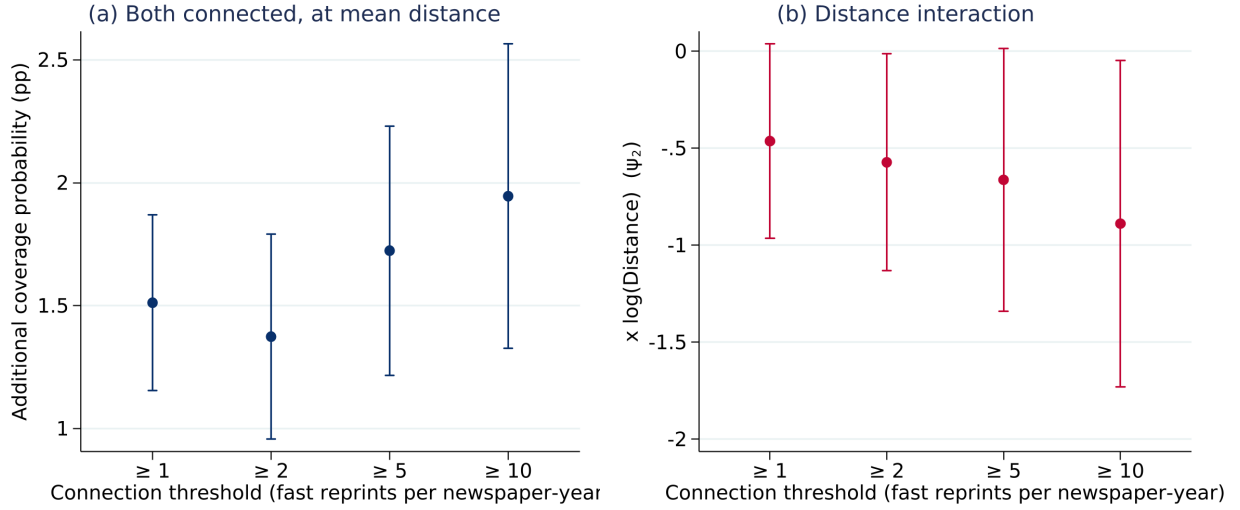
Notes: Column (1) repeats column (4) of Table 1. Column (2) drops every dyad whose speech origin is ever a state capital or a hub county (New York, Kings, Cook, District of Columbia) within the window; column (3) drops such counties from either side of the pair.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by speech county x coverage county pairs.

D.3 Stricter Connection Thresholds

Appendix B.3 shows that a single fast reprint is thin evidence of agency access and defines stricter measures that require two, five, or ten fast reprints per newspaper-year. Figure D5 re-estimates column (4) of Table 1 under each definition on the same sample, so only the treatment coding changes. The coverage effect of joint connection is positive and statistically significant at every threshold, and it is largest at ten reprints, where the marginal connections that Appendix B.3 identifies as the weakest leave the treated group. The distance interaction steepens somewhat as the threshold rises, though it is measured imprecisely under the stricter definitions. The coverage result therefore does not depend on the one-reprint definition.

Figure D5: Coverage regression estimates under stricter connection thresholds



Notes: This figure shows the specification of column (4) of Table 1 re-estimated with the connection indicators recoded at each threshold; the estimation sample is identical across thresholds (673,435 dyads), so only the treatment coding moves. Panel (a) reports the additional coverage probability when both endpoints are connected, evaluated at the mean log distance ($\hat{\psi}_1 + \overline{\log(\text{Distance})} \hat{\psi}_2$, with the linear-combination standard error); panel (b) reports the distance interaction $\hat{\psi}_2$. Bars are 95% confidence intervals from standard errors clustered by county pair. Estimates that hold or rise at stricter thresholds are consistent with a coverage effect that does not rest on the weakest connections.

D.4 Identification Tests: Leads, Lags, and Event Studies

Two features of the setting limit the timing tests. First, I observe coverage of presidential speeches only in campaign years, while I observe connection every year. A lead of four years is the connection status at the time of the next campaign, so it is not a placebo. Informative leads are therefore limited to one or at most two years. Second, the data cover ten campaigns (1880–1916), and an event study in campaign time requires that every adoption cohort is observed over the full window. Widening the window from ± 1 to ± 2 to ± 3 campaigns reduces the number of eligible cohorts from eight to six to four. For these reasons, I focus on horizons of one and two campaigns and interpret the ± 3 window with caution.

Throughout, y is the calendar year of the election, so “connected by $y - 1$ ” is the treatment in column (4) of Table 1, “by y ” adds the pairs that connect during the election year, and “by $y + 1$ ” and “by $y + 2$ ” are pairs that connect only after it. Connection is an absorbing state, so the indicators for “both counties connected by calendar year $y + k$ ” are nested. A pair that is connected by $y - 1$ is also connected by y , $y + 1$, and $y + 2$. When these indicators enter the regression together, each coefficient measures the increment for the pairs whose connection date falls between two adjacent indicators. The cumulative sum, starting from the last lead, gives the difference in coverage between each group of pairs and the pairs that are still unconnected two years after the campaign. Table D4 reports these estimates.

Table D4: Leads and Lags of Joint Connection Status

Dep. Var.: 1[Coverage > 0] x 100	(1)	(2)
Both wired by calendar $y+2$		0.80*** (0.30)
Both wired by calendar $y+1$	0.41 (0.34)	-0.30 (0.45)
Both wired by calendar y	0.40 (0.53)	0.43 (0.53)
Both wired by calendar $y-1$	0.76* (0.40)	-0.50 (0.57)
Both wired by calendar $y-2$		1.35*** (0.42)
Origin x Year FE	X	X
Destination x Year FE	X	X
Origin x Destination FE	X	X
Observations	692,312	692,312
R-squared	0.442	0.443

Notes: Each regressor is an indicator for the pair being ever connected by the stated calendar year, so the set is nested and each coefficient is the increment for pairs whose connection date falls between two adjacent indicators. The specification is otherwise column (4) of Table 1 without the distance interaction.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by speech county x coverage county pairs.

Under the identifying assumption, the lags can have an effect, because coverage can build up after connection. The leads should have no effect, because pairs that are not yet connected in campaign year y are untreated. A positive coefficient on the indicators for “by calendar year $y + 1$ ” or “by calendar year $y + 2$ ” would indicate anticipation. Table D4 shows two things. First, connecting *during* the campaign year itself adds almost nothing, which supports the use of the previous year’s connection status in the main specification. Second, there is some anticipation. Adding up the coefficients in column (2), the pairs that will connect in the two years *after* the campaign already have coverage 0.7 to 0.9 percentage points higher than the pairs that remain unconnected, against a difference of 2.0 for pairs connected at least two years earlier. Between a third and a half of the difference is therefore present before the connection date. The rest of this appendix asks where that anticipation comes from.

Table D5 separates the one-year leads and lags of the origin and of the destination, holding the other county’s status at calendar year y . Column (1) is the baseline. The other columns repeat the specification with the full set of pair-level trend controls, excluding capital and hub counties, and restricting the sample to pairs that eventually connect, as in Appendix D.2.

Table D5: Leads and Lags by Side, Under Controls, Exclusions, and the Ever-Adopter Restriction

Dep. Var.: 1[Coverage > 0] x 100	(1) Baseline	(2) Pair-trend controls	(3) Excl. capital/hub	(4) Ever-adopter pairs
Origin y+1, destination y	-1.29** (0.57)	-1.33** (0.57)	-0.84 (0.55)	1.06 (1.08)
Origin y, destination y+1	1.48*** (0.43)	1.46*** (0.43)	1.12** (0.44)	1.32 (0.88)
Both wired by calendar y	-0.34 (1.08)	-0.28 (1.08)	-0.78 (1.04)	-1.90 (2.02)
Origin y-1, destination y	1.22* (0.64)	1.23* (0.64)	0.59 (0.61)	1.20 (1.05)
Origin y, destination y-1	0.50 (0.51)	0.47 (0.51)	0.70 (0.45)	0.30 (1.05)
Origin x Year FE	X	X	X	X
Destination x Year FE	X	X	X	X
Origin x Destination FE	X	X	X	X
Pair-trend controls x Year		X		
Observations	692,312	692,308	557,687	86,087
R-squared	0.442	0.444	0.435	0.445

Notes: All columns enter the one-year leads and lags of each endpoint's connection status jointly, holding the other endpoint at calendar y , in the specification of column (4) of Table 1 without the distance interaction; outcomes are capped at 1916. Column (1) is the baseline; column (2) adds every pair-level trend control of Table D1; column (3) excludes capital/hub counties from either side; column (4) keeps only pairs ever both-connected in the window.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by speech county x coverage county pairs.

The anticipation comes entirely from the destination. In column (1), the lead of the origin's connection is negative and insignificant, while destinations that will connect within a year already cover the origin 1.43 percentage points more. This is the direction of selection discussed in Section 4.1, in which newspapers connect *because* their demand for outside news is rising. The other columns show where it comes from. The destination lead barely moves with the full set of trend controls, so observable trends by type of pair do not explain it. It falls by about a third when I exclude capital and hub counties, and it is larger but imprecise among pairs that eventually connect. About a third of the lead therefore comes from capitals and hubs. The rest is most consistent with error in the date of connection. I date connection by the first *observed* fast reprint, which can occur after the start of the service. The event study without capitals and hubs, presented below, shows no pre-trends, which is not consistent with selection that is specific to a pair. Table D8 at the end of this appendix tests the dating-error explanation directly.

Table D6, together with Figure 10 in the main text, presents event studies around the first campaign in which both counties of a pair are connected. Event time is measured in campaigns (four years each), the reference period is the campaign immediately before connection, and pairs that never connect serve as controls. I estimate each window twice, in the full sample (columns (1)–(3)) and excluding capital and hub counties from either side of the pair (columns (4)–(6)). Figure 10 plots the ± 2 and ± 3 windows. The ± 1 window has no pre-period coefficient beyond the

reference, so the table alone reports it.

Table D6: Event Studies Around First Joint Connection: Baseline and Excluding Capitals/Hubs

Dep. Var.: 1[Coverage > 0] x 100	Baseline			Excluding capitals and hubs		
	± 1	± 2	± 3	± 1	± 2	± 3
$k = -3$			-3.05*** (0.83)			-1.54** (0.77)
$k = -2$		-0.11 (0.39)	-0.11 (0.47)		0.18 (0.37)	0.36 (0.47)
$k = 0$	1.08*** (0.42)	0.78* (0.40)	0.37 (0.48)	0.39 (0.39)	0.34 (0.36)	0.13 (0.43)
$k = 1$	1.76*** (0.39)	1.44*** (0.41)	0.88** (0.45)	1.33*** (0.36)	0.75** (0.38)	0.03 (0.40)
$k = 2$		0.37 (0.35)	-0.13 (0.37)		0.41 (0.32)	-0.25 (0.32)
$k = 3$			0.72* (0.42)			-0.08 (0.39)
Campaigns before/after	1	2	3	1	2	3
Eligible adoption cohorts	8	6	4	7	6	4
Origin x Year FE	X	X	X	X	X	X
Destination x Year FE	X	X	X	X	X	X
Origin x Destination FE	X	X	X	X	X	X
Joint pre-trend p-value	.	0.7752	0.0007	.	0.6239	0.0527
Observations	622,007	629,735	628,221	512,540	517,586	516,167
R-squared	0.481	0.471	0.468	0.474	0.461	0.461

Notes: Event time k counts campaigns (four years each) relative to the first campaign in which the pair is both-connected; $k = -1$ is the omitted reference in every column. Treated pairs contribute observations only inside the stated window; never-treated pairs remain controls; pairs already connected in 1880 and cohorts without full calendar support are excluded. Columns (1)–(3) use the full sample; columns (4)–(6) drop every dyad with a state capital or hub county (New York, Kings, Cook, District of Columbia) on either side, as in Table D3. The joint pre-trend test covers all pre-period coefficients beyond the reference; the ± 1 window has none.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by speech county x coverage county pairs.

Under the identifying assumption, the coefficients before connection ($k < 0$) should be zero, and the effect should appear at $k = 0$ and persist. In the full sample, the ± 1 and ± 2 windows are consistent with this pattern. Coverage increases at the first connected campaign and remains higher afterwards, and the coefficient at $k = -2$ is small and insignificant (joint test of pre-trends, $p = 0.24$). The ± 3 window is less clean. Relative to the campaign before connection, coverage was significantly *lower* two and three campaigns earlier, which means that coverage was rising before connection, and the joint test rejects the absence of pre-trends ($p = 0.001$). Three caveats apply to this finding. First, the window has only four eligible adoption cohorts. Second, its earliest periods depend on the front-porch campaigns, which have very few speech origins (Appendix D.1). Third, these are conventional two-way fixed-effects event studies, so with staggered adoption the coefficients can combine the effects of different cohorts.

The columns without capitals and hubs show where the pre-trend comes from. When I exclude these counties, the increase at the time of connection remains, the coefficients before connection shrink toward zero, and the joint test for the ± 3 window no longer rejects ($p = 0.30$). Table D7 extends this comparison to the four samples and specifications of Table D5. Its baseline and

exclusion columns repeat those of Table D6.

Table D7: Event-Study Pre-Trends, Under Controls, Exclusions, and the Ever-Adopter Restriction

Dep. Var.: 1[Coverage > 0] x 100	(1) Baseline	(2) Pair-trend controls	(3) Excl. capital/hub	(4) Ever-adopter pairs	(1) Baseline	(2) Pair-trend controls	(3) Excl. capital/hub	(4) Ever-adopter pairs
$k = -3$					-3.05*** (0.83)	-3.09*** (0.83)	-1.54** (0.77)	-1.86 (3.70)
$k = -2$	-0.11 (0.39)	-0.11 (0.39)	0.18 (0.37)	0.59 (1.09)	-0.11 (0.47)	-0.11 (0.47)	0.36 (0.47)	-0.41 (1.29)
$k = 0$	0.78* (0.40)	0.76* (0.40)	0.34 (0.36)	0.17 (0.75)	0.37 (0.48)	0.33 (0.48)	0.13 (0.43)	1.15 (0.99)
$k = 1$	1.44*** (0.41)	1.42*** (0.42)	0.75** (0.38)	1.90** (0.78)	0.88** (0.45)	0.85* (0.45)	0.03 (0.40)	2.11** (1.02)
$k = 2$	0.37 (0.35)	0.34 (0.36)	0.41 (0.32)	0.00 (.)	-0.13 (0.37)	-0.16 (0.38)	-0.25 (0.32)	-0.07 (1.00)
$k = 3$					0.72* (0.42)	0.67 (0.42)	-0.08 (0.39)	0.00 (.)
Campaigns before/after	2	2	2	2	3	3	3	3
Joint pre-trend p-value	0.7752	0.7813	0.6239	0.5910	0.0007	0.0006	0.0527	0.8727
Pair-trend controls x Year		X				X		
Origin x Year FE	X	X	X	X	X	X	X	X
Destination x Year FE	X	X	X	X	X	X	X	X
Origin x Destination FE	X	X	X	X	X	X	X	X
Observations	629,735	629,731	517,586	23,440	628,221	628,217	516,167	21,936
R-squared	0.471	0.472	0.461	0.545	0.468	0.469	0.461	0.526

Notes: The event-study design of Table D6 ($k = -1$ is the reference; eligible cohorts need full calendar support), estimated in the ± 2 window (first four columns) and the ± 3 window (last four), each for the four samples and specifications of Table D5. In the ever-adopter columns the never-connected pairs are dropped, so the controls are not-yet-connected adopter pairs only; without a never-treated group, event time is collinear with the year and pair fixed effects, so one post-period cell ($k = 2$ in the ± 2 window, $k = 3$ in the ± 3 window) is absorbed and reported as zero, and those columns are effectively normalized to two reference periods.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by speech county x coverage county pairs.

The pre-trend in the ± 3 window comes from capitals and hubs. The joint test rejects in the full sample and with the full set of trend controls, so observable trends by type of pair do not explain it. It does not reject when I exclude capital and hub counties or when I restrict the sample to pairs that eventually connect, and the ± 2 window does not reject in any of the four cases. In the sample without capitals and hubs, the connection effect is still 1.05 percentage points (Table D3). The pattern of rising coverage before connection is therefore concentrated in the few counties that Section 4.1 identifies as the exception.

Finally, I use the connection thresholds to test the dating-error explanation. I date connection by the first *observed* fast reprint. Requiring two, five, or ten fast reprints per newspaper (Appendices B.3 and D.3) therefore moves every connection date *later*. If the remaining anticipation reflects connection dates that are too late, it should become stronger as the threshold rises. Selection that is specific to a pair gives no reason for the anticipation to depend on the number of reprints. Table D8 re-estimates the one-year leads by side and the joint pre-trend tests at each threshold, with and without capitals and hubs.

Table D8: Anticipation Diagnostics Under Stricter Connection Thresholds

Connection threshold (fast reprints)	Baseline				Excluding capitals and hubs			
	1	2	5	10	1	2	5	10
<i>Panel A: One-year leads by side (leads-and-lags design, annual timing)</i>								
Origin y , destination $y + 1$	1.48*** (0.43)	1.18*** (0.41)	1.28*** (0.46)	1.60** (0.66)	1.12** (0.44)	0.65 (0.39)	0.57 (0.46)	0.56 (0.62)
Origin $y + 1$, destination y	-1.29** (0.57)	-1.52** (0.64)	-1.87*** (0.61)	-2.18*** (0.70)	-0.84 (0.55)	-1.00 (0.62)	-1.26** (0.60)	-1.91*** (0.71)
Observations	692,312	692,312	692,312	692,312	557,687	557,687	557,687	557,687
<i>Panel B: Joint pre-trend p-values (campaign event studies)</i>								
± 2 campaign window	0.7752	0.6963	0.9869	0.5257	0.6239	0.7608	0.2782	0.8318
± 3 campaign window	0.0007	0.0010	0.0000	0.0002	0.0527	0.0482	0.2252	0.4904

Notes: Each column recodes connection at the stated number of fast reprints per newspaper (Appendix B.3) and re-estimates the timing diagnostics under that definition, with pair adoption dates re-derived accordingly; the one-reprint columns are the paper's definition and repeat the corresponding estimates of Tables D5 and D7. The last four columns exclude capital and hub counties from either side of the pair, as in Table D3. Panel A reports the one-year lead coefficients from the by-side leads-and-lags design of Table D5, estimated separately at each threshold; the lag terms and the contemporaneous term are included in every regression but not shown. Panel B reports the joint pre-trend p-values from the ± 2 and ± 3 campaign event studies of Table D6, re-estimated at each threshold.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by speech county x coverage county pairs.

Table D8 is consistent with the dating-error explanation. The one-year lead of the destination increases monotonically from 1.43 to 2.32 percentage points as the threshold rises from one to ten reprints, the joint test for the ± 2 window goes from not rejecting to rejecting, and the rejection in the ± 3 window becomes stronger. In contrast, the lead of the origin is insignificant at every threshold. This asymmetry is what dating error predicts. I infer the connection date of a destination from its own reprints, so a destination that is about to cross the threshold is already reprinting fast. Nothing links the recorded connection date of an origin to the coverage that it receives elsewhere.

Raising the threshold also changes which pairs are treated, so this evidence supports the dating-error explanation but does not prove it. Without capitals and hubs, the increase is smaller and the ± 3 window does not reject at the five percent level at any threshold. At ten reprints the ± 2 window does reject, which suggests that a dating error of that size also affects ordinary pairs. However, all three diagnostics move monotonically in the predicted direction, and only on the predicted side of the pair, which is difficult to explain with selection. The results also imply that the baseline definition of one reprint is the *least* exposed to late dating, so the anticipation under the baseline definition is the smallest.

E Robustness and Identification of the Media-Reach Response

The identifying assumption of equation (3) is that, conditional on a county’s own connection, the timing of connections *elsewhere* is unrelated to the factors that make the county a campaign target. This appendix examines the estimates of Table 2 and this assumption in three steps. I first report the full regressions behind Table 2, including the two main effects that the table omits, for each version of the outside audience and for two further versions that remove the influence of state capitals and hub counties (Appendix E.1). I then test whether campaigns anticipated the growth of the network (Appendix E.2). Finally, I re-estimate the results under stricter definitions of connection (Appendix E.3).

E.1 The Full Decomposition

Table E1 reports all three coefficients of equation (3). Media reach, the coefficient β on the interaction of the outside audience with the county’s own connection, is the coefficient of interest and the only one that Table 2 displays. The coefficient on reachable voters, α_2 , measures how the outside audience relates to visits among counties that were not connected. A speech in these counties could not travel through the network, so under the mechanism this coefficient should be zero, and I use it as a test of the design. The coefficient on the local wire, α_1 , is the change in the probability of a visit when a county connects, evaluated at the average outside audience.

Table E1: Own Connection versus Reachable Audience Elsewhere, by Audience Definition

Dep. Var.: 1[Speech > 0] x 100	(1) Baseline	(2) Baseline + Controls	(3) 100-Mile Buffer	(4) Fixed 1880 Population	(5) All Aged 21+	(6) No Capital/Hub Destinations	(7) Purged Weights	(8) Both Purges
Media reach (Reachable voters × Local wire)	2.73*** (0.59)	2.26*** (0.57)	2.06*** (0.56)	2.15*** (0.60)	2.24*** (0.57)	2.18*** (0.56)	2.04*** (0.57)	1.98*** (0.56)
Reachable voters (std)	1.38 (1.79)	1.05 (1.92)	-8.63*** (2.01)	1.17 (1.66)	1.10 (1.88)	-0.27 (2.01)	0.50 (1.10)	-0.22 (1.13)
Local wire	2.13** (0.94)	1.69* (0.90)	1.60* (0.91)	1.90** (0.90)	1.71* (0.90)	1.70* (0.91)	1.94** (0.89)	1.91** (0.90)
Wired-origin response	4.11 (1.79)	3.31 (1.95)	-6.57 (2.08)	3.31 (1.69)	3.34 (1.91)	1.91 (2.03)	2.55 (1.18)	1.75 (1.20)
Controls	None	All	All	All	All	All	All	All
County FE	X	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X	X
Observations	23,048	23,048	23,048	23,048	23,048	23,048	23,048	23,048
R-squared	0.315	0.320	0.321	0.320	0.320	0.320	0.320	0.320

Notes: Each column estimates equation (3) and displays all three coefficients. Columns (1) through (6) are the regressions of Table 2. Columns (3) through (8) keep the controls of column (2) and change the construction of the outside audience. Column (3) counts only destinations more than 100 miles away. Column (4) holds destination electorates at their 1880 levels. Column (5) counts all adults aged 21 and over. Column (6) drops state capitals and hub counties as destinations. Column (7) uses coverage weights from equation (1) re-estimated without those counties on either side of the pair, and column (8) applies both changes. Media reach, the interaction of reachable voters with the local wire, is the coefficient of interest. The coefficient on reachable voters is the association between the outside audience and visits among counties that were not connected to a news agency, where a speech could not travel. It is a test of the design and should be zero. The wired-origin response is the sum of the two. It measures the association among connected counties only when the coefficient on reachable voters is zero, which is not the case in column (3).

* p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors clustered by county.

Columns (1) through (6) repeat the regressions of Table 2. The media reach coefficient is stable across them, with or without controls and whichever version of the outside audience enters

the regression. The test coefficient is statistically indistinguishable from zero in five of the six columns.³⁸ The exception is column (3), where the outside audience only counts destinations more than 100 miles away and the test coefficient is significantly negative.

Part of this coefficient reflects what the buffer removes. The buffer drops the destinations within 100 miles of each county. These destinations are a small part of the outside audience of most counties, so in levels the buffer measure is almost the same variable as the baseline measure. The regression, however, compares each county with its own average and with the other counties in the same year. When newspapers near a county connect, they enter the buffer measure of every county more than 100 miles away but not the measure of the county itself. The county therefore falls behind the rest of the country in buffer reach, even though its total reach rises. Net of the fixed effects, the nearby connections that the buffer drops push the buffer measure and the baseline measure in opposite directions. The wired-origin response, the sum of the two coefficients, is negative in this column for the same reason, and it should not be read as a negative response to media reach. The media reach coefficient itself is close to the baseline.

Columns (6) through (8) turn to the largest cities. State capitals and the hub counties of New York City, Chicago, and the District of Columbia add large electorates to the outside audience of every county. As discussed in Section 4.1, their connections are also the least likely to be exogenous. Column (6), which also appears in Table 2, drops these counties as destinations, so that their voters do not count toward the outside audience of any county. They account for about a quarter of total reach. Column (7) keeps every destination but uses coverage weights from equation (1) estimated without capital and hub counties on either side of the pair (Appendix D.2). These weights are lower at every distance, and lower by more at longer distances, so the resulting measure gives relatively more weight to nearby destinations. Column (8) applies both changes. The media reach coefficient barely moves in the three columns, and the test coefficient stays close to zero. The audiences of the largest cities therefore do not drive the response.

E.2 Anticipation of Network Growth

Table E2 tests the identifying assumption of Section 4.3 directly. Among connected counties, media reach varies only with connections elsewhere, and the assumption fails if candidates visited counties whose outside network was about to grow. To test this, I add to columns (1) and (2) of Table E1 the growth of the outside audience over the following year among counties that are already connected, with a value of zero for all other counties. I use the growth of the outside audience because the future *level* of media reach would also move with the county’s own upcoming connection. Among counties that are already connected, own connection cannot change, so the growth over one year only reflects connections elsewhere. I use a horizon of one year for the reason discussed in Appendix D.4. Connections that arrive three or four years later coincide with the *next* presidential campaign and can respond to it, so longer leads are not placebos.

³⁸The county and year fixed effects absorb all but about 1% of the variance of the outside audience. They absorb even more when distance decay is flatter, since the outside audience of every county then approaches the connected electorate of the whole country. As a result, over the 95% confidence interval of the decay ratio (footnote 26), the standard error of the test coefficient in column (2) rises from 1.0 to 5.8. The test coefficient remains indistinguishable from zero over this interval in every column except column (3), where it is significantly negative throughout.

Table E2: Future-Network-Growth Placebo

Dep. Var.: $1[\text{Speech} > 0] \times 100$	(1)	(2)
	No Controls	Controls
Growth of reachable audience, y to $y + 1$ (wired, std)	0.05 (0.28)	0.06 (0.28)
Reachable voters elsewhere (std)	1.38 (1.79)	1.04 (1.92)
Local wire availability	2.04* (1.05)	1.59 (1.02)
Reachable voters elsewhere $\times$ Local wire	2.69*** (0.61)	2.21*** (0.59)
County FE	X	X
Year FE	X	X
Observations	23,048	23,048
R-squared	0.315	0.320

Notes: “Growth of reachable audience, y to $y + 1$ ” is the one-year-ahead change in the outside audience among counties already connected by the previous year. It is zero for all other counties and standardized. Because connection is absorbing, own connection cannot change over this window, so the regressor moves only with connections elsewhere in the network. Column (1) has no controls beyond the displayed terms, and column (2) adds the full set of controls of Table 2. Both columns use the estimation sample of Table E1, and every regressor is standardized on it, so the other rows are in the same units as that table. Because the growth regressor is defined only among connected counties, its inclusion shifts the local wire and interaction coefficients relative to Tables 2 and E1. The coefficient on reachable voters among unconnected counties barely changes.

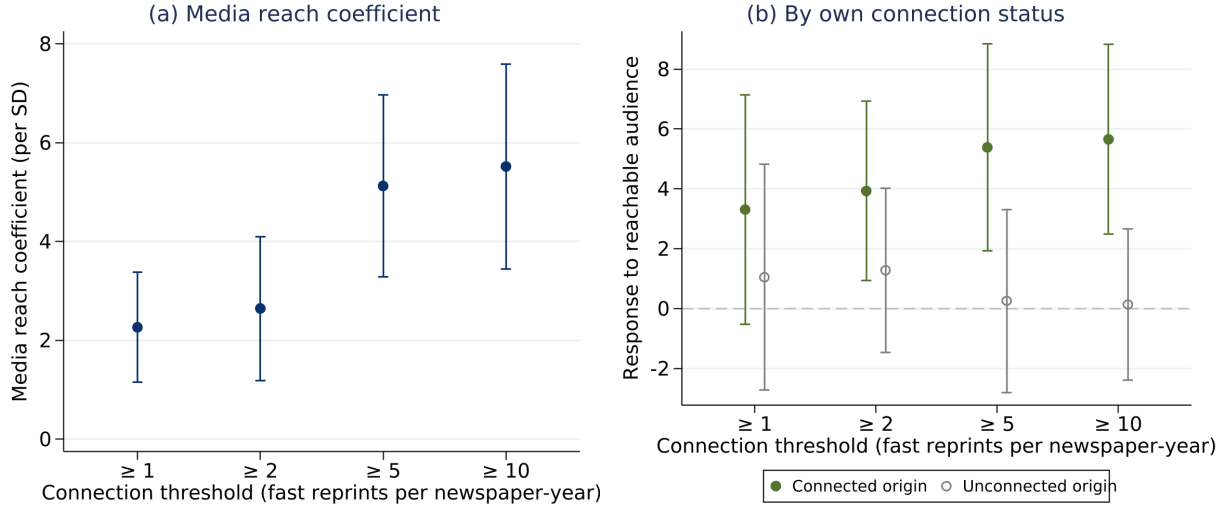
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by county.

The coefficient on future growth is negative and statistically significant with and without controls, while media reach remains positive and significant in the same regressions. Candidates were therefore slightly *less* likely to visit counties whose outside network was about to grow. The sign matters, because anticipation that would bias the estimates upward requires a *positive* coefficient. The pattern fits a network that expanded toward places that campaigns had not yet reached.

E.3 Stricter Connection Thresholds

Figure E1 re-estimates the media-reach regressions under the stricter connection definitions of Appendix B.3. Under each definition, media reach is rebuilt from that definition’s own coverage coefficients and connection network, and column (2) of Table E1 is re-estimated on the same sample. Panel (a) plots the media reach coefficient. Panel (b) plots the response of visits to the outside audience among connected and among unconnected counties, which is the test of the design. The response to media reach is positive and statistically significant at every threshold, and the response among unconnected counties is indistinguishable from zero at every threshold. The result and its test therefore do not depend on the one-reprint definition.

Figure E1: Media reach estimates under stricter connection thresholds



Notes: This figure shows the estimates of column (2) of Table E1 under each connection threshold, with media reach rebuilt from that threshold's own coverage coefficients and connection network (equation (2)) and standardized within its own definition. Panel (a) reports the media reach coefficient, the interaction of the outside audience with the local wire. Panel (b) reports the response of campaign stops to the outside audience among counties that are themselves connected ($\alpha_2 + \beta$, filled) and among those that are not (α_2 , hollow). One standard deviation of the outside audience corresponds to roughly 44,500 expected reachable voters at the baseline threshold, 38,700 at two articles, 40,800 at five, and 44,500 at ten. All estimates use the 23,034 county-years of Table E1, so the baseline points reproduce its column (2). Bars are 95% confidence intervals from standard errors clustered by county. A response that holds or rises at stricter thresholds, with a test coefficient at zero throughout, is consistent with a design that does not rest on the weakest connections.

The estimates grow as the definition becomes stricter, which is informative about what the binary indicator measures. The intensity of connection is very skewed. The newspapers above the threshold of ten articles typically print hundreds of fast reprints in a year, while the newspapers with one or two fast reprints are at the other end of the distribution. The marginal connections are still informative, since pairs and counties that only meet the one-article definition have clearly higher coverage than unconnected ones. However, the response is concentrated among the newspapers with the most intense connection, and each higher threshold restricts the treated group to them, so the estimated difference grows. The binary indicator is therefore a coarse measure of the intensity of news-agency access. The baseline estimates with one article are the most conservative, because they average weak and intense access in a single treated group.

To compare magnitudes across definitions, I express the coefficients in a common unit. Media reach is measured in expected reachable voters, so I can convert the coefficient under each definition from its own standard deviation to an effect per voter. Per 10,000 additional reachable voters, the media reach coefficients in panel (a) are 0.48, 0.64, 1.09, and 1.08 percentage points of the probability of a visit as the threshold rises from one to ten articles, against a visit probability of 5.6 percent in the estimation sample.³⁹ Per reachable voter, the response therefore roughly doubles from the baseline to the stricter definitions, in step with the plotted coefficients.

³⁹The probability of a visit is 4.8 percent among unconnected and 10.3 percent among connected counties. The average county has about 8,200 eligible voters, so 10,000 reachable voters are roughly one average county's electorate, and one standard deviation of the outside audience roughly five.

F Additional Content Results

This appendix reports the two further content margins described at the end of Section 5.2. The values labels and the other content measures are defined in Appendix C.5.

Values. Table F1 reports the same specification as Table 3. I estimate equation 3 for each share with state and year fixed effects and the full control set of Table 2. References to a community change with media reach. The share of passages that refer to the nation as “we” increases, and the share that refer to a class or group decreases. No individual share is statistically significant, but the three community shares change jointly ($p = 0.08$, Wald test across equations with standard errors clustered by county). The composition of warrants does not change ($p = 0.35$). The outside audience itself has no effect among unconnected counties (joint $p = 0.71$ for the community shares and 0.47 for the warrant shares), although this test has little power in this small sample. The shift is in the direction the audience-composition mechanism predicts.

Table F1: Media Reach and the Values Content of Speeches

Dep. Var. (x100)	Community references			Warrant		
	Nat. inward	Nat. outward	Group-class	Stake	Identity	Party record
Media reach	6.07 (5.30)	5.58 (3.74)	-3.57 (2.82)	-3.94 (4.84)	6.75 (5.14)	-2.81 (5.43)
State FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Mean dep. var.	22.6	12.1	6.3	25.3	29.7	45.1
Joint axis test (p)	0.094	0.094	0.094	0.403	0.403	0.403
Observations	366	366	366	366	366	366
R-squared	0.177	0.360	0.239	0.205	0.264	0.216

Notes: Each column reports β from a separate estimate of equation 3 for the named share of a county-year’s usable direct speech content (multiplied by 100), estimated on the 365 county-years with at least one usable direct passage in the five campaigns for which I build the values labels (1896, 1900, 1908, 1912, 1916). The community shares classify the collective a passage invokes as “we” (the nation in domestic terms, the nation facing the world, or a class or group); the warrant shares classify the ground it offers for support (material stakes, identity or principle, or the party’s record) and sum to one. All regressions include state and year fixed effects, the outside audience, and the control set of columns (2)–(6) of Table 2. “Joint axis test” is a county-clustered Wald test that media reach moves all shares of the group jointly (3 degrees of freedom for the community shares; 2 for the warrant shares, whose three coefficients sum to zero by construction). Standard errors, in parentheses, are clustered by county.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, based on unadjusted p -values.

Topic–interest alignment. I measure alignment by comparing the topics of a candidate’s speeches in a county with the economic composition of the county (from census variables) or with its interests (from the content of its non-political news). I compare the match with the distribution that I obtain by randomly re-assigning the same speeches to the same counties, and the statistic is the excess of the match over this benchmark. Because the statistic is built from permutations, conventional standard errors do not apply to it, and I do not report inference for its coefficient. I regress the statistic on media reach with the specification of Table 3. The point estimates are close to zero for every definition of alignment, every set of controls, and every set of fixed effects. However, they are based on 261 county-campaign observations between 1896 and 1916, and the outside audience varies almost only across state-years within this sample, so the test has little power.

G Robustness to the Labeling Model: The GPT-4o Labels

I first built the speech measure with OpenAI’s `gpt-4o`, then one of the leading models. As better models became available, the limitations of the original prompt became clear. The most important one is that it did not identify the speaker, and surrogates delivered many of the speeches in the early years. I therefore rewrote the prompt and labeled every article again. This appendix describes the original run and prompt (Appendix G.1), the differences between the two measures (Appendix G.2), and the main results under the original labels (Appendix G.3). The main results hold under both sets of labels, but the original labels likely assign speeches to counties that the candidate did not visit in the early years.

G.1 The Prompt and the Run

I labeled all articles between April and June of 2025 with OpenAI’s `gpt-4o` model through the Batch API, using the same sample as in Appendix C.1.⁴⁰ The total cost was nearly \$400.

The GPT-4o prompt differs from the current prompt in four ways. First, it asks a single yes-or-no question, whether the article refers to a public speech, address, or rally by the candidate. An article that reports a speech and an article that announces one for the following week therefore receive the same label. Second, it does not ask who spoke, so a rally addressed by a running mate or by a local speaker on behalf of the candidate counts as a speech. Third, it returns locations, dates, and crowd sizes as separate lists rather than as events, so a town cannot be linked to its speaker or to its date. Fourth, it places the candidate’s name in the instructions rather than in the message with the article, so the instructions differ across requests. In addition, the request set a temperature of 0.1 (the current model uses a reasoning effort) and required an answer in JSON without a fixed schema, so a malformed field was possible. As in the current prompt, the instructions ask for a one-sentence justification before each answer (Korinek, 2025) and end with four examples. However, the examples only distinguish speeches from articles that quote or criticize the candidate. None of them covers a surrogate speaker or an announced event, which are the two most important cases.

G.2 Differences in the Resulting Measure

Table G1 classifies the 7,110 articles that the GPT-4o labels call speeches and the current definition excludes. Three quarters of them describe a speech that was announced for a later date or had been given earlier. In one in six a speech occurred but the speaker was not the candidate, and in most of these cases the current labels identify a surrogate. The remaining tenth report no speech-like event at all. The two measures therefore differ mainly on whether the speech had occurred, and less on who delivered it. A single yes-or-no question cannot capture this distinction.

⁴⁰Model version `gpt-4o-2024-08-06`.

Table G1: Why the Current Labels Drop an Article the GPT-4o Labels Called a Speech

Current labels' reading of the article	Articles	Share
A speech is announced for the future, not yet delivered	3,221	45.3%
The article refers back to a speech given earlier	2,090	29.4%
A speech occurred, but the speaker was not the candidate	1,141	16.0%
<i>of which the speaker is an identified surrogate</i>	960	13.5%
No speech-like event is reported at all	658	9.3%
Total dropped	7,110	100%

Notes: The 15,679 article–candidate observations that the GPT-4o labels classify as reporting a speech, of which the current definition keeps 8,569 and excludes the 7,110 shown. The categories come from the current labels and not from an independent review. Rows are mutually exclusive and use the article-level classification. An article can contain several events, and the current definition keeps it if any single event is a delivered speech by the candidate.

Each set of labels counts articles that the other leaves out. The GPT-4o labels classify 15,679 articles as speeches, against 9,503 under the current definition. However, the broadest definition under the current labels, which counts any type of event by any speaker, classifies 19,484 articles as speeches (Appendix C.2), more than the GPT-4o labels do. Most of the GPT-4o speeches fall under this broad definition, but 5,707 articles that the GPT-4o labels exclude report a speech-like event under it. In total, 6.2 percent of the articles change classification between the two. Switching to the current labels therefore drops some articles and picks up others.

Differences by campaign. The differences between the two measures vary across campaigns, and this variation is the main reason for the change of labels. Table G2 counts the candidate–county pairs with at least one speech under each set of labels. The current labels keep four fifths of these pairs in Bryan’s 1896 tour and in the 1908 campaign of Bryan and Taft, but only about one in eight in 1888 and one in three in 1892. The differences follow *campaign activity* rather than the calendar. In 1904, a late year in which neither Roosevelt nor Parker campaigned actively, the current labels keep a share close to that of 1892. This pattern is consistent with Table G1: when the candidate does not travel, the articles that mention him are more often an announcement, a reference to an earlier speech, or a rally held on his behalf.

Table G2: Treated Candidate–County Cells by Election Year, Both Label Sets

	1880	1884	1888	1892	1896	1900	1904	1908	1912	1916	Total
GPT-4o	33	126	60	76	449	262	40	471	248	278	2,043
GPT-5.6-terra	19	99	8	27	371	183	15	381	167	195	1,465
Retained	58%	79%	13%	36%	83%	70%	38%	81%	67%	70%	72%

Notes: A cell is a candidate–county pair in an election year with at least one detected speech, counting the two major-party candidates separately, so a county visited by both contributes two. GPT-5.6-terra is the paper’s measure, described in Appendix C.1. The denominator is 5,656 cells per year (2,828 counties $\times$ 2 major-party candidates), and 8,484 in 1892 and 1912, when a third candidate is labeled; the third candidate’s cells are excluded from both rows. This is not the *speech-origin* county count that identifies the coverage regressions of Section 5, which collapses the candidate dimension: 1908, for instance, is 381 cells across 298 counties, because 83 counties hosted both Bryan and Taft.

The same pattern affects the paper’s main descriptive fact. Between 1892 and 1896, the number

of candidate–county pairs with a speech increases by a factor of 5.9 under the GPT-4o labels and by a factor of 13.7 under the current labels. The GPT-4o labels would therefore have understated the transition from front-porch campaigning to touring.

Examples. The six newspaper pages in Figures 5 and C4 illustrate the same comparison. The two sets of labels agree on the three panels of Figure 5 and disagree on panels (a)–(c) of Figure C4, where the GPT-4o label is the incorrect one in all three. Two are false positives. Panel (b) is a Topeka meeting three weeks away, which the GPT-4o labels read as a delivered speech in Shawnee County, Kansas. Panel (c) is a Los Angeles rally addressed by Senator White on Bryan’s behalf. One is a false negative. Panel (a) is the list of Bryan’s twenty-one Iowa speeches of October 31, in which the digitized text joins Albia with the neighboring towns, so the GPT-4o labels assign no speech to Monroe County, Iowa. Panels (d)–(f) are errors of the current labels and play no role in this comparison. A fourth error appears in no figure. For the speech at Clinton, Tennessee, the GPT-4o labels return Clinton, *Kentucky*, and assign it to Hickman County, while the current labels return the town without a state and assign it to no county. The current labels get the first three cases right and leave the fourth unassigned.

G.3 Results under the GPT-4o Labels

The two main estimates of the paper hold under the GPT-4o labels. Table G3 repeats the coverage regression of Table 1. Table G4 repeats the media reach regression of Table 2 with the outside audience rebuilt from the coefficients of Table G3, and displays all three coefficients of the equation, as Table E1 does. The specification, the period, the sample rules, and the definition of connection are those of the main text, so only the speech labels change. The sign, the significance, and the shape of the distance gradient are the same in every column of both tables.

Table G3: Coverage of Presidential Speeches, GPT-4o Labels

Dep. Var.: 1[Coverage > 0] x 100	(1)	(2)	(3)	(4)
Both endpoints wired	1.48*** (0.07)	1.80*** (0.08)	1.83*** (0.15)	1.99*** (0.16)
Distance (logs), demeaned		-0.15*** (0.02)		
Both endpoints wired x Distance (logs)		-1.44*** (0.12)		-0.87*** (0.22)
Same state	1.05*** (0.06)	0.69*** (0.06)		
Origin x Year FE	X	X	X	X
Destination x Year FE	X	X	X	X
Origin x Destination FE			X	X
Observations	1,377,285	1,377,285	1,015,605	1,015,605
R-squared	0.157	0.159	0.433	0.433

Notes: As in Table 1, estimated on the GPT-4o speech labels. Distance is the straight-line distance between the two counties’ centroids, demeaned, so the coefficient on “Both endpoints wired” in column (4) is the effect at the sample mean log distance.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by speech county x coverage county pairs.

The coverage estimates are larger under the GPT-4o labels because those labels place speeches in more counties. The connection effect at the average log distance is about a quarter larger than in Table 1, and the distance interaction is steeper by half. The extra origin counties give the regression nearly half again as many county pairs, and they are concentrated in the campaigns in which the candidate stayed home (Table G2). A blind re-reading of a sample of these articles indicates that they are mostly surrogate rallies and announcements rather than places where the candidate spoke. The GPT-4o measure is therefore closer to “a campaign event was reported here” than to “the candidate spoke here”, and the difference is largest before 1896.

The media reach estimates barely move. The coefficient on media reach, in the first row of Table G4, is within a tenth of its value in Table 2 in every column, on the same estimation sample, since the county-year panel does not depend on the labels and only the outcome and the outside audience change. However, the coefficient on the outside audience among unconnected counties, in the second row, which Appendix E.1 uses as a test of the design and finds indistinguishable from zero under the current labels, is positive and significant at 5 percent in five of the six specifications under the GPT-4o labels. This pattern is consistent with the old measure counting rallies and announcements in the larger places near the network, whether or not the county itself was connected.

Table G4: Own Connection versus Reachable Audience Elsewhere, GPT-4o Labels

Dep. Var.: $1[\text{Speech} > 0] \times 100$	(1) Baseline	(2) Baseline + Controls	(3) 100-Mile Buffer	(4) Fixed 1880 Population	(5) All Aged 21+	(6) No Capital/Hub Destinations
Media reach (Reachable voters $\times$ Local wire)	2.66*** (0.66)	2.20*** (0.65)	2.08*** (0.63)	1.82** (0.71)	2.15*** (0.65)	2.12*** (0.63)
Reachable voters (std)	3.55*** (1.38)	4.01*** (1.49)	-4.05*** (1.48)	3.68*** (1.35)	4.00*** (1.46)	3.07** (1.54)
Local wire	3.18*** (1.08)	2.60** (1.03)	2.47** (1.04)	2.98*** (1.03)	2.63** (1.03)	2.59** (1.03)
Wired-origin response	6.21 (1.48)	6.21 (1.59)	-1.97 (1.57)	5.50 (1.44)	6.15 (1.56)	5.19 (1.63)
Controls	None	All	All	All	All	All
County FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Observations	23,048	23,048	23,048	23,048	23,048	23,048
R-squared	0.361	0.366	0.366	0.366	0.366	0.366

Notes: As in Table E1, with the dependent variable built from the GPT-4o speech labels and the outside audience built from the coefficients of Table G3. Each column estimates equation 3 and displays all three coefficients. The first row is the coefficient that Table 2 reports, with the outside audience standardized over all counties as there. The columns follow those of Table 2. Columns (2) through (6) add its controls, and columns (3) through (6) recalculate the outside audience as described there. The coefficient on reachable voters is the association between the outside audience and visits among counties that were not connected, where a speech could not travel, and should be zero under the mechanism. The wired-origin response is the sum of the two. The estimation sample is that of Table 2, because the county-year panel does not depend on the speech labels.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by county.