Information Access in 19th Century America

Levi Boxell, *Stanford University*[*]

January 2019

[Preliminary and Incomplete]

Abstract

The telegraph fundamentally altered the information environment in 19th century America. To examine this change, I build a novel dataset that documents the early growth of the telegraph and measures the time delays in the diffusion of information between cities. Using this dataset and Donaldson's (2018) method of estimating transportation costs, I estimate the speed at which information travels across various types of 19th century infrastructure.

[*]E-mail: E-mail: lboxell@stanford.edu. I would like to thank Matthew Gentzkow, Aaron Honsowetz, Richard Hornbeck, Jesse Shapiro, and participants at the 2017 EHA Conference for comments and suggestions. I acknowledge funding from the National Science Foundation (grant number: DGE-1656518), the Institute for Humane Studies, the History Project at Harvard University, and the Institute for New Economic Thinking. *Please note that this is a work in progress and that errors may be present within this draft. Do not cite without permission.*

"Have you any news" – Samuel Morse,
1844 Inaguration of the Washington-Baltimore Telegraph Line

1 Introduction

The telegraph fundamentally reshaped the information environment in 19th century America. Instead of being restricted to reporting on local events, nationwide news could easily be reported with minimal delay (Blondheim 1994, 38). The nationalization of news reporting and the associated fixed costs quickly gave rise to news wire services that provided an information exchange on events of national interest to all subscribing newspapers (Sotirovic and McLeod 2009). What was the quantitative impact of the telegraph on the news environment? And, how did these changes impact electoral behavior?

To answer these questions, I construct two novel datasets. Using historical maps, I construct a georeferenced dataset documenting the initial expansion of the telegraph network between 1844 and 1852. I then combine the telegraph network with readily available data on transportation networks during the same period which measures the extent of the communication network at each point in time. In addition, I use recently digitized newspapers to construct a dataset on the origin location and date along with the reporting destination for several thousand newsworthy events during the same period. This dataset gives a measurement of the time it takes for an event in one location to be reported in another location—that is, a measurement of information access between the two locations.

To allow for spillovers and heterogeneity in the impact of the telegraph, I directly model information flows as having a fixed cost and a variable cost dependent on the type of communication infrastructure required to traverse between the two locations. I then estimate the speed at which information travels across each type of infrastructure by nesting Dijkstra's least-cost-route algorithm within a nonlinear least squares optimization problem. This is similar to Donaldson's (2018) approach for estimating transportation costs across various types of infrastructure in 19th century British India. Given these parameter estimates, I obtain a prediction of the time it takes for information to travel between any two locations within the network. The results suggest that between 1844 and 1848, areas in the midwest experienced a decrease in the length of time it takes for news to reach them from Washington, DC of several days. This gives a quantitative measure of the impact of the telegraph, and infrastructure developments more generally, on access to national politican information.

The (estimated) time it takes for news from Washington, DC to reach counties in the United States is a measure of access to national political information. This can then used to examine the impact of the telegraph-induced changes in information access on county-level voting outcomes. Examining the impact of the telegraph and information access on voting outcomes would build on a growing literature examining the impact of various information technologies on voter behavior and beliefs, including the radio (e.g., Stromberg 2004), the television (e.g., Gentzkow 2006; Martin and Yurukoglu 2017), and the internet (e.g., Lelkes et al. 2017; Boxell et al. 2017; Boxell et al. 2018; Campante et al. 2018). With respect to the telegraph, Garcia-Jimeno et al. (2018) examine the diffusion of Temperance protest activity in the late 19th century and argue that the railroad and telegraph networks helped facilitate this diffusion. My measurement of information access and empirical approach also provide a methodological contribution to literature on information impacts more broadly by directly modelling and estimating the spillovers and heterogeneity. Lastly, the novel datasets collected provide a useful starting point for future research examining the impact of the telegraph on political and economic outcomes.

The remainder of the paper is outlined as follows. Section 2 describes the data. Section 3 outlines the measurement and estimation of information access. Section 4 concludes.

2 Data

2.1 Communications Infrastructure

2.1.1 Telegraph Stations

I build a data set of the location of telegraph stations at three points in time: 1844, 1848, and 1852. In 1844, there were essentially no telegraph stations. The first test line between Baltimore and Washington, DC was not built until May 1844 by Alfred Vail and Samuel Morse after getting approval for funding from Congress. I therefore use 1844 as a base year with no telegraph stations. In subsequent years, the telegraph network grew quickly.

To determine the location of these telegraph stations, I digitize two maps. The first map from 1848 exhibits the lines of Morse's electro-magnetic telegraph. It documents telegraph lines that were in operation, being constructed, and proposed. It depicts the lines and labels the location of telegraph stations along these lines. A key feature of the telegraph network is that a station or receiver is required to send and receive messages. Furthermore, the telegraph network was a connected network. That is, for the vast majority of the time, a message from one station could be sent to any other station.[1] Lastly, information on the telegraph network travels along the lines at near instantaneous speeds.[2] Therefore, only information about the location of the telegraph stations is required and one can assume that each telegraph station is connected to all other stations. Using the 1848 map, I obtain 64 cities with a telegraph station whose locations I then geocode.I then repeat this procedure using the 1853 Barr map of telegraph stations in the United States, Canada, and Nova Scotia. This map gives 635 cities with telegraph stations which I then geocode for locations.

2.1.2 Other Infrastructure

In addition to communicating across the telegraph network, transportation infrastructure was frequently used to send news and information. For the railroad network, I use the railroad lines defined to be in operation for each year and assume information can be transhipped on and off the line at any point (Atack 2016).[3] I also utilize information on all canals and steam-boat navigable rivers for each year and make similar assumptions regarding transhipment at any point. Additionally, oceanic travel was another important source of news. For this, I utilize the coastline as a separate, ocean-based communication network. Lastly, to help conntect the communication network, I construct a road network that is defined to be straight-line connections between all locations within 40 kilometers of each other.[4]

I define my locations of interest over which the communication network is constructed to be all locations with telegraph stations, all 1850 county centroids, and cities for which I have information flows data.

Figure 1 depicts the growth of the telegraph and transportation networks across my time period. While there is growth in the railroad network and canal network during this time, these are largely preceded by and overshadowed by the rapid growth of the telegraph network. The rapid diffusion of the telegraph network can be explained by its low cost to build relative to rail and canal networks. A given mile of the telegraph network is estimated to have cost $150 to

[1] While lines may have been temporarily unconnected during their construction, the large incentives for having an entirely connected telegraph network minimized these gaps between lines.

[2] While the low power of the early systems often required re-sending of information for long distance messages, these costs were minimal relative to other communication network options.

[3] The railroad, river, and canal networks were downloaded from Jeremy Atack's website https://my.vanderbilt.edu/jeremyatack/data-downloads/ on May 21, 2016.

[4] This creates a fully connected network for the eastern portion of the United States, which is the focus of this paper.

construct (DuBoff 1980), whereas a given mile of railroad would often cost over $2,000 (Wicker 1960).

2.2 Information Flows

An important component for estimating how quickly information moves across different types of communication infrastructure is data on information flows between locations—particularly the location and time of origin along with the location and time of destination. Given the relatively slow pace of information flows in 19th century America, newspapers frequently cited the date and origin of the news they printed. This allows for systematically collecting information on the time delays between which a news event occurred in one location and when it was reported in another location.

To do this, I focus on five different, daily newspapers: The New York Herald (New York, NY; 1844, 1848, and 1852), The Daily Crescent (New Orleans, LA; 1848 and 1852), The Evansville Daily Journal (Evansville, Indiana; 1848 and 1852), The Daily Union (Washington, DC; 1848 and 1852), and The American Republican and Baltimore Daily Clipper (Baltimore, MD; 1844). I then collect the location of origin, the date of origin, the type of news, and the date of printing for events reported in each newspaper-year combination between October 15 and November 14 (inclusive). For many events, only the city is reported rather than the city and state. For these events, the state is imputed if clear. I exclude events with an unclear origin state along with those originating in foreign locations, California, Texas, Oregon, and a few other remote locations in the United States.[5]

3 Measuring Political Information Access

To account for the heterogeneous impact and spillovers from the construction of telegraph stations, I take a more structural approach of modelling how information flows across infrastructure and then estimate this model using the data on time delays in news reporting between locations. This approach is similar in spirit to Donaldson's (2018) estimation of transportation costs across different types of infrastructure in 19th century India. This also builds on Pred's (1973) analysis of pre-telegraphic information delays in the United States.

[5]"Key West, FL, USA," "Havre, MT, USA," "Arendahl, MN, USA," and "Scott County, IA, USA." I also drop events which have negative time delays, e.g., due to printing errors, and time delays that are greater than 30 days.

The time it takes for a given news event originating in location i to be reported in location j is modelled as

$$d_{ijt} = \alpha_t + \omega_{ij} + \beta \begin{pmatrix} x_{ijt}^{tele}(\gamma) \\ x_{ijt}^{rail}(\gamma) \\ x_{ijt}^{canal}(\gamma) \\ x_{ijt}^{river}(\gamma) \\ x_{ijt}^{ocean}(\gamma) \\ x_{ijt}^{road}(\gamma) \end{pmatrix} \cdot \boldsymbol{\gamma} + \epsilon_{ijt} \tag{1}$$

where d_{ijt} is the minimum time delay (in days) of news reporting between locations i and j in year t, α_t are year fixed effects, ω_{ij} are origin-destination pair fixed effects, $\boldsymbol{\gamma}$ is a vector of relative travel time costs (where I normalize road travel speeds to be one, assume the telegraph has instantaneous travel speeds, and restrict rivers and canals to have the same trave speeds), $\mathbf{x}_{ijt}(\gamma)$ is a vector containing distance on each type of infrastructure required for information to travel between locations i and j as a function of the parameter vector γ, and β is a scalar coefficient. Since time delays in news reporting are endogenous, I use the minimum time delay for each origin-destination-year triplet in order to capture how quickly information *could* travel on a given type of infrastructure.

A key input in estimating this model is the vector of travel distances $\mathbf{x}_{ijt}(\gamma)$. The time delay data does not include this information. However, conditional on a set of parameter values γ, the least cost route between any two locations can be calculated using the communication network and Dijkstra's algorithm. This gives rise to a nonlinear least squares problem which I solve using the following steps:

1. Choose $\boldsymbol{\gamma}$.
2. Determine $\mathbf{x}_{ijt}(\gamma)$.
   (a) Represent infrastructure as a weighted graph (nodes and edges).
   (b) Get least-cost path from i to j using Dijkstra's algorithm for graphs.
3. Estimate α_t, ω_{ij}, and β via OLS.
4. Repeat steps 1-3 and choose $\boldsymbol{\gamma}$ to minimize $\sum_{ijt} \epsilon_{ijt}^2$.

Due to the computationally intensitve nature of this routine, I perform a grid search over potential parameter values to determine γ.[6] With the origin-destination fixed effects, the identifying variation for γ and β will be coming from infrastructure changes overtime that alter the vector of travel distances $\mathbf{x}_{ijt}(\gamma)$ required to traverse the two locations optimally. The estimates values for γ are 2 for railroads, 2.3 for rivers and canals, and 2 for ocean. These estimates suggest that these infrastructure improvements increase information speed by large margins.

Given the model estimates, I define the *political information access* of county j in year t to be

$$\hat{d}_{kjt} = \hat{\alpha}_t + \hat{\beta}\mathbf{x}_{ijt}(\hat{\gamma}) \cdot \hat{\boldsymbol{\gamma}}.$$

where k denotes Washington, DC. This is the model's prediction of the time it would take for information from Washington, DC to reach county j driven by changes in the infrastructure and annual shocks. Note that since the data used to estimate the model does not include points in every county, I have to exclude the pair fixed effects from the measure.

Importantly, this will be a notion of access to *national* political information rather than local political information. Figure 2 shows the change in political information access between 1844 and 1848 along with between 1848 and 1852. By 1848, regions in the midwest had experienced a decrease of several days in the time it takes for information from Washington, DC to be printed locally. The northeast region, which also experienced rapid diffusion of the telegraph by 1848, experienced a much smaller change in information access due to their higher baseline levels of information access. By 1852, these gains in information access began spreading to the south as well. As already noted in figure 1, these changes are going to largely driven by the rapid diffusion of the telegraph network.

4 Conclusion

Future work should continue examining the impact of the telegraph and information access more broadly on related political and economic outcomes—the empirical strategy and associated datasets in this paper provide a useful starting point for this future work.

[6]These parameter values are $\{2, 3, 6, 9, 12, 15\}$ for railroads, the same for ocean, and $\{1.75, 2.3, 2.7, 3, 3.3, 4\}$ for rivers/canals. I then estimate the model on all parameter triplets.

References

Atack, Jeremy. 2016. Historical Geographic Information Systems (GIS) database of U.S. Railroads, Steamboat Navigable Rivers, and Canals during the 19th Century in the United States.

Blondheim, Menahem. 1994. *News over the wires: The telegraph and teh flow of public information in America, 1844–1897.* Harvard Studies in Business History. Harvard University Press.

Boxell, Levi, Matthew Gentzkow and Jesse M. Shapiro. 2017. Greater internet use is not associated with faster growth in political polarization among US demographic groups. *Proceedings of the National Academy of Sciences.* 114(4): 10612–10617.

Boxell, Levi, Matthew Gentzkow and Jesse M. Shapiro. 2018. A note on internet use and the 2016 U.S presidential election outcome. *PLoS ONE*. 13(7): e0199571.

Campante, Filipe, Ruben Durante, and Francesco Sobbrio. 2018. Politics 2.0: The multifaceted effect of broadband internet on political participation. *Journal of the European Economic Association*. 16(4): 1094–1136.

Donaldson, Dave. 2018. Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review.* 108(4-5): 899–934.

DuBoff, Richard B. 1980. Business demand and the development of the telegraph in the United States, 1844–1860. *Business History Review*. 54(4): 459–479.

Garcia-Jimeno, Camilo, Angel Iglesias, and Pinar Yildirim. 2018. Women, rails, and telegraphs: An empirical study of information diffusion and collective action. *NBER Working Paper No. 24495.*

Gentzkow, Matthew. 2006. Television and voter turnout. *Quarterly Journal of Economics*. 71(3): 931–972.

Gentzkow, Matthew, Jesse M. Shapiro, and Michael Sinkinson. 2011. The effect of newspaper entry and exit on electoral politics. *American Economic Review*. 101(7): 2980–3018.

Lelkes, Yphtach, Gaurav Sood, and Shanto Iyengar. 2015. The hostile audience: The effect of access to broadband internet on partisan affect. *American Journal of Political Science*. 61(1): 5–20.

Martin, Gregory J. and Ali Yurukoglu. 2017. Bias in cable news: Persuasion and polarization. *American Economic Review*. 107(9): 2565–2599.

Pred, Allan R. 1973. *Urban growth and the circulation of information: The United States system of cities, 1790–1840.* Harvard University Press: Cambridge, MA.

Sotirovic, Mira and Jack McLeod. 2009. Media coverage of U.S. elections: Persistence of tradition. In, *The handbook of election news coverage around the world.*

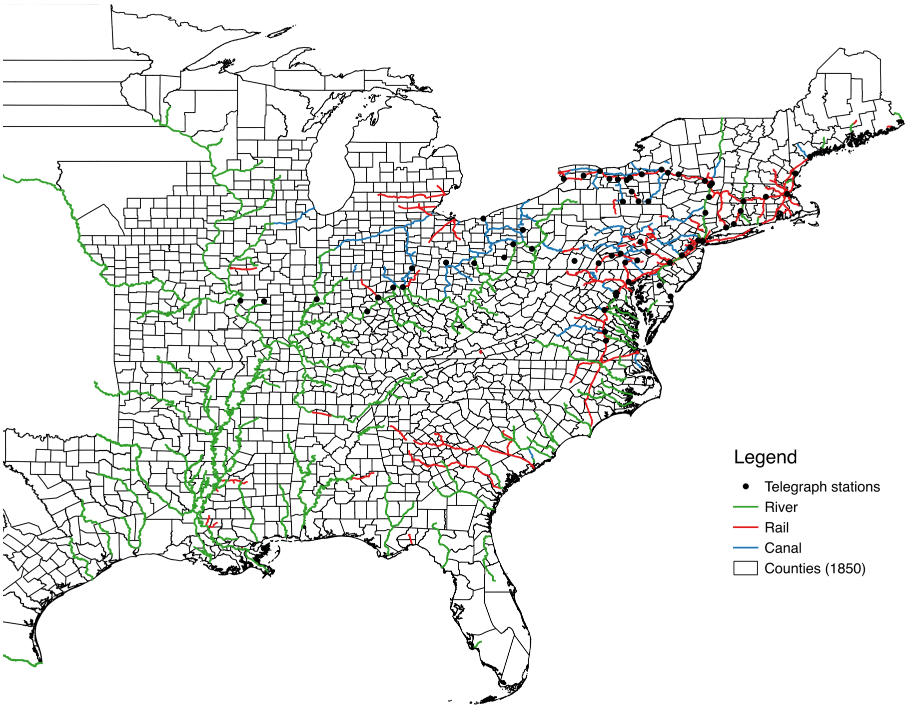
Stromberg, David. 2004. Radio's impact on public spending. *Quarterly Journal of Economics*.

119(1): 189–221.

Wicker, E.R. 1960. Railroad investment before the Civil War. In, *Trends in the American economy in the nineteenth century*. Princeton University Press.

Figure 1: 19th Century Communications Infrastructure

Panel A: 1848



Panel B: 1852

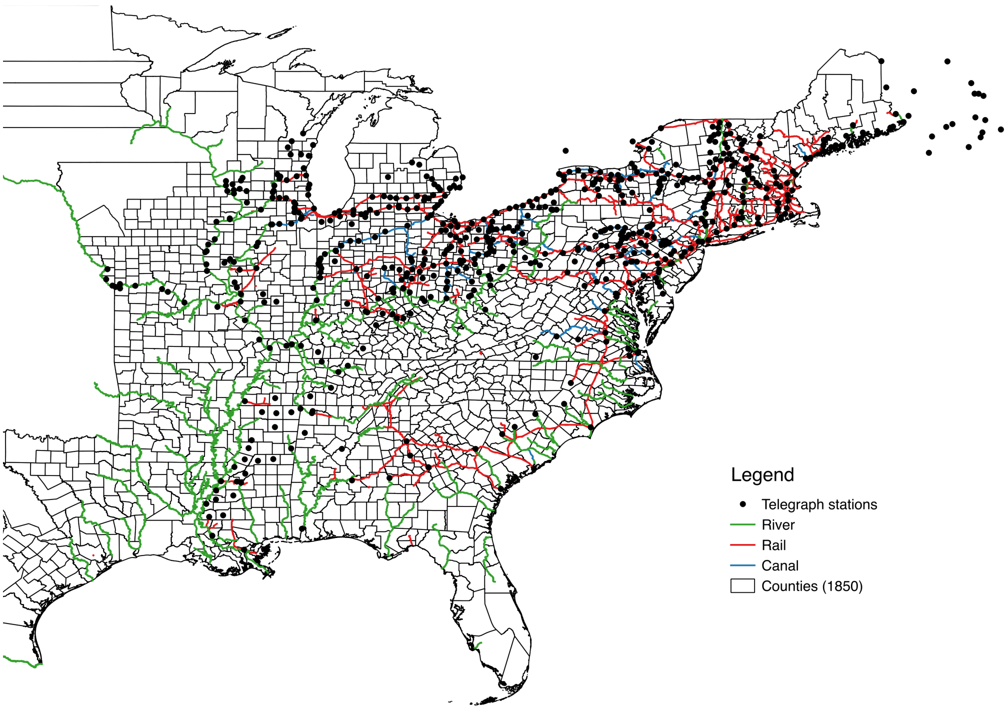
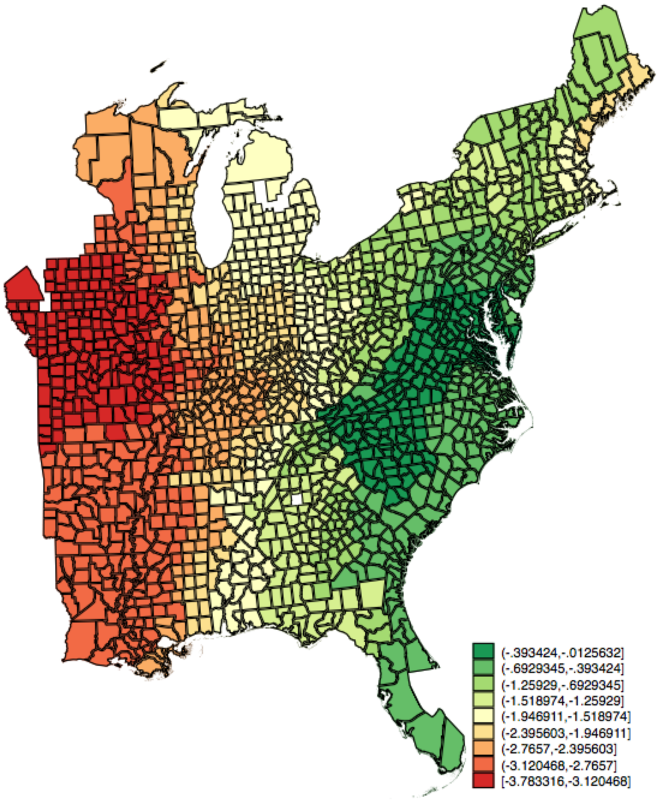


Figure 2: Change in Information Access

Panel A: 1844–1848



Panel B: 1848–1852