Reassessing Railroads and Growth:

Accounting for Transport Network Endogeneity

Scott N. Swisher IV*

May 2014

Abstract

Motivated by the seminal work of Robert Fogel on U.S. railroads, I reformulate Fogel's original counterfactual history question on 19th century U.S. economic growth without railroads by treating the transport network as an endogenous equilibrium object. I quantify the effect of the railroad on U.S. growth from its introduction in 1830 to 1861. Specifically, I estimate the output loss in a counterfactual world without the technology to build railroads, but retaining the ability to construct the next-best alternative of canals. My main contribution is to endogenize the counterfactual canal network through a decentralized network formation game played by profit-maximizing transport firms. I perform a similar exercise in a world without canals. My counterfactual differs from Fogel's in three main ways: I develop a structural model of transport link costs that takes heterogeneity in geography into account to determine the cost of unobserved links, the output distribution is determined in the model as a function of transport costs, and the transport network is endogenized as a stable result of a particular network formation game. I find that railroads and canals are strategic complements, not strategic substitutes. Therefore, the output loss can be quite acute when one or the other is missing from the economy. In the set of Nash stable networks, relative to the factual world, the median value of output is 45% lower in the canals only counterfactual and 49% lower in the railroads only counterfactual. With only one of the transportation technologies available, inequality in output across cities would have been lower in variance terms but sharply higher in terms of the maximum-minimum gap. Such a stark output loss is due to two main mechanisms: inefficiency of the decentralized equilibrium due to network externalities and complementarities due to spatial heterogeneity in costs across the two transport modes.

JEL Classification Codes: E22, O11, N71, L92, R42

Keywords: economic growth, transport infrastructure, network formation games, strategic complements, railroads, counterfactual history, multiple equilibria, computation, simulation

*Department of Economics, University of Wisconsin-Madison. I would like to acknowledge my main advisor Steven Durlauf and dissertation committee members Dean Corbae, Michael Gofman, and Randall Wright for their helpful comments. I also acknowledge constructive comments from Charles Engel, Fran Flanagan, Kenichi Fukushima, Theresa Gutberlet, Joseph Han, Wontae Han, Erik Hembre, Kevin Hutchinson, Bjorn Jorgensen, Hsuan-Chih Lin, Kurt Lunsford, Jaime Luque, Irina Merkurieva, Kayuna Nakajima, Dan Quint, Ariel Roginsky, Marzena Rostek, Ananth Seshadri, Teng Kok Tan, Noah Williams, Cheng-Ying Yang, Jake Zhao, and Yu Zhu. I would also like to thank helpful seminar participants at Cambridge University, Oxford University, Virginia Commonwealth University, Wisconsin Macroeconomics Working Group, UW-Whitewater, Midwest Economics Association, and Wisconsin Economics Association. Contact information: e-mail snswisher@gmail.com. Any remaining errors are my own.

1 Introduction

The United States experienced rapid but volatile economic growth in 19th century which was correlated with the expansion of transportation networks, specifically railroads and canals. I present evidence on the concurrence of output growth and transport network expansion in the 19th century U.S. case in Figure 1 (Appendix, part A), in levels for output and transport network mileage, and Figure 2, in growth rates; 1861 is highlighted as the year used for the empirical work in this paper.[1] Note that the rate of railroad network expansion slowed after 1860 and railroad mileage growth is positively correlated with output growth. I am primarily interested in estimating the empirical effect of transport networks on economic growth. Note the eventual dominance of railroads over canals via competition and technological change in Figure 3, and the resulting price gap between the two transport modes in Figure 4 due to the service quality advantage of railroads. Specifically, I estimate the effect of railroads on U.S. economic growth in the 19th century from 1830 to 1861 and the associated output gain due to railroads. Alternatively this is equivalent to measuring the output loss when railroads are removed from the technology set of the economy. Although the previous literature found that one could use additional canals relative to the observed canal network to replicate the effect of railroads, I find that the output loss is large when either railroads or canals are removed. Thus I conclude that both railroads and canals were essential to 19th century U.S. economic growth.

Relative to the previous literature (Fogel, 1964; Donaldson and Hornbeck, 2013), the main contribution of this paper is to endogenize the counterfactual canal network when railroads are unavailable through a decentralized network formation process that uses costs derived from the data. I develop a computational strategic model of transport network formation in which transport firms choose where to build links in order to maximize their individual profits. For computational tractability, I employ a parsimonious gravity trade model to determine the relevant firm-level payoffs. Specifically, observed transport networks are used to bound type-specific link costs for railroad and canal firms; the model is then estimated using these cost bounds. Then I employ the estimated model to run counterfactuals: Fogel's question of canals only without railroads discussed extensively in the literature, and the reverse scenario with railroads only which is novel. Due to multiplicity of equilibria, I provide both a point estimate and associated 90% coverage interval for the marginal output gain due to railroads, such that 90% of equilibria are associated with an output loss within the coverage interval. The model can also be used to answer policy questions, such as the optimal general subsidy for transport firms or optimal subsidies for targeted links that the policymaker wants to build in the network.

One of the mechanisms driving the large output gain due to both railroads and canals is that construc-

[1] Historical U.S. time-series data from Carter et al. (2006).

tion of a transport link has positive external effects on other agents in the economy that are not captured by the firm building the link. Specifically, addition of a transport link lowers trade costs globally and increases the magnitude of trade flows that do not necessarily utilize the new link directly.[2] Thus the firm does not fully capture the gains from trade facilitated by the link and one would expect underbuilding of transport links in a decentralized equilibrium relative to what a social planner would choose. Therefore the First Welfare Theorem does not hold in this environment since the planner's solution differs from the decentralization, although it may be possible to decentralize the planner's solution by subsidizing link construction. Since an inefficiently low number of links are built in the decentralized economy, this inefficiency may be mitigated or amplified when fewer transport technologies are available. I find that, due to complementarities between railroads and canals, the underbuilding inefficiency is amplified when only railroads or canals are available, contributing to large output losses in such counterfactuals. For parsimony, I treat transport network expansion as only source of growth, thus one can interpret these estimates as an upper bound on the true output loss due to the removal of railroads from the economy.

The key friction in the model is that trade is costly across space; even if trade is possible, it might be very costly due to high transport costs between source and destination cities. Additional transport links lower trade costs and increase trade flows and output. In the decentralized equilibrium and associated stable transportation networks, transport firms will not build additional links if the revenue they capture through traffic over their own links is insufficient to cover costs. I find evidence of complementarities between canals and railroads which are not directly observable but rather emerge naturally from cost heterogeneity across types: when one technology is missing from the economy, firms operating the other see a substantial decline in profits. Therefore, they choose to build fewer links, so fewer cities are connected and the remaining connected cities trade at a higher cost. Increased trade costs lower trade flows and output in the counterfactuals when one of the two transport modes, railroads or canals, is never invented. These complementarities are a result of cost heterogeneity due to spatial variation in geography, which is not assumed but emerges empirically from the structural model of link construction costs developed in the paper.

In the ongoing debate on the contribution of railroads to U.S. economic growth, historians such as Jenks (1944) initially claimed that railroads were essential for industrialization and thus indispensable for growth, although they could not quantify the output gain due to railroads alone. Fogel (1964) and the subsequent literature (Fishlow, 1965; Donaldson and Hornbeck, 2013) found that the 1890 railroad network could have been replicated with more canals, relative to the observed canal network, at a low cost to society. Thus they conclude that the output loss associated with removing railroads is small, on the order of 3-15% of

[2]Trade flows for which the new link is not on the least-cost path between source and destination cities may increase as a result of link construction, but the firm building the new link sees no direct benefit from these additional flows.

1890 output. However, I find that canals would not have filled the gap left after the removal of railroads since the canal network would have developed differently in the absence of railroads. Therefore, canal firms find it unprofitable to replicate the observed railroad network with canals alone. Canals and railroads are complements, thus removing railroads disincentivizes canal construction since transport costs increase and output falls, lowering revenue for canal firms.

A discontinuous innovation in the transportation sector, such as the development of a new class of infrastructure, could be interpreted as the introduction of a new type of capital goods. In the canonical growth model, physical capital could be augmented with an array of "transportation capital" types that enter the production function in the usual way.[3] This is analogous to the introduction of human capital without spillovers in Lucas (1988). After the one-time innovation, agents can now invest in and accumulate a new type of transport capital that obeys the standard assumptions.[4] Such an innovation implies that the resulting above-trend growth rate is high initially and monotonically decreasing to the new steady state, and thus observationally equivalent to a positive productivity shock.[5] In order to capture the unique characteristics of transport capital, such as the nonlinear effect of building an additional link on global transport costs due to network externalities, I develop a model of economic growth where the transport sector plays a central role through changes in the economy's transport network.

The premise of this paper is that, using a network micro-structure, it is possible to say something quantitative about historical growth episodes where transport networks rapidly expanded, specifically the U.S. experience in the 19th century with railroads. After the introduction of a new class of transport capital goods, I find a positive relationship between transport network expansion and economic growth. Direct modeling of the individual transport firm's profit motive in determining the network structure captures nonlinearities in link formation and network effects. Households and firms in cities that produce tradable goods are passive demanders of transport services. Output depends on the matrix of bilateral transportation costs derived from the transport network as least-cost routes between locations. Thus the resulting set of stable networks depends on relative costs for railroads and canals, which are determined in the model based on observed transport networks.

Motivated by the seminal work of Robert Fogel on U.S. railroads, I reformulate Fogel's original counterfactual history question on 19th century U.S. economic growth without railroads by treating the transport network as an endogenous equilibrium object. I quantify the effect of the railroad on U.S. growth from its introduction in 1830 to 1861. Specifically, I estimate the output loss in a counterfactual world without

[3]That is, the production function is increasing and concave in each type of transport capital, with the relevant cross-partial sign restrictions to guarantee the existence of an interior solution.

[4]Provided that the economy starts with some nonzero initial level, part of the innovator's endowment.

[5]Of course, observational equivalence only holds if transport capital (i.e. the transport network) is not directly observed, with the econometrician restricted to standard macroeconomic aggregates such as output, capital, labor, and the like.

the technology to build railroads, but retaining the ability to construct the next-best alternative of canals. My main contribution is to endogenize the counterfactual canal network through a decentralized network formation game played by profit-maximizing transport firms. I perform a similar exercise in a world without canals. My counterfactual differs from Fogel's in three main ways: I develop a structural model of transport link costs that takes heterogeneity in geography into account to determine the cost of unobserved links, the output distribution is determined in the model as a function of transport costs, and the transport network is endogenized as a stable result of a particular network formation game.

I find that railroads and canals are strategic complements, not strategic substitutes. Therefore, the output loss can be quite acute when one or the other is missing from the economy. See Table 8 (Appendix, part B) for the main results. In the set of stable networks, relative to the 1861 factual world, I find that output is 45% lower in the canals only counterfactual and 49% lower in the railroads only counterfactual on average. I also find that, relative to the 1861 factual world, in the counterfactual with canals only the variance of output decreases by 69%, the number of links built by canal firms increases by 508%, the number of active canal firms increases by 43%, and total canal firm revenue increases by 716%. In the counterfactual with railroads only relative to the observed world, I find that the variance of output also decreases by 69%, but the number of links built by railroad firms decreases by 45%, the number of active railroad firms decreases by 83%, and total railroad firm revenue increases by 7%. With only one of the two transportation technologies available, inequality across U.S. cities would have been lower in variance terms but sharply higher in terms of the maximum-minimum gap. Such a stark output loss is due to two main mechanisms: inefficiency of the decentralized equilibrium due to network externalities and complementarities due to spatial heterogeneity in costs across the two transport modes.

Finally, I compare the output loss results in the counterfactuals to previous efforts to measure the effect of railroads in the literature. For example, Fogel (1964) estimates a counterfactual output loss of 3-5% of 1890 U.S. output, Fishlow (1965) finds a larger output loss in the 5-15% range, and Donaldson and Hornbeck (2013) arrive at a loss of 3.4% similar to Fogel. I find a substantially larger output loss of 45% relative to 1861 U.S. output when only canals are available. Although it is difficult to make a direct comparison due to different base years for the analysis, regarding Fogel's work and the previous axiom of indispensability he found lacking merit, mainly that railroads were "indispensable" to economic growth in the United States, I conclude in opposition that both railroads and canals were indispensable for U.S. economic growth in the 19th century.[6]

There are three main channels that can explain the difference between the findings in this paper and

[6] Although my base year is earlier, 1861 compared to 1890, one would expect that starting earlier would tend to underestimate the growth contribution of railroads relative to canals due to technological progress in the railroad sector, although in principle the bias could go either way (canals also experienced productivity gains).

that of previous studies, including Fogel: (1) the method used to determine the counterfactual transport network, (2) the method to calculate the loss, and (3) the year of analysis. As I have emphasized before, the counterfactual transport network used in this paper differs from that of Fogel and previous authors, who also took as given the counterfactual network that Fogel developed. Fogel's counterfactual network is the observed canals in 1890 plus some additional canals that Fogel thought would be built in the absence of railroads. With railroads gone, canal projects that were proposed but not built due to pessimistic estimates of their profitability would have been built to replace the missing railroads. I allow a large collection of profit-maximizing transport firms to determine the alternate canal network when railroads are gone, thus the counterfactual transport network is determined within the scope of the model. Although direct comparisons are difficult due to the reasons previously mentioned, this leads to more canal building that what Fogel had but insufficient substitution through canal construction to completely mitigate the output loss due to the removal of railroads.[7] Canals alone are unable to fully replace the missing railroads.[8]

To illustrate how my output calculation is different from that of previous studies, first recall Fogel's method. For Fogel, the output loss is due to farms no longer being able to bring their goods to market since they now operate too far away from a transport link. Thus for Fogel what matters is the distance to the nearest canal, not the cost of shipping goods from the farm to the destination city as computed over the entire canal network. Fogel computes the output loss as the annualized total reduction in land value for all farms forced to shut down due to being too far from a canal in the counterfactual transport network without railroads. These losses do not affect the resulting alternate transport network in any way.[9]

The output loss method that I employ requires that a reduced form version of the gravity equation of trade holds across all productive locations within the U.S. in both the observed and counterfactual worlds.[10] All trade is between locations within the borders of the U.S. according to the gravity equation. Locations are spatially fixed, but the level of output at each location is allowed to vary and depends negatively on trade costs. Thus when rail links are removed in the counterfactual, least-cost routes over the transport network increase in total cost and trade costs within the U.S. increase, leading to output declines according to the gravity equation.[11] The method used here is continuous, as output at a location continuously falls

[7] Dropping a typical endogenous canal network from this paper into Fogel's model of determining the output loss results in an output loss of less than 10%, comparable to what Fogel originally found. Again, due to other differences, this comparison is approximate.

[8] The set of endogenous counterfactual canal networks developed in this paper still have a multitude of coverage gaps that result in isolated locations and pockets of locations with low output, relative to what was observed in 1861 with both railroads and canals.

[9] In other words, there is no feedback from the output loss to the counterfactual transport network and the causal relationship strictly goes the other way, from the network to the losses.

[10] The parameters in the gravity equation are estimated to match as accurately as possible the observed cross-sectional output distribution in 1861, and they do not change in the counterfactual. Alternatively one can think of this as a direct causal interpretation of the standard reduced form gravity equation.

[11] Output falls the most for locations that are isolated from the counterfactual canal network and are forced to use the relatively costly outside option of roads when railroads are no longer available.

due to increases in trade costs, while Fogel's is discrete: when railroads are gone, as a farmer you are either sufficiently close to a canal to continue to operate or not. Also this method allows for two-way feedback between the output loss and the counterfactual transport network: as the output loss increases, firms find canal construction less profitable since trade volume over their links and thus their revenue is reduced, so they continue to build fewer links which can lead to an even higher output loss. The endogenous canal network that results depends critically on the output loss method assumed. This is why a clean decomposition between the two is not possible: they are interdependent.[12]

Finally, the difference in year used for the analysis is crucial: previous studies looked at 1890, a year that has been focused on due to Fogel's original work and data limitations that he faced, while I use 1861 for the analysis, which is also motivated by data limitations. I need data on firm ownership of links in the railroad industry, which is to my knowledge only available cleanly for 1861 and not 1890.[13] Without repeating my analysis in 1890 instead of 1861, which cannot be done without additional data work, I am currently unable to quantify explicitly the importance of the base year to the output loss difference.

A brief outline of the paper is as follows. Section 2 provides motivation by presenting a stylized example to build intuition for why the output loss can be so large when railroads are removed due to the presence of complementarities between railroads and canals. Section 3 reviews the relevant literature in economic history, networks, and economic geography. Section 4 presents the novel data developed in this paper, including data on observed transport networks, which allows the network formation model to be solved at the firm level. Section 5 develops the model, a strategic model of transportation network formation which has implications for macroeconomic aggregates. Section 6 discusses the simulation method, which allows for rich spatial heterogeneity across railroads and canals, used to sample from the set of stable counterfactual transport networks. Section 7 discusses the quantitative results, including the main result that output falls by 45% in the counterfactual without railroads; in the counterfactual without canals, output falls by 49%. Section 8 concludes. Figures and tables are provided in the Appendix, parts A and B respectively. Additional motivation for the empirical approach taken in this paper is provided in the Appendix, part C.

2 Motivation

The empirical exercise of this paper is to estimate the marginal growth contribution of 19th century U.S. railroads relative to a baseline with canals only. Roads, turnpikes or wagon trails, are present in both

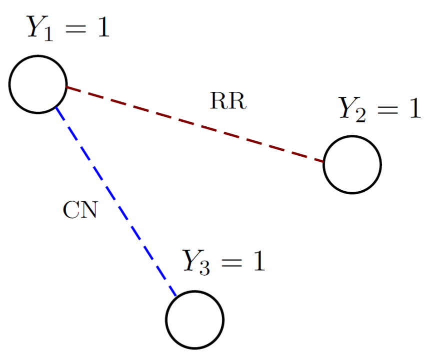
[12]However, if one drops Fogel's counterfactual canal network into the output loss model of this paper, the output loss is roughly 80%. This is because Fogel's network is relatively sparse compared to the typical endogenous network in the current framework. This is not a truly appropriate comparison since Fogel's network must be roughly approximated to fit with the past.

[13]The entire railroad network is certainly known for 1890, but not how it breaks down into pieces owned by each firm. Also it is easier to deal with fewer links in 1861 relative to 1890.

scenarios as a costly outside option for the transport of goods, along with natural waterways and coastal shipping. Relative to the previous literature, my main contribution is to endogenize the counterfactual canal network through a network formation game played by transport firms, specifically canal firms in the counterfactual scenario where the railroad was never invented.

To motivate this exercise, I provide a stylized example of how the output loss from the removal of railroads can be large due to strong railroad-canal complementarities. The figure below provides the initial condition of the economy: the empty network with no links, three cities, and two possible links represented as dashed lines; actual links will be denoted by a solid line. The color convention is red for a railroad link (RR) and blue for a canal link (CN). The only cost of building a link in this example is its construction cost. For simplicity assume no heterogeneity in link construction costs such that $c_{RR} = c_{CN} = 1$, so both links cost exactly one unit of output to build. The trade flow protocol is $X_{ij} = Y_i Y_j$ if connected for bilateral trade between cities i and j. If the cities are not connected, as in there does not exist a path between them on the transport network using railroad and canal links, then they do not trade. Cities 2 and 3 do not trade for simplicity. Thus in the initial condition no trade takes place and $X_{ij} = 0$ for all i, j pairs; this is autarky.

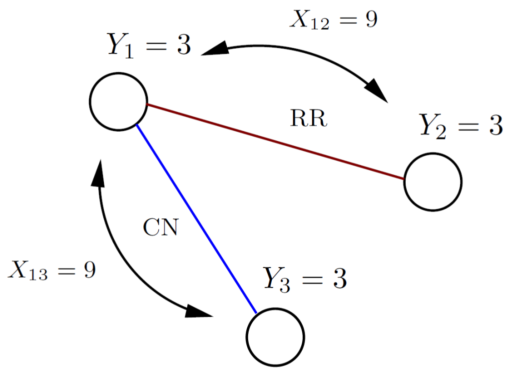
Figure: Empty Network



This example captures the idea that output increases when trade costs fall. Let city output Y_i equal the number of cities connected to city i via canal and railroad links, including i itself. In the empty network, the baseline output level for each city is $Y_i = 1$ in the worst-case scenario where no transport links are present. Due to spatial heterogeneity in geography, the set of potential transport links is such that the only possible canal link is between cities 1 and 3 and the only possible railroad link connects 1 and 2. One possible geography story to support this is a mountain between cities 2 and 3, thus no possible link there, a river between 1 and 3, supporting the construction of a canal, and flat plains between 1 and 2, allowing easy railroad access. Shutting down variation in costs across types, geography dictates that the set of potential links varies across types. Of course, if the set of potential links was the same for railroads and canals, they would be perfect substitutes here and the output loss due to removing railroads would be zero.[14]

The next figure shows the observed network from the factual world where railroads and canals coexisted.

Figure: Observed Network



Due to the reduction in trade costs with both links present, output at each city has increased to $Y_i = 3$. Assume only two firms, one railroad type and one canal, that own each link of the same type: one railroad firm owns the railroad link (RR) and one canal firm owns the canal link (CN). Although this does not matter for the counterfactuals, assume further that firms cannot own links of the other type, so a railroad firm owning canal links or vice versa. Thus a firm is completely characterized by the set of links it owns and its type. Assume that firms charge uniform price $p = 0.2$ which can be interpreted as a usage fee per unit of traffic on the link. So one finds that both firms make revenue $p \sum_{i,j \neq i} X_{ij} = 0.2(9) = 1.8$ individually since everything is symmetric across firms except for the specific identity of the link they own.

To go through the revenue calculation in detail, there are two trade flows in total that need to be accounted for: cities 1 and 3 trade to generate $X_{13} = Y_1 Y_3 = 9$ and 1 and 2 trade $X_{12} = Y_1 Y_2 = 9$. Assume that all trade flows take the shortest possible path over the transport network. Since X_{13} and X_{12} are direct flows, the shortest path is obvious: the direct path between cities. For the canal firm, only X_{13} uses the canal link from city 1 to 3. The railroad firm draws traffic X_{12} on its railroad link from city 1 to 2. Thus both firms have the same traffic flow over links and link price, and the revenue calculation follows directly. Abstracting away from the pricing decision, firms have an incentive to build links that are as central as possible to attract as much traffic as possible to maximize revenue and profits.

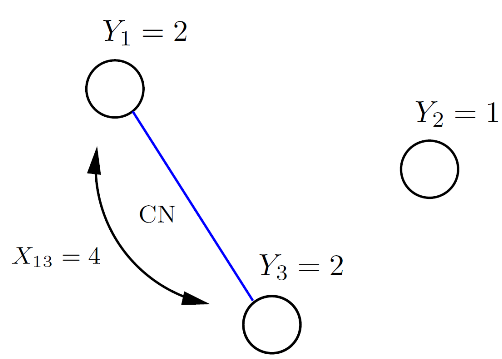
To check stability of this network, note that both firms earn revenue 1.8 which covers link construction costs of 1. Normalizing the value of firm exit to zero, where exit is removing your last link, transport firms do not want to remove the links they already have. Since no unbuilt potential links are available, no more links can be constructed by either firm. Thus the network is stable against unilateral deviations by either

[14] This relies on the link construction cost being identical across types.

firm. This is the stable network result one would expect in a decentralized equilibrium with both transport technologies available. Since trade costs are as low as possible with all potential links built, output is as high as possible.[15]

One possible interpretation of the exogenous counterfactual canal network presented by Fogel or the previous literature, where railroads are removed but the resulting canal network is fixed in some prior configuration, is given in the next figure. Here the canal network used is the observed network. Assume that the remaining canal link is priced exactly the same as in the real-world case with $p = 0.2$ fixed. Output falls across cities since trade costs increase when railroad technology is withdrawn from the economy. Now the canal firm's revenue can be computed as $p \sum_{i,j \neq i} X_{ij} = 0.2(4) = 0.8$, which is insufficient to cover its costs of 1. Since the canal firm cannot operate at a profit it would choose to shut down if it was allowed to in the protocol that determines the network, but since this is the exogenous network counterfactual the canal link remains in place by assumption. The key disagreement in the literature is the profitability of remaining canal firms when railroads are removed. According to the trade protocol described here, the remaining canal actually becomes less profitable and would prefer to exit if permitted to do so in the network solution concept. However, by assumption the firm continues to operate, its link remains in place, and the output loss of $9 - 5 = 4$ units of total output or 44% of observed output is mitigated as a result.

Figure: Exogenous Canal Network Counterfactual

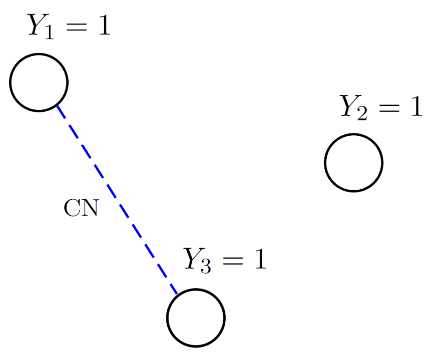


Although the output loss was nontrivial when the canal firm was forced to continue operating in the exogenous network counterfactual, the loss is amplified when the canal firm is allowed to exit in the endogenous network counterfactual presented below. This network is stable in the sense that no individual firm has a unilateral incentive to revise its links through either link addition or link removal. It turns out that this is the stable counterfactual network when the canal firm is allowed to be strategic and exit the industry due

[15]This does not take into account link construction costs by deducting them from total output across cities, since transport firms maximize their profits not total output.

to its operating loss. As the canal shuts down, the economy reverts to the empty network with total output 3. The resulting large output loss of $9 - 3 = 6$ units of total output, or 67% of observed output, is due to the additional increase in trade costs and reduction in total output when the canal ceases operation.

Figure: Endogenous Canal Network Counterfactual



Thus it appears that railroads and canals are not substitutes but rather complements. This complementarity is not assumed directly in any functional form, but emerges naturally from spatial heterogeneity in costs. Specifically in this example I present the stark case where the sets of potential railroad and canal links differ in a substantial way due to geography: a river is in one place (between cities 1 and 3) but not another, so canals cannot be built everywhere. Again, with no spatial variation in link construction costs and identical prices across modes of transport, railroads and canals would be perfect substitutes. However, geography alone is enough to differentiate railroads and canals such that they go from substitutes to complements, resulting in a large output loss when railroads are removed; the same is true if one was to remove canals.[16]

Geography manages to keep the two transport modes at arm's length such that trade creation, the increase in total output when I build a link, exceeds trade diversion, the share of competitors' traffic I divert when I build a link near the links of other firms. Note that ex ante one cannot tell which effect is going to dominate, trade creation or diversion, from a purely theoretical perspective. This is an empirical question which requires a model to answer. See the Appendix, part C for additional motivation, including a discussion of historical data on macroeconomic aggregates, evidence from preliminary reduced-form vector autoregression models, a discussion of why the canonical growth model is difficult to take to the data in this case, and a stylized example of how a search algorithm can fail to find the globally optimal solution to the social planner's problem of selecting the transport network to maximize total consumption.

[16] Again this result is driven by cost heterogeneity across types or transport modes (railroads, canals).

3 Literature Review

This section provides a brief summary of the relevant existing literature on economic growth and transport network expansion, specifically the case of U.S. railroads. First I discuss the work of Robert Fogel, which is central to the empirical exercise of this paper on the growth contribution of railroads. In his book *Railroads and American Economic Growth*, Fogel (1964) estimates the marginal effect of railroads on U.S. output, where the next-best alternative is the counterfactual canal network generated by Fogel in Figure 44. One can think of this figure as Fogel's solution to the social planner's problem of where to build extensions of the existing canal network in the absence of railroad technology, taking the observed output distribution over cities as given. He emphasizes the role of federal government land grants, state and local government subsidies, the transition from iron to steel rails for heavy freight, and the standardization of track gauge in the development of the 19th century U.S. railroad network. The "axiom of indispensability" is defined by Fogel as the claim, made by historians such as Jenks (1944), that the railroad was indispensable or essential for American economic growth in the period. Fogel concludes that the U.S. economy would have proceeded on a similar growth path with or without railroads. The story goes as follows: since the United States was endowed with an extensive system of rivers and other navigable waterways, relatively few counterfactual canals are necessary in order to match the coverage of the combined canal-railroad network using canals alone. Fogel performs his primary counterfactual experiment in 1890: a computation of the "social savings" or marginal growth contribution of railroads.

The "primary effect" of railroads is a reduction in transportation costs, while the "secondary effect" acts through forward and backward linkages, specialization, and the division of labor. Fogel finds that the 1890 road-canal price differential was large, while the canal-railroad price differential was small.[17] He assumes that waterways are a constant-cost industry, so prices do not change as waterway traffic is scaled up in the counterfactual without railroads. Based on this assumption, Fogel's social savings estimate across all transportable products is less than 5% of 1890 U.S. output, at most 4.7% under one set of assumptions, a result that contradicts previous indispensability claims for railroads. The inter-regional savings from East-Midwest and East-South "trunk lines", although highly praised by historians, is estimated to be less than the intra-regional savings due to high transport costs on roads.

As a closing assessment of the "take-off hypothesis" due to Rostow (1956), which was influential at the time, Fogel states that there exists no identifiable unique take-off period in terms of a sudden burst of economic growth since the increase in manufacturing's share of total output was smooth and continuous from 1807 onwards. No single innovation was vital for 19th century U.S. economic growth; the Industrial

[17] See Figure 4 (Appendix, part A).

Revolution, as a consequence of the previous Scientific Revolution, was associated with multiple solutions to cost-reduction problems in a series of incremental innovations. All sectors of the economy grew concurrently due to a series of sector-specific technological innovations, which rejects the "leading sectors hypothesis" that certain leading sectors expanded first. In the United States for the period 1830 to 1890, Fogel concludes that economic growth can be attributed to technological change, lower transportation costs, and increases in market size due to population growth and urbanization. However, no single innovation can claim a large share of this growth.

An overview of Fogel's work and the subsequent literature is provided in Atack and Passell (1994). Fogel's effort to measure the social savings of railroads has been subject to extensive criticism (Lebergott, 1966; Nerlove, 1966; McClelland, 1968, 1972; David, 1969; Williamson, 1974). Passenger service is ignored and the assumption of canals as a constant-cost industry is strong. By 1890, canals had already been out-competed by railroads and driven to the edge of economic viability; waterway prices were pushed down as far as possible to maintain limited competitiveness with railroads. Canals began shutting down in response to entry by railroads as early as 1840. Fogel does not make an extensive attempt to check that the waterways in question have sufficient capacity to support the traffic flows implied by the counterfactual canal network. Additionally, Fogel assumes that producers can compensate for the seasonality of canals by holding excess inventories, and the only cost associated with this activity is the storage cost of additional raw materials at factories; for time-sensitive goods or for goods with a multi-tiered supply chain, this assumption may not hold.[18] He discounts the speed and reliability differential between canals and railroads; in 1890, canals were specialized and not in direct price competition with railroads for most manufactured goods. By 1890, 60 years of growth aided by railroads had occurred, and in that time railroads had eliminated canals as a viable competing method of transit in terms of canal firm profitability.

Finally, the counterfactual system of canals may cover the same geographic area as the real-world canal-railroad network, but this does not guarantee that average travel distances or transport costs are comparable. Chicago and New York are connected in both real-world and counterfactual networks, but the route length is much longer in Fogel's network without railroads. Therefore, the counterfactual canal system understates the actual cost of routing goods through the network. This line of criticism, addressed in Donaldson and Hornbeck (2013), is associated with an additional output loss of a few percentage points at most when accounted for. Thus updating Fogel's binary transport coverage convention, that as a farmer you are either close enough to a transport link or not, to continuous coverage in the form of least-cost routes over the

[18]Waterways freeze in winter, rendering them inoperable; railroads, if properly maintained, can operate year-round. This contributes to a service quality differential between railroads and canals that is reflected in an observed price differential, with railroads priced higher than canals.

transport network does not make railroads essential to growth.[19]

Fishlow (1965) provides an alternative estimate of the social savings due to railroads of at least 15% of U.S. output in 1890. This estimate is larger than Fogel's due to less generous assumptions on the available set of transport alternatives, specifically that the canal network would not have expanded and roads would not have improved in the absence of railroads. In Goodrich (1960), the author documents the 19th century history of U.S. government intervention in the transport sector at both the federal and state level. He finds that policy exhibited large variation over time in terms of the level of government subsidization and control. Insofar as government subsidized canal or railroad construction, since this paper generates an estimate for link costs within the model based on observed transport networks, I am capturing the average effect of such subsidies on observed link costs.

Fogel responds to his critics in Fogel (1979), which addresses their main points of attack on his work. In response to Fishlow's use of the observed canal network instead of Fogel's extended network, Fogel says that the actual canal network is the least likely outcome without railroads and thus his expanded version is more plausible since the proposed extensions would have been profitable. Regarding criticism of the assumption that the canal industry exhibits constant costs as traffic increases, Fogel states that the industry has constant costs at worst and decreasing costs are more likely, thus assuming constant costs provides an upper bound on the losses when railroads are removed and traffic is rerouted to use canals. Fogel assumed a low elasticity of substitution between railroads and canals, which generates a modest social savings due to railroads; although critics claimed that this elasticity could have been higher, reducing the social savings, Fogel provides empirical evidence that his assumed value near one is appropriate.[20] To address the line of thinking that observed railroad prices were too high due to local monopolies and pricing power, Fogel's retort is that the railroad industry did not earn an excessively high rate of return on capital and taking modest markups into account does not substantially affect the resulting social savings number. Finally, to mollify critics that are concerned with the long-run indirect growth effects of railroads, Fogel replies by saying that accounting for economies of scale or structural change in the economy due to railroads would probably have a small effect on the derived social savings.

The networks literature in economics has expanded considerably in the past few decades (Jackson, 2008, 2010; Shy, 2011). Generally, agents maximize some individual objective function in the presence of a network structure that is either exogenously imposed or an endogenous outcome of some network formation game. The actions of other agents in the network are classified as either strategic substitutes or strategic

[19] Unless combined with other factors, such as an endogenous output distribution or counterfactual canal network in the absence of railroads.

[20] In this paper, to be conservative in terms of the output loss due to removing railroads and to abstract away from quality differences between transport modes, I assume that canals and railroads are perfect substitutes from the perspective of the demand side (i.e. user of transport services). Thus this implies an infinite elasticity of substitution between types.

complements in the payoff function. Link formation is costly, but the gains derived from membership in the network are increasing in the number of connected nodes or the intensity of the links (Aumann and Myerson, 1988; Jackson and Wolinsky, 1996; Galeotti et al., 2010; Elliott and Golub, 2013). However, typically in the literature players of the network formation game are nodes in the network and not firms, which I conceptualize as disjoint sets of connected links with a permanent type, railroad or canal. The network formation game used in this paper is a modified version of the simultaneous link announcement game of Myerson (1977) where the players are firms. In terms of econometric theory and estimation, the most relevant paper is Sheng (2012), but again the author uses pairwise stability as the solution concept which depends on interpreting the players as nodes, thus their framework does not apply directly to the network formation game with firms studied in this paper.

One of the closest antecedents of this paper in the literature is the framework developed by Kelly in a series of papers on transportation networks and economic growth (Kelly, 1997, 2001). In Kelly (1997), a branching tree network structure is exogenously imposed: two potential links branch out from every node to two nodes further down the tree hierarchy. The links are waterways or canals that are initially closed to trade; the network is initially empty. The nodes are interpreted as cities with a continuum of intermediate goods producers and a unit mass of identical households that consume the final good. A specialized low-cost producer is located at each node for one of the intermediate goods, so the nodes are not ex-ante identical. Firms at a particular node have the option of opening up two nearby waterways at a cost, and once open the canals impose no additional transport costs. In the complete network, firms specialize and produce only the low-cost intermediate good, trading for other intermediate goods to produce output and provide households with consumption. Kelly obtains a logistic transition path for output, which is consistent with historical episodes of rapid economic growth driven by transport network expansion, such as growth of the Chinese waterway network during the Song Dynasty.

The empirical framework used here is also similar to that employed by Donaldson and Hornbeck (2013). They use a trade model to estimate the output loss generated by removing railroads from the observed 1890 transportation network. In the no railroads counterfactual, they add to the factual canal network the set of counterfactual canal links proposed by Fogel (1964). They find that the no railroads counterfactual induces an output loss of 3.4% relative to the observed network without Fogel's proposed links. When Fogel's proposal is included in the canal network, the output loss is mitigated by only 13%. Solving the social planner's problem to determine the optimal transport network from the perspective of aggregate welfare, Perez-Cervantes (2012) finds that output decreases by 9.6% with only canals available. Other related papers in the literature include Donaldson (2012) on railroads as a source of exogenous variation in trade costs in colonial India, Banerjee et al. (2012) on railroads in China, Feyrer (2011) on the closing of the Suez canal as exogenous variation in world trade costs, Dell (2012) on drug trafficking in Mexico over the road network via shortest paths, and Holmes (2011) on the distribution network of Wal-Mart.[21] These papers assume, as in this paper, that trade occurs over least-cost routes on the transport network: for Banerjee et al., the Chinese railroad network, for Donaldson, the railroad network of colonial India, for Feyrer, the world network of oceanic shipping routes, and for Dell, the road network of Mexico. However, although in this paper trade also occurs over least-cost routes, I differ from the previous literature by endogenizing the counterfactual transport network.[22]

Atack and coauthors have written a series of papers on the U.S. railroad network and economic growth (Atack et al., 2010; Atack and Margo, 2011; Atack, 2013). They find that, in agreement with Fishlow (1965), railroads were not built ahead of demand in the Midwest but rather their construction followed pre-existing patterns of economic development by building railroad lines that would generate the most traffic first. However, those railroad lines increased market access to large swathes of new agricultural land, especially in the Midwest, which were brought into production as a result; it is unclear if this land would have been cultivated as intensively without railroads. Finally, Atack documents the reduction in price dispersion across cities due to railroad and canal expansion in a novel time-series dataset on U.S. transportation networks. However, since I need information on firm ownership and not just the spatial disposition of transport network links, this novel data is not of immediate use for this paper.

See Gramlich (1994) for a summary of the literature on the link between transportation investment, as a subset of infrastructure investment, and economic growth. The new economic geography literature builds models where space and location matter due to transportation costs and local productivity spillovers (Krugman, 1991, 1998). For the purposes of this paper, I am primarily interested in spatial spillovers induced by local reductions in transport costs via transport network expansion and the role that location plays in determining economic outcomes. The main role that geography plays in this paper is to generate spatial heterogeneity in link construction costs for railroads and canals, thus the two transport modes are not perfect substitutes.[23] The resulting strategic complementarities that I find between railroads and canals are related to the coordination failure model of Cooper and John (1988).

The vast majority of the literature on economic growth does not use a network micro-structure to model transportation network expansion. Instead, these papers explain self-reinforcing economic growth in a variety of settings with demand externalities or production spillovers (Rostow, 1956; Murphy et al.,

[21] For more on transportation and economic outcomes, also see Fernald (1999); Baum-Snow (2007); Redding and Sturm (2008); Duranton and Turner (2012); Duranton et al. (2013); Faber (2013).

[22] Thus I do not treat the transport network as exogenous variation in transport or trade costs.

[23] Again note that in this paper the two transport modes are perfect substitutes from the demand side (by assumption, to abstract away from service quality differences that show up in observed prices), but not the supply side (due to heterogeneity in costs). Spatial spillovers due to a reduction in transport costs are captured by taking the gravity equation of trade literally as a protocol to determine output across cities in the model.

1989; Durlauf, 1993; Galor and Weil, 2000; Ngai, 2004). The literature on endogenous growth is closely related due to the implausibility of matching a short burst of rapid economic growth using a sequence of independent exogenous productivity shocks, even if persistence is allowed. With endogenous productivity growth, it is plausible that short-term rapid growth is a domestically-generated and self-perpetuating event that eventually self-corrects (Romer, 1990; Aghion and Howitt, 1992; Kremer, 1993; Segerstrom, 1998). This process could be observationally equivalent to a transport network expansion episode if the network is not directly observed.

Economic growth can occur through a number of channels, including extensive expansion of transportation systems to utilize more land in production and increase trade. In the literature, the most commonly explored channels are market scale and human capital (Lucas, 1988; Stokey, 1988; Becker and Murphy, 1992; Lucas, 1993). The general result of specialization in production, going back to Adam Smith's *Wealth of Nations*, is that the division of labor is limited by the size of the market. Economic growth occurs through market scale when increased market size encourages specialization, increases output per worker, lowers production costs, and increases output (Stigler, 1951). International trade, transport networks, product standardization, and economic integration through free trade agreements are all potential examples of growth through market scale. Typically, the assumption of increasing returns to scale in production captures scale effects through specialization and the division of labor (Matsuyama, 1991; Yang and Borland, 1991; Ades and Glaeser, 1999). These scale effects are strong enough to overwhelm diminishing marginal returns to individual factors.

North (1966) discusses the role of transportation in early 19th century U.S. territorial expansion, settlement, trade, and growth. Market scale effects, specialization, and the division of labor were important drivers of early American economic growth. Economic activity in the late 18th century transitioned from household production to local and export production as the U.S. had an expanding home market due to land purchases and settlement and an attractive external market due to foreign demand driven by European conflict and the Napoleonic Wars. As a barrier to expansion, significant lags were present in the incorporation of new frontiers into the corpus of U.S. territorial holdings; lags were also severe in the construction of new transportation links.[24]

With a total population of 3.9 million, the United States in 1790 was a rural decentralized economy.[25] Household production was the dominant form of economic activity: self-sufficient farmers produced for their own subsistence with occasional local trade that generated a small supplemental market income. Due to the high cost of land transportation, it was infeasible to move bulky goods or food on long overland routes;

[24] At this stage, roads (turnpikes) and canals.

[25] It possessed only two cities above 25,000 in population: Philadelphia and New York.

the U.S. needed navigable waterways for the export of food or raw materials to England or continental Europe. At this time, the U.S. economy was characterized by small-scale domestic markets and limited international demand; the American advantage in primary products was insufficient to overcome European barriers to trade and trans-Atlantic shipping costs. Despite a high domestic annual birth rate of 5.5% in 1800, both capital and labor were scarce relative to land with few domestic financial intermediaries to support investment. Additionally, the potential for large-scale immigration into the United States was limited.[26]

Land was the most important factor for early growth, specifically the Louisiana Purchase in 1803. Local internal trade developed, connecting major seaports with the hinterland and internal periphery. The largest U.S. cities all had direct access to the Atlantic: Philadelphia through the Delaware River and Bay, Baltimore through the Chesapeake Bay, and New York via the Hudson River. Turnpikes extended the size of local markets, but were too costly to have a major impact on internal trade flows; only high-value goods were transportable by land at a profit. Coastal trade expanded at a rapid rate and the interregional flow of of labor to the West intensified in the early 1800s. Due to external European demand, shipping and the re-export trade dominated American economic activity in cities on the Atlantic coast, leading to the urbanization of ports and a widening of local markets around ports as urban centers in a core-periphery system. The subsequent effort to reduce local transportation costs involved the construction of viable roads in the Northeast; however, the U.S. continued to import most manufactured goods from Europe until industrialization in the 1820s, which was also centered in the Northeast.

Taylor and Neu (1956) present empirical evidence on the growth of the U.S. railroad network from 1861 to 1890. For the purposes of this paper, the most notable part of the book is its endnotes which contain detailed maps of the 1861 railroad network. According to Taylor and Neu, "The first railroads in the United States were built, as were most of the early turnpikes [roads] and canals, to serve nearby and local needs" (p. 4). As an example, the authors cite the Boston and Worcester railroad line, which was originally constructed to divert trade from the nearby Blackstone Canal. They also provide evidence that railroad construction in the United States was done in a decentralized manner by many small firms; for additional evidence see Poor (1860). Early railroads were locally financed to provide local transportation for the primary products of farmers. However, after the Civil War, railroads were viewed as "a coordinated network whose primary function was to facilitate transportation" (pp. 5-6).[27] By digitizing the firm-level data in Taylor and Neu, I provide novel evidence on the firm distribution for railroad and canal firms which is used to bound costs and run counterfactual experiments.

[26]Immigration flows were small at 4,000-6,000 per year.

[27]Previously, "early American railroad promoters looked upon the railroad as primarily a means for short-haul transportation, and therefore saw little need for uniformity of railroad gauge" (p. 13); gauge was not completely standardized until 1900, but by 1880 roughly 81% of all U.S. railroad mileage was in standard gauge.

4 Data

I now discuss four distinct transportation networks which are observed in the data: railroads, canals, natural waterways including rivers and lakes, and coastal shipping consisting of the Great Lakes and Atlantic Ocean. They are treated separately in the model: railroad and canal networks are endogenous while waterways and coastal shipping are exogenous, considered as part of the natural resource endowment of the United States. First I compute centrality measures for the railroad network to characterize its structure and connectivity.

See Figures 5 and 7-10 (Appendix, part A) for a visual representation of the 1861 U.S. railroad network from Taylor and Neu (1956).[28] Nodes are cities and links are railroad lines. Nodes are placed in the territory of the U.S. through latitude and longitude data for cities. Hub cities Boston, New York, Philadelphia, and Chicago, highlighted in Figure 8 when one weights nodes by their degree, are connected to nearby smaller peripheral cities.[29] However, by inspection, the U.S. railroad network as a whole does not have hub nodes. The hub cities listed previously are local or regional hubs, not national hubs for the entire network. The railroad network has 595 nodes and 767 links.

Nodes are labeled with city and state identifiers, and most links are labeled by firm ownership. There are 251 railroad firms, and the individual pieces of the network or subnetworks they own have on average 2.31 links each. In Figure 5 each firm's subnetwork is highlighted in a different color to emphasize that the industry was comprised mainly of small firms at this time; mergers and consolidation occurred later. Trunk lines, nodes and therefore adjacent links with high betweenness centrality in Figure 7, are observed from the East to the Midwest through mid-Pennsylvania and from the East to the South through the Philadelphia-Washington-Baltimore corridor.[30] No primary Midwest-South rail link exists yet, since that corridor was dominated by Mississippi barge and steamboat traffic due to a relative cost advantage for canals over railroads that persisted for decades after the introduction of the railroad to the United States in 1830.[31] The network takes on a grid-like appearance in some areas, specifically the Midwest, especially Illinois and Indiana, and Massachusetts. These grid-like substructures appear near major regional hubs such as Chicago and Boston.

The 1861 U.S. canal network is presented in Figure 11. The data source is the American Canal Society, and the complete list of canals operating in 1861 is provided in Table 1 (Appendix, part B). Again each firm is highlighted in a different color. The network has 595 nodes by construction, with the set of nodes consistent

[28]Although included in the maps of Taylor and Neu (1956), nodes and links associated with the Canadian railroad network are excluded to focus on the U.S. counterfactual question.

[29]In this context, the term "hub" refers to nodes with high relative degree. These cites are the top four in terms of degree and are weighted in Figure 8 accordingly. Node degree is defined as the number of links that count the node as either origin or destination of the link in an undirected network. By assuming that transport networks are undirected, I impose the restriction that all links are reciprocated: if a link between A and B exists, then a link between B and A also exists and vice versa.

[30]Betweenness centrality is defined as the fraction of shortest paths that pass through a particular node in the network. Nodes on more shortest paths are considered more central by this measure, which is a proxy for traffic volume.

[31]Track gauge is identified in Taylor and Neu (1956) if this information is needed for later work. Modeling track gauge choice and "gauge wars" is beyond the scope of this paper.

throughout from the railroad network data, and 88 links. Major canals include the Erie, Ohio & Erie, Wabash & Erie, and Illinois & Michigan. In total, 42 canals are included in the data. Canal construction is concentrated almost exclusively in the North, with the South relying on its system of navigable waterways. The Pennsylvania subnetwork appears to be the most grid-like part of the canal system.

The network of navigable U.S. waterways, which includes major lakes and rivers, excluding the Great Lakes which are counted in the coastal network discussed next, is detailed in Figure 13. The Mississippi River corridor plays a prominent role, with a number of tributaries that extend through Iowa and Pennsylvania, providing extended connectivity from the Gulf of Mexico to the inland United States. Although the Pennsylvania-Mississippi corridor could be used to transit goods to the South and Europe, it is a circuitous route that would not be optimal when shipping to the East. Relative to the North, the South possesses a superior set of natural waterways and inland rivers. Although similar in number, the waterways of the South provide more direct access to coastal areas and the Atlantic Ocean, whereas the waterways of the North and Midwest require significant investment in canals in order to obtain connectivity from the Atlantic seaboard through to the Great Lakes and inland areas. Table 2 provides a list of major U.S. waterways included in the data, which are mostly rivers.

Figure 14 shows the network of U.S. coastal areas. This is a synthetic network that captures the use of coastal shipping in transport, made possible by the Great Lakes and Atlantic seaboard. In general, each coastal city is connected to its nearest 2-3 neighbors both above and below the coastline. The primary features of the network are Lake Michigan, Lake Erie, and the Atlantic Ocean. The Welland Canal, which opened in 1829, is assumed to provide the necessary waterway link between Lake Erie and Lake Ontario to connect the Great Lakes to the Atlantic. The waterway and coastal networks, including the Welland Canal, are part of the natural resource endowment of the United States, and as such they are present at no construction cost in both observed and counterfactual scenarios.

I merge these four networks into a single combined multimodal transport network. This network is used in order to compute bilateral transport costs and thus output and welfare, which requires combining knowledge of the network structure with information on observed freight rates or link usage prices. For the real-world case, the transport network structure is taken directly from the railroad, canal, waterway, and coastal networks. In order to assign a price to usage across these four modes of transport, I abstract away from price differentials. These prices, which will be used throughout the paper to compute both factual and counterfactual trade costs, are listed in Table 3 in the rightmost "Assumed cost (normalized)" column. In order to prevent isolation of any cities due to withdrawn railroad or canal links, I assume that roads, also termed wagon trails or turnpikes, are operating in the background over all possible links as a costly alternative to the four modes discussed previously. As can be seen in Table 3, one would only use a road as a transport option of last resort since its price, and thus its usage cost in trade, is an order of magnitude higher than the other options.

Normalize the per-mile usage cost of wagon trails to one, so the relevant prices are now the relative cost of using railroads, canals, waterways, and coastal shipping for transport. These relative prices are also reported in Table 3. Note that the price differences between the four modes are small relative to the price gap between the four modes and roads as the outside option. These prices are reported in per-mile terms, so total transport link usage costs scale up linearly in the distance between source and destination cities. The computational procedure to construct the transport cost matrix is as follows. First, fill in roads for every possible link since they are the high-cost outside option when other links are absent. The non-road modes are assumed to all have the same usage fee or link price. So, for each link in the non-road networks, reduce the cost of using the same link in the multimodal network by the cost gap between roads and the other modes or 90% of the road price. The resulting network has link weights that are used to compute shortest paths and thus transport costs for shipping goods between cities.

Given the matrix of transport costs, one can then compute the payoffs that accrue to transport firms in the model. Figure 6 plots the distribution of railroad firm profits under the observed multimodal transport network; Figure 12 is analogous for canals. Profits appear uniformly distributed over firms except for a few highly profitable firms in the Midwest. From a firm's perspective it is best to own links in the center of the United States that are on many shortest paths for trade and thus attract high traffic and revenues, holding construction costs equal across links. The model, developed in the next section, provides a structural method to generate link costs for unobserved counterfactual links, and from that generate stable equilibrium networks and compute the output loss associated with those networks when railroads are removed.

5 Model

Development of the model proceeds in three steps. First I outline the trade model, the in-means gravity trade model to be precise, which determines trade flows and subsequently revenues and payoffs for transportation firms in the estimated network formation game used to find stable networks. Next I define and discuss the terrain cost function, derived from a structural model of link construction costs that takes deep heterogeneity in geography into account to determine the cost of building links that are not observed in the real-world transport network. Third, in combination with the terrain cost, I bound type-specific firm cost parameters called type costs using observed railroad and canal networks. Finally I employ simulation methods to sample from the set of stable counterfactual transport networks to answer the empirical question on railroads and growth originally posed by Fogel and reformulated in this paper.

5.1 Trade Model

Consider a static model with a fixed set of cities L which can be interpreted as nodes in the network. Although the set of cities is fixed, the output distribution across cities is endogenous, thus the output level of a particular city can fall arbitrarily close to zero.[32] The economy is closed to the outside world, as in no international trade, but cities within the economy trade with each other according to a standard gravity equation. A parsimonious specification of the gravity trade model for trade between cities $i, j \in L$ is

$$\phi Y_i = b \sum_{j \in L \setminus \{i\}} Y_i^{\alpha_1} Y_j^{\alpha_2} c_{ij}^{\alpha_3}$$

where Y_i is the output of city $i \in L$, c_{ij} is the bilateral transport cost between cities i and j, ϕ is the share of exports or tradables in output, b is a constant of proportionality, α_1 is the sensitivity of exports to exporter's output, α_2 for importer's output, and α_3 for transport cost.[33] I assume that transport cost c_{ij} is taken directly as the trade cost between cities i and j according to the least-cost route over the multimodal transport network where the options are road, natural waterway, coastal shipping, canal, and railroad.[34] The purpose of the trade model is to determine trade flows, and then subsequently traffic on each firm's subnetwork and firm revenue. First solve for "home" output Y_i as a function of $\{Y_j\}_{j \in L \setminus \{i\}}$, the "foreign" output distribution, and $\{c_{ij}\}_{j \in L \setminus \{i\}}$, the set of bilateral transport costs originating at home city i with foreign destination j. From this point forward ϕ and b can be dropped as scale factors of the firm payoff function without loss of generality; alternatively ϕ is absorbed into b, which is calibrated to observed total output, and b is subsequently dropped.[35] Thus one obtains

[32] To justify a fixed set of city locations, think about a fixed set of underlying natural resources like coal deposits and fertile farmland that demands a particular settlement pattern for resource exploitation. Households can migrate away from locations that become undesirable due to high trade costs, and this occurs endogenously in the model, but a city cannot be abandoned completely thus output at a location never goes to zero. Since roads are available everywhere as a costly outside option, individual locations will not become completely cut off from the transport network and production will still take place there; locations cannot experience infinite bilateral trade costs due to roads.

[33] One can also think of labor as the only factor of production with perfect labor mobility across cities. The motivation for trade between cities is a taste for variety story where each city produces a differentiated good demanded by all the others and agents trade goods over the network to satisfy demand. Alternatively all trade is in intermediate goods and lowering trade costs provides productivity and thus output gains as agents have access to a more balanced portfolio of intermediates (or more from the least-cost supplier of each intermediate good). For examples of empirical support for the canonical gravity equation see Anderson and van Wincoop (2003), Silva and Tenreyro (2006), and Chaney (2008).

[34] The cost of a least-cost route is in terms of dollars, not time, by assumption (or convention). If one was to judge routes according to their travel time, the output loss due to removing railroads would only increase since railroads had a huge speed advantage over all other modes of transport in the 19th century; at 15 miles per hour for railroads, their nearest competitor was coastal shipping at 5 miles per hour (3 for canals and 1 for roads).

[35] This is because the trade model is used to determine firm revenue. Any scalar multiple of firm revenue is absorbed in the type cost parameters (for railroads and canals) which are bounded in the estimation procedure. Therefore as a scalar factor the particular value of ϕ has no effect on model outcomes because costs are endogenous (specifically their level relative to revenue) and not taken exogenously from the data in absolute terms. For example, scaling ϕ by some factor c results in the cost bounds also scaled by c thus on net there is no effect. Invariance in ϕ is also due to taking the gravity equation literally as a protocol to determine the output distribution and calibrating the model such that total output $\sum_{i \in L} Y_i$ in the model for the observed transport network equals the level of output observed in the data. What matters for determining stable networks is how firm payoffs vary across space due to geography, not their absolute level. Alternatively, ϕ is absorbed into b and cannot be separately recovered from the calibrated value of b.

$$Y_i = \left(\sum_{j \in L \setminus \{i\}} Y_j^{\alpha_2} c_{ij}^{\alpha_3} \right)^{\frac{1}{1-\alpha_1}}$$

which to be precise is a proportionality and not an equality due to dropping b, but this does not affect model outcomes and I will continue to use equality throughout.

This expression presents difficulty in solving for the output distribution across cities $\{Y_i\}_{i \in L}$ since it generates a nonlinear system of $|L|$ equations in $|L|$ unknowns that is relatively slow to solve numerically. As an approximation to maintain computational tractability, approximate Y_j by $\bar{Y}$ and c_{ij} by $\bar{c}(i)$, where $\bar{Y}$ is average individual city output and $\bar{c}(i)$ is average transport cost between home city i and all other foreign cities $j \in L \setminus \{i\}$. This can be interpreted as an in-means approximation that takes the expected value of both sides of the previous expression

$$\mathbb{E}(Y_i) = \mathbb{E} \left[\left(\sum_{j \in L \setminus \{i\}} Y_j^{\alpha_2} c_{ij}^{\alpha_3} \right)^{\frac{1}{1-\alpha_1}} \right]$$

and assumes that Y_j and c_{ij} are independent for all $j \in L \setminus \{i\}$, thus the expectation on the right-hand side can be replaced with[36]

$$Y_i = \left(\sum_{j \in L \setminus \{i\}} \bar{Y}^{\alpha_2} \bar{c}(i)^{\alpha_3} \right)^{\frac{1}{1-\alpha_1}}$$

which is an approximation.[37] This approximation is reasonable because Y_j depends on the entire transport cost matrix $[c_{ij}]_{i,j \in L}$, the object which endogenously determines the distribution of output across cities in the model, and thus only weakly on individual elements c_{ij} of that matrix. Any reasonable structural model which specifies output distribution $\{Y_i\}_{i \in L}$ endogenously and exactly given the matrix of bilateral transport costs $[c_{ij}]_{i,j \in L}$ would compute substantially slower than the approximation described here, making it computationally infeasible to characterize the set of stable transport networks via simulation of the firm payoff function.

Simplify by dropping the term that depends on $|L|$ as another scalar factor like ϕ

$$Y_i = (|L| - 1)^{\frac{1}{1-\alpha_1}} \left(\bar{Y}^{\frac{\alpha_2}{1-\alpha_1}} \bar{c}(i)^{\frac{\alpha_3}{1-\alpha_1}} \right)$$

[36] While replacing $\mathbb{E}(Y_i)$ with Y_i since there is no uncertainty.

[37] It turns out that with estimated parameter values for α_1, α_2, and α_3, the functions f and g in $f(\bar{Y})g(\bar{c}(i))$ on the right-hand side of the original model without the approximation are concave and convex, respectively. Thus via Jensen's inequality one should not expect equality exactly when moving the expectation inside the right-hand side terms. Assuming equality when replacing foreign output and transport cost with their averages from the perspective of city i is part of the approximation. One can think of the approximation as an alternative in-means model which is distinct from the original parsimonious statement of the gravity model of trade; again this is done to maintain computational tractability.

aggregate the previous expression across all $i \in L$

$$|L|\bar{Y} = \sum_{i \in L} Y_i = \bar{Y}^{\frac{\alpha_2}{1-\alpha_1}} \sum_{i \in L} \bar{c}(i)^{\frac{\alpha_3}{1-\alpha_1}}$$

and solve for average city output level $\bar{Y}$, again dropping another scale factor in $|L|$

$$\bar{Y} = \left(\sum_{i \in L} \bar{c}(i)^{\frac{\alpha_3}{1-\alpha_1}} \right)^{\frac{1-\alpha_1}{1-\alpha_1-\alpha_2}}$$

which yields average individual city output $\bar{Y}$.[38] With $\bar{Y}$ known, one obtains the output distribution $\{Y_i\}_{i \in L}$ from the following equation

$$Y_i = \bar{Y}^{\frac{\alpha_2}{1-\alpha_1}} \left(\bar{c}(i)^{\frac{\alpha_3}{1-\alpha_1}} \right)$$

and then can compute total output across all cities Y as either $\sum_{i \in L} Y_i$ or $|L|\bar{Y}$.

Parameters in the gravity trade model are estimated based on the observed population distribution across cities from the 1860 U.S. Census such that $\alpha_1 = 0.1$, $\alpha_2 = 0.09$, and $\alpha_3 = -1.86$. The population distribution is equivalent in rank order to the output distribution under the assumption of a single factor production function which is linear in labor employed at a city.[39] What is used in this estimation procedure is not the level but rather relative output across cities under the normalization $Y_1 = 1$. Again the value of ϕ or $|L|$ does not matter for model outcomes because they are scale parameters that do not affect the vector of relative city output used for estimation. I set $\phi = 0.2$ as a convention and $|L| = 595$ is taken from the data since, as scale parameters, they are absorbed into link costs determined entirely in the model. These costs are bounded according to both built and unbuilt links in observed railroad and canal networks. The parametric restriction $\alpha_1 + \alpha_2 < 1$ is required in order to generate the appropriate negative relationship between transport costs and output. Fortunately, in minimizing the distance between the observed and theoretical output distributions, the estimation procedure chooses $\alpha_1 + \alpha_2 = 0.1 + 0.09 = 0.19 < 1$ thus it is unnecessary to impose this restriction directly.[40]

Data on the observed output distribution is presented in Figure 28 (Appendix, part A). This is actually

[38] Again, alternatively $|L|$ is absorbed into b, which is calibrated to match total output in the data.

[39] This is imposed due to data limitations, specifically that the 1860 population distribution is observed but not the output distribution at the city level.

[40] One reason that the estimated model delivers large output losses is that the estimated value for $\alpha_3 = -1.86$ is large and results in a large implied trade elasticity. In order to match the dispersion in output across cities, the model wants to heavily penalize cities in terms of output for high transport costs. In the trade literature when discussing exports between countries, a typical value used is $\alpha_3 = -0.5$ which results in a smaller trade elasticity for international trade. However, when dealing with trade within the United States between cities at a low level of aggregation, one should not necessarily expect that the model will yield an identical value for α_3 as in the international trade case. It is intuitive that the trade elasticity increases as the unit of aggregation becomes finer, but by how much is an empirical question.

1860 U.S. Census data on the population distribution, but the estimation method only requires relative output and the population distribution is mapped cleanly into the output distribution via a linear production technology in labor for each city. Clustering occurs around large Eastern hub cities like Boston, New York, and Philadelphia. This behavior can be explained by a number of factors outside the model, including agglomeration economies, local economies of scale, the history of U.S. settlement (East to West), and path dependence. However, a model with local economies of scale results in multiplicity in terms of large cities which are not uniquely determined.[41] Presumably historical events outside the model selected the real-world equilibrium with agglomerations located at Boston, New York, and so on. In general it is difficult to recover the observed output distribution unless productivity differences are innate to a particular city, which is close to assuming the result since there is no reason to believe that the productivity distribution is an invariant object in the counterfactual. For this reason I abstract away from differences in productivity across cities, permanent or otherwise, and focus only on output dispersion due to migration driven by differences in transport costs.

Figure 29 shows the exact output distribution according to the trade model without resorting to the in-means approximation that uses average transport cost. Dispersion in output across cities is due entirely to migration resulting from differences in transport costs. One can think of this as a frictionless model in terms of the ability of settlers to migrate across cities. The set of cities is fixed, which presumably depends on the exogenous distribution of farmland and resources across space. Insofar as cities were chosen to be close to natural resources, a fixed set of cities is reasonable since there is no reason to believe that cities would shift in the counterfactuals; the underlying spatial distribution of resources is invariant. Since the level of output is determined endogenously at each city, peripheral cities will have relatively low output and one can think of this as the city being removed from the economy. Compared to the approximate theoretical distribution, the exact distribution exhibits sharper spatial contrast with more spiked, local clusters of high economic activity in the center of the United States.

Finally, the approximate theoretical output distribution is presented in Figure 30. One observes a more uniform distribution of output across cities, as output is spread evenly across the middle of the United States. Perhaps this is a better fit of the data than local spikes in the Midwest at river junctions from the exact distribution, but still it is undeniable that the exact and approximate solutions differ substantially. Alternatively one could conceptualize the approximate model as distinct from the exact, whereas the former is computationally feasible and the latter is not. The approximate model retains the key mechanism that lowering transport costs results in an increase in output. Both the exact and approximate models concentrate

[41] Their position in terms of location $i \in L$ is not uniquely determined in space; the large urban agglomeration cities could have been located almost anywhere with sufficient transport access (river and coastal).

output in central areas of the U.S. that have the lowest average distance to all other cities.

5.2 Cost Heterogeneity Within Types: Terrain Cost Function

Now turn towards the structural model of link costs, specifically the component that depends on spatial heterogeneity in geography. The purpose of this part of the model is to infer costs for links that are not observed in factual railroad and canal networks. First, I import the observed railroad, canal, waterway, and coastal shipping transportation networks. The observed railroad network is a list of cities with position (x, y) in U.S. territory and a list of links with firm ownership specified. The railroad network consists of roughly 750 links and 250 firms. The process of importing the factual canal network is similar. I do not have canal ownership data, but I do observe canal names. I assume that each canal of a different name is owned by a separate firm.[42] This generates a canal network with approximately 100 links and 40 firms.

The waterway network is a list of links, rivers and inland lakes, that includes 30 major U.S. rivers. Waterway links are not owned by firms, so the result is a network with 250 links. Finally, I import the coastal shipping network. It is a synthetic network which consists of a list of links that proxy for the connectivity provided by the Great Lakes and Atlantic Ocean, including the Gulf of Mexico, in transporting goods by boat. This generates a network with 160 links.

Next, I merge these four separate networks together to form a single multimodal transportation network. To do this, I first import the grayscale terrain elevation map shown in Figure 17 (Appendix, part A); the black state borders and white coastal boundary are for illustrative purposes. The main geographical feature to notice in this map is Appalachian Mountains running from northern Georgia to Pennsylvania. Also note that the region with the roughest terrain in terms of elevation variability is the Northeast, so one expects that link costs are higher there. For the New York section of this map in more detail, see Figure 18. Difficult terrain, the hills and mountains that impede link construction, is represented by rapid changes in the grayscale gradient of the elevation map: the color of the pixels in the map changing from white to black rapidly, for example, as one ascends a mountain in the Appalachians. In the New York map, note that the observed path taken by the Erie Canal, the vertical then horizontal valley originating from New York City in the southeast corner of the New York map, was over relatively level terrain for the state: flat, without dramatic changes in elevation and along naturally-occurring waterways, and thus relatively cheap for canal construction. I also import a version of this map with rivers and lakes marked in blue, provided in Figure 19. This map identifies background natural waterways, termed "minor waterways" to distinguish them from the waterway network composed of navigable rivers that can be used as a means of transport, that affect

[42]This affects the counterfactuals only through its effect on the cost bounds determined from the observed canal network. Based on the ownership structure observed in the data, private firms with varying low to moderate levels of government assistance in the form of subsidies and loans, this assumption is reasonable.

canal and railroad link costs in the locality or cell, to be defined below. Figure 20 shows the version of the grayscale terrain elevation map used to mark coastal areas, denoted by light blue shading, which are both very favorable to canals and extremely unfavorable to railroads in terms of link construction costs.[43]

The terrain elevation map is divided into a grid of cells. I define a cell as a 6×6 array of pixels; the entire terrain map measures 220 by 260 cells. Now define the specialized functions used to compute link construction costs; this comprises the structural model of link costs that is used to determine the cost of building links that are not observed in the factual transport networks. Define $WaterDetect(i, j)$ as a function that returns the number of pixels of minor waterways in cell (i, j), where pixels of minor waterways are marked in blue in Figure 19. Let $CoastDetect(i, j)$ define an analogous function for coastal areas marked in light blue in Figure 20.

Now define the function used to compute the individual cell terrain penalty, which determines link construction costs in the model. Define **cell penalty function** $\Phi_{RR}(i, j)$ for railroads and $\Phi_{CN}(i, j)$ for canals at the cell (i, j) level, thus for $\Phi_\theta(i, j)$, type $\theta \in \{RR, CN\}$

$$\Phi_\theta(i,j) = 1 + (WaterPenalty_\theta)\,\frac{WaterDetect(i,j)}{36} + (CoastPenalty_\theta)\,\frac{CoastDetect(i,j)}{36} + (LandPenalty_\theta) \sum_{k \in cell(i,j)} \delta_k$$

where δ_k is the **terrain difficulty function** to be subsequently defined. Functions $\Phi_{RR}(i, j)$ and $\Phi_{CN}(i, j)$ are interpreted as the cost of crossing cell (i, j) using a railroad or canal in link construction, respectively. Type-specific parameters in these functions, which include $WaterPenalty_\theta$, $CoastPenalty_\theta$, and $LandPenalty_\theta$, are calibrated to the paths taken by observed link construction projects such that they obtain the values listed below.[44]

Table: Calibrated Cost Parameters

	Railroad ($\theta = RR$)	Canal ($\theta = CN$)
$WaterPenalty_\theta$	3	-1.5
$CoastPenalty_\theta$	9	-1.5
$LandPenalty_\theta$	10	10

[43] This is done to both aid canal building and prevent railroad construction over oceans and the Great Lakes, which is nonsensical. One can view these coastal areas as a nearly free medium for canal construction. However, the canal firm still needs to operate its service over the open ocean, which is costly.

[44] With three parameters for each type of link, I calibrate to three canal construction projects and three railroad projects observed in the data where their path in space is known directly. As a specific example, consider the case of the Erie Canal. These parameters are set such that in the model canal firms wanting to build the same link from New York City to Buffalo take the same least-cost path as seen in the data: up and through the center of the state using the Mohawk Valley. The same path for railroads takes a more direct route through a hilly intermediate region, also in correspondence with the data.

To interpret, railroad builders want to avoid all three categories of terrain obstacle: minor waterways, coastal areas, and uneven terrain; they especially dislike the coast since it is intractable for railroad construction. However, canal builders should stick to natural minor waterways and coastal areas while avoiding hills. This difference is a key component in the cost heterogeneity between railroads and canals as two competing modes of transport available in the 19th century. At this juncture, note that what matters for the final cost estimate is not the relative value of $\Phi_{RR}(i,j)$ compared to $\Phi_{CN}(i,j)$ but rather how costs vary individually for each type as i and j are varied due to geography.[45]

To compute terrain difficulty term $\sum_{k\in cell(i,j)} \delta_k$, associate each pixel in each cell with a grayscale index on the unit interval $[0,1]$. Difference this grayscale index pixel-by-pixel in sub-blocks, where each cell is split into two sub-blocks of 3×6 pixels each. This operation results in a length 18 array of differences $\{\delta_k\}$. To compute the total terrain penalty term in $\Phi_\theta(i,j)$, add across all pixels in the length 18 array associated with cell (i,j) and multiply by the appropriate type-specific $LandPenalty_\theta$ parameter. The terrain difficulty function is interpreted as the degree of terrain elevation unevenness or flux present in a particular cell; as the elements of $\{\delta_k\}$ increase, the cell's terrain is more hilly or mountainous and thus less conducive to inexpensive link construction.

Now translate this cell crossing cost $\Phi_\theta(i,j)$ into a link and type-specific cost measure for links between neighboring cells, termed the **neighbor terrain cost function** $\eta_\theta(\ell_{AB})$. For each interior cell, draw four links in four directions to adjacent cells: north, northeast, northwest, and west. The cost of building a link from cell A to neighboring or adjacent cell B, link ℓ_{AB} with associated cost $\eta_\theta(\ell_{AB})$ for type $\theta\in\{RR,CN\}$, is the average of the cell crossing costs $\Phi_\theta(i_A,j_A)$ for cell A and $\Phi_\theta(i_B,j_B)$ for cell B defined previously. This is

$$\eta_\theta(\ell_{AB}) = \frac{1}{2}\left(\Phi_\theta(i_A,j_A)+\Phi_\theta(i_B,j_B)\right)$$

for a link of type θ. To get a sense of what the neighbor terrain costs look like, see Table 4 (Appendix, part B) for summary statistics. Again, level differences between railroads and canals in this neighbor terrain cost are not necessarily meaningful; what matters is the dispersion within each type as i and j vary spatially.

Construct a weighted graph with these neighbor links and link weights given by neighbor terrain cost $\eta_\theta(\ell_{AB})$. This procedure generates a connected grid graph. Using their spatial position, I place all cities in the cell grid architecture and compute least-cost link construction paths between all 595 cities on the grid graph. This yields type-specific **terrain cost function** $\tau_\theta(i,j)$ for $\theta\in\{RR,CN\}$, which returns one

[45] Any arbitrary difference in levels between $\Phi_{RR}(i,j)$ and $\Phi_{CN}(i,j)$ will be absorbed in the type-specific link construction cost parameter c_θ to be discussed next. This is what is meant by $\Phi_\theta(i,j)$ only characterizing within heterogeneity for each type separately.

component of the total cost of building a railroad or canal link from city i to city j or $TC_\theta(i,j) = c_\theta \tau_\theta(i,j)$. See Table 5 for terrain cost function summary statistics; to repeat, level differences between types are not necessarily informative. For some selected examples of the paths taken by least-cost railroad and canal links using this cost minimization procedure on the part of transport firms, see Tables 21 and 22 for railroads and canals, respectively. To interpret, railroad firms want to avoid waterways and rough terrain when building links, while canal firms want to make use of waterways as much as possible in the construction of their links. This is why the example canal links in Figure 22 seek out the nearest body of water as soon as possible when routing from origin to destination city. It turns out that such behavior, driven by the meaningful spatial heterogeneity discussed previously, is why railroads and canals are not perfect substitutes in the model but rather complementary transport modes.

Figures 23 to 27 provide additional detail on the distribution of the terrain cost. Figure 23 shows railroad terrain cost function $\tau_{RR}(i,j)$ for a set of randomly selected links, where thicker red shading of lines indicates that the cost is relatively higher within the distribution of $\tau_{RR}(i,j)$. Railroads are most costly to build in the Northeast due to uneven terrain since minor waterways are more evenly distributed across the United States. I provide the actual distribution of $\tau_{RR}(i,j)$ in the following illustration, Figure 24. Compared to canals, railroads have relatively more inexpensive links and very expensive links among the set of all links that could potentially be built between city pairs. The railroad terrain cost empirical distribution is single-peaked at a moderate cost level. Figures 25 and 26 are analogous for canals. Note that the canal terrain cost $\tau_{CN}(i,j)$ is less uniform across space; this is reflected in both the map and empirical distribution. Canals are only penalized by the uneven terrain prevalent in the Northeast, with moderate to low construction costs everywhere else in the country; costs in the Midwest and South are especially low relative to the Northeast. Figure 27 compares the two type-specific empirical distributions under the normalization that each of their means is individually set equal to one. Again, canal costs are less uniform and the peak occurs at a higher relative cost while railroad links are either cheap or expensive with a larger right tail of the distribution. This right tail for the railroad distribution corresponds to building links over the waterways and hills of the Northeast in general and Vermont in particular.

5.3 Transport Firms and Equilibrium

Assume an initial set of transportation firms T with $|T|$ members. Although $|T|$ is a fixed parameter that is part of the stable network search algorithm, firms are allowed to remove their last link and exit the economy so the number of firms in the final stable network is endogenous. Transport firms own and operate connected pieces of the transport network or their individual subnetworks. The objective of a transport firm is to maximize its profits by choosing the links in its subnetwork subject to the choices made by all other firms. Firms incur costs by building links and collect revenue from the traffic that uses their links in trade. There are two permanent types of transport firms: railroad (RR) and canal (CN). Each type can only build links of its type, so no firm can own links of both types.[46]

All transport firms are randomly seeded with one or more initial links individually when the network stability algorithm is initialized.[47] Thus every firm starts with at least one transport link. Link seeding is uniformly random over the set of potential links, which differs from the complete set of links in two ways. First, a firm cannot build a link where one already exists.[48] This is equivalent to saying that firms cannot directly build on the links of competitors to divert traffic from them.[49] Second, I restrict the set of potential links such that every link in the set lies below some maximum allowable distance threshold. This restriction is motivated by the data and computational tractability: observed links are usually short and connect neighboring cities, and restricting the set of potential links means fewer link construction deviations by firms to check. However it is important to note that the resulting set of potential links goes roughly an order of magnitude beyond nearest neighbor links in terms of maximum allowable distance between origin and destination cities.

For computational feasibility, I consider equilibria where firms only consider unilateral single link deviations when trying to improve their profits in the stable network search algorithm.[50] At each iteration of the algorithm that searches for stable networks, there is a uniformly random arrival of a decision opportunity for some transportation firm. When a transport firm has a decision opportunity, it can either do nothing and retain its current subnetwork, remove a link it already has in its subnetwork, or build a new link in its set of candidate links to add. At most one link can be removed or built given a decision opportunity. The firm always has the option to take no action, maintaining the status quo. The active firm with a decision opportunity makes its discrete choice over the firm-specific set of links to maximize its profits. Firms are not

[46] For the counterfactuals this assumption is never used because only one of the two transport technologies is available, thus one firm owning links of both types is impossible. In the data, which motivates this assumption, one does not observe a firm owning links of both types.

[47] Any additional links beyond the first are randomly assigned to firms to guarantee that the initial network is connected. This is an imposed convention for the stable network search algorithm that one can think of as an equilibrium selection device in the presence of multiplicity. It is a particularly generous equilibrium selection protocol since firms are less likely to remove links that isolate cities, cutting them off from the rest of the transport network. This generosity implies that one can think of the subsequent output loss results as a lower bound on the true output loss in the presence of multiplicity.

[48] Alternatively, multigraphs are not allowed; the transport network is simple graph.

[49] Presumably, if such behavior by firms was allowed, a firm could replicate a competitor's link and then receive half of their previous traffic in an equal split. This type of action is ruled out since it is not observed in the data. However firms can easily build parallel links that divert much of competitor traffic on any particular transport link.

[50] The single link deviation assumption is made to render the problem computationally feasible for simulation. The firm has to consider all of the options available and choose the optimal one through exhaustion, which is a computationally-intensive combinatorial problem. The firm has to run all possible link removal and construction scenarios in order to act fully optimally, where deviations consist of sets of links. It is not possible to check all multiple link deviations in plausible running time for the algorithm. The single link assumption is reasonable because many of the links available in the set of potential links are long links between distant cities that would have been considered as the union of multiple single links if links were defined as only between close neighbors.

allowed to act without a decision opportunity. There is no time dimension here and one should not think of one iteration of the algorithm as a discrete time unit, but rather a computational convention where one firm is randomly selected to act next.

The firm's link construction problem is constrained in the following way. For any pair of cities $i, j \in L$, only one link ij can be built directly between them. If a link already exists between a pair of cities, no additional links can be built there by any transport firm.[51] New links must be built from a city that is the terminus of a pre-existing link owned by the firm to another city.[52] To rephrase, a firm's links cannot jump around from one end of the United States to the other; firm subnetworks are individually connected. Again this requirement is motivated by the data and computational feasibility: one rarely observes firm subnetworks in the data that are not connected and such a restriction limits the number of link construction alternatives to check. Thus the set of links a firm can add is conditioned on its identity, which is fully summarized by the links of its subnetwork and its type.

Let $g = \{\mathcal{S}_t\}_{t \in T} \cup G$ denote the state of the multimodal transportation network, where $\mathcal{S}_t$ is the **subnetwork** owned and operated by firm t and G is the union of the road, waterway, and coastal shipping networks as part of the natural resource endowment. Thus G is exogenous, taken as given by transport firms. Denote the set of existing links by Λ. Let p_{ij} denote the **link price** per unit of value shipped across link ij per mile, defined as the hauling price for goods from city i to j. By convention p_{ij} is expressed as a nominal price in cents per mile per dollar of the value of goods being shipped.[53] The interpretation is that p_{ij} is charged to the shipper of the goods at each step of the journey from the source city to the destination, where the total resulting usage fee depends on the initial value of the goods determined at the origin city i and the length of link ij as measured by the Euclidean distance between cities i and j.[54] These prices are not decision variables for transport firms by assumption, thus firms take them as given and their choice is restricted to the set of links to build and not how they are priced.[55] Links are undirected: $ij \in \Lambda$ implies that $ji \in \Lambda$ so the transport network is symmetric with respect to links between individual pairs of cities. Thus no shipping routes are more costly in a particular direction: for all $ij \in \Lambda$ $p_{ij} = p_{ji}$.[56] Therefore one can drop the ij subscript and denote the nominal link price by p since it does not vary across links.

[51] However, that link could be removed by the firm and then built by another firm.

[52] One can think of this restriction as granting firms a sort of local monopoly on link construction since they will not stray too far from their initial links (which are randomly seeded to initialize the algorithm).

[53] This price can also be expressed as an iceberg cost that is bounded on $(0, 1)$. The interpretation is that fraction p_{ij} of the goods in transport are taken as a usage fee by the appropriate transport firm.

[54] This implies that the total cost of travel for a route is additive in the individual costs incurred at each link along the route. The link owner at the start of a route will receive the same payment in usage fee for traffic as the owner at the end of the same route, provided that their links are the same length in Euclidean distance; they will always be paid the same per mile per dollar of cargo shipped.

[55] Again this is done for computational tractability.

[56] One can conceptualize this as no "uphill" or "downhill" routes, where the firm would want to charge a higher price for uphill routes since they use more fuel and effort to traverse, or a limit on price discrimination.

Let R_{ij} denote the set of routes from city i to j given g. Thus g implies Λ, which in turn implies R_{ij}. If $R_{ij} = \emptyset$, then i and j are not connected and no goods can flow between them. This cannot occur in the model by assumption since roads are assumed to exist everywhere as the costly outside option of last resort. Route $r_{ij} \in R_{ij}$ of length $\ell + 1$ is of the form

$$r_{ij} = (ik_1, k_1k_2, \ldots, k_\ell j)$$

with intermediate cities $k_1, k_2, \ldots, k_\ell \in L$ on the route; the associated total cost of the route is

$$C(r_{ij}) = p \sum_{ij \in r_{ij}} d_{ij}$$

where p is the link price, d_{ij} is the Euclidean distance from city i to j, and $C(r_{ij})$ is imposed on each unit of value of goods shipped via r_{ij}. Thus total usage cost is additive in the individual link prices per unit distance.

The total cost of shipping x dollars of a good from i to j using route r_{ij} is $C(r_{ij})x$ dollars, or pDx dollars where p has units dollars per mile per dollar shipped, total route length D is in miles, and cargo value x is in dollars. The minimized unit cost of shipping from i to j or vice versa is

$$C^*(i,j) \in \arg\min_{r_{ij} \in R_{ij}} C(r_{ij})$$

with associated least-cost route $r^*_{ij} \in R_{ij}$, which is not necessarily unique given g but regardless does not depend on cargo value x. Given the set of least-cost routes resulting from the multimodal transport network, traffic flows over individual links are computed from the total volume of trade flows since trade between cities i and j is assumed to always take the least-cost route r^*_{ij} in terms of shipping cost $C^*(i,j)$ dollars as fees paid to the transport sector and transport firms on the route per unit of cargo value. Total profits for transport firm t with subnetwork $\mathcal{S}_t$ are determined based on endogenous traffic flows, which depend on the multimodal transport network to determine trade costs and the output distribution, and exogenous link prices $\{p_{ij}\}_{ij \in \mathcal{S}_t}$ or simply p as a parameter the firm takes as given.

Heterogeneity in geography within types is captured by terrain cost function $\tau_\theta(i,j)$ which affects link construction and maintenance costs for transportation firms.[57] Construction and operation of a link between cities $i, j \in L$ has associated total cost $TC_\theta(i,j) = c_\theta \tau_\theta(i,j)$ that scales linearly in $\tau_\theta(i,j)$, where c_θ is a scalar parameter termed the **type cost** that varies across link type for $\theta \in \{RR, CN\}$. This cost is paid

[57] Intuitively $\tau_\theta(i,j)$ captures the idea that canals are less expensive and railroads more expensive to build due to the local presence of small-scale waterways (streams, rivers, lakes, etc.). Both transport modes are penalized by hills, mountains, valleys, and so on (steeper terrain gradient).

by transport firm t for each link ij built in its subnetwork $\mathcal{S}_t$. **Total cost** for firm $t \in T$ of type θ with subnetwork $\mathcal{S}_t$ is

$$TC_\theta(t) = \sum_{ij \in \mathcal{S}_t} c_\theta \tau_\theta(i,j)$$

where $TC_\theta(t)$ is interpreted as total annualized link construction and operating cost for the firm.[58] Firm t computes its profits $\Pi_\theta(t)$ by taking total revenue $TR_\theta(t)$ derived from traffic over its links and subtracting total cost $TC_\theta(t)$ based on its subnetwork $\mathcal{S}_t$, type cost c_θ, and terrain cost $\tau_\theta(i,j)$. The objective of the firm is to maximize its profits $\Pi_\theta(t)$ by choosing its subnetwork or links $\mathcal{S}_t$, taking as given the subnetworks of all other firms.

The stable network search algorithm is designed to find equilibria of a modified version of the simultaneous link announcement game originally proposed by Myerson (1977). This is a static one-shot network formation game where the players are profit-maximizing transport firms. Unlike the original game proposed by Myerson, players are firms characterized by their subnetworks and not cities or nodes in the network. In the simultaneous link announcement game, each firm announces a set of links $\mathcal{S}_t$ to build in the first stage. The multimodal transport network g is built in the second stage and firm profits $\Pi_\theta(t)$ are determined according to the approximate gravity model for total firm revenue $TR_\theta(t)$ and the structural model of link costs for total firm costs $TC_\theta(t)$ defined previously. The decision space of the firm is restricted according to the rules discussed previously: firms must choose connected subnetworks to build and their links cannot overlap with those of their competitors.[59]

The solution concept I employ is a weaker version of Nash stability where the network is robust to single link deviations by individual firms, termed **Single Link Nash Stability** (**SLNS**). Denote the resulting stable network by $g^* = \{\mathcal{S}_t^*\}_{t \in T} \cup G$, where G is the exogenous network of natural waterways and roads, which satisfies SLNS such that no individual firm has a unilateral incentive to revise its links by changing one link at at time. In practice the modified link announcement game under SLNS induces multiple equilibria, so one can refer to a general class of stable networks $\mathcal{G}^*$ where $g^* \in \mathcal{G}^*$ is a member and the stable network

[58] Annualized at a yearly interest rate of 8% (consistent with U.S. data around 1861), where annually recurring operating costs are assumed to be 20% of initial link construction cost (which is annualized on a 30-year basis using the interest rate). To rephrase, the firm pays 8.8% of the initial construction cost per year as an annual payment on a 30-year basis and also pays 20% of the construction cost in operating costs per year, for an annualized total of 28.8% of the initial construction cost every year. The story one should have in mind is the stable network prevailing forever with no change, but the firms build their links in the first period and then repay those costs on a 30-year basis plus incurred operating costs. Use of a particular 30-year basis for repayment of initial costs is not important for the results, and in general annualization only affects interpretation of the total cost estimates in Tables 6 and 7.

[59] Think of the firm incurring an arbitrarily large penalty if these rules are not followed. Thus in equilibrium the set of links announced by the firm to build will conform with the requirements of connected individual subnetworks and disjoint subnetworks across firms. Again these restrictions on the firm's decision space are imposed for computational tractability when simulating the model to find stable networks. These restrictions could equivalently be embedded in the solution concept as an equilibrium refinement or in the strategy space directly instead of the payoffs.

search algorithm is exploring this space and finding some member g^* each time it is run. I do not claim to fully enumerate the members of $\mathcal{G}^*$ since there are likely millions if not trillions of them and a full accounting is not feasible to compute in fixed running time. Rather I sample on the order of 100 members from $\mathcal{G}^*$ and construct a sample of the set of stable networks that satisfy the SLNS criterion.[60]

Now I provide the definition of equilibrium used in this paper, termed **Decentralized Transport Network Equilibrium** (**DTNE**), which makes use of the SLNS criterion to restrict the set of permissible transport networks.

Definition (Decentralized Transport Network Equilibrium) An equilibrium in the decentralized transport network formation model consists of output distribution $\{Y_i\}_{i\in L}$, transport costs $\{c_{ij}\}_{i,j\in L}$, and transport network $g^* = \{\mathcal{S}_t^*\}_{t\in T} \cup G$ such that conditions (1)-(3) hold:

(1) Transport costs $\{c_{ij}\}_{i,j\in L}$ are from g^* using least-cost routes

$$c_{ij} \in \underset{r_{ij}\in R_{ij}}{\arg\min}\, C(r_{ij})$$

(2) Given transport costs $\{c_{ij}\}_{i,j\in L}$, output distribution $\{Y_i\}_{i\in L}$ is from approximate gravity trade model

$$Y_i = \bar{Y}^{\frac{\alpha_2}{1-\alpha_1}}\, \bar{c}(i)^{\frac{\alpha_3}{1-\alpha_1}} \quad , \quad \bar{Y} = (\sum_{i\in L} \bar{c}(i)^{\frac{\alpha_3}{1-\alpha_1}})^{\frac{1-\alpha_1}{1-\alpha_1-\alpha_2}}$$

(3) Transport firms $t \in T$ of type $\theta \in \{RR, CN\}$ maximize profits

$$\Pi_\theta(t) = TR_\theta(t) - TC_\theta(t) = \sum_{ij\in\mathcal{S}_t^*} (pd_{ij}X_{ij} - c_\theta \tau_\theta(i,j))$$

in the transport firm link announcement game such that all individual members of $\{\mathcal{S}_t^*\}_{t\in T}$ satisfy the Single Link Nash Stability (SLNS) criterion, taking G, the combined road/waterway/coastal network, as given.

Note that the restriction on the set of stable networks imposed by SLNS is weak such that the multiplicity problem is rather sharp here.[61] To resolve this problem I develop a simulation method via the stable network search algorithm to find members of the set of networks satisfying SLNS. I then report the median outcome for the output loss among members of this set as an answer to the empirical question posed at the start of the paper. This point estimate is accompanied by 90% coverage intervals that cover 90% of the sampled

[60] This sample is not truly random but rather conditional on the initial network allocating as least one link to each firm and sufficient additional links to make the initial network connected. This is imposed as a convention or alternatively an equilibrium selection rule.

[61] This is because of the high dimensionality of the problem in terms of players, cities, and possible networks to form (not just the network as a whole, but individual firm subnetworks as well). Additionally, forming a link is a lumpy investment discrete choice problem, you either build/remove a link or not, and a large class of opponent actions will keep your payoff relatively unaffected such that "do nothing" is still the optimal choice, aggravating the multiplicity problem.

equilibria to quantify the uncertainty associated with the multiplicity due to weakness of the SLNS solution concept.

6 Simulation Method

6.1 Initialization: Computing Factual Objects of Interest

With terrain cost function $\tau_\theta(i,j)$ in hand for each type $\theta \in \{RR, CN\}$, I define the set of potential links from which counterfactual links are drawn as options for link construction in the search algorithm for stable networks. For railroads and canals, the set of potential links is a random sample of the set of all links below a specified distance threshold.

The usage cost of the outside transport option, roads or wagon trails, is normalized to one. Under this normalization, according to Table 3 (Appendix, part B), in 1860 the relative cost of railroad usage in the data is 0.13 and, for canals, 0.06. However, the observed gap between railroad and canal prices is due to quality differences, specifically that railroads were faster and more reliable than canals.[62] Certainly the first-order difference is that between roads and all other transport modes, not between canals and railroads; this is also seen in Figure 4 (Appendix, part A) on observed prices across transport modes. To abstract away from differences in quality between the two transport modes affecting their relative price, I shut down the price differential channel entirely by setting the usage cost of both modes equal to the average of their observed values or 0.098 relative to roads.[63] This is indicated in the rightmost column of Table 3 as the assumed normalized cost for both modes, including natural waterways and coastal shipping, which is interpretable as 9.8% of that of roads and used to run the counterfactuals.[64]

The computational procedure used to generate transport cost matrix $[c_{ij}]_{i,j\in L}$, which enters the gravity trade model discussed previously to determine the output distribution, is as follows. Given railroad, canal, waterway, and coastal networks, for each pair of nodes, reduce their bilateral distance by the appropriate relative freight rate if a non-road link exists. Thus the baseline is 1 since roads are present everywhere, and the presence of a railroad or canal reduces the bilateral trade cost to 9.8% of its initial value. In the case of more than one link for a single pair of nodes, use the lowest-cost option.[65] The least-cost routes of the resulting weighted network are the observed transport costs c_{ij} as seen in Figure 16.[66] Since the observed

[62] Railroads were roughly five times faster than canals and could operate throughout the winter when canals would freeze over and become unusable.

[63] This is the midpoint between railroads, 0.13, and canals, 0.06, as seen in the data.

[64] This is done to focus on the effect of the presence of links or not rather than the relative price effect of how much links of different modes cost to use for the shipper relative to each other.

[65] For example, a pair of cities has both a rail link and a canal. This contingency is not observed in the data and cannot occur in the counterfactual by assumption. Also by assuming that all non-road options are identically priced, it is unnecessary to break ties. All that matters is if there is a non-road option present over a link or not.

[66] All that matters for the least-cost routing problem here is relative transport costs, not their absolute level.

network is connected, transport costs are almost universally low with some minor exceptions in the Midwest.

With observed transport costs in hand, now discuss computation of transport firm payoffs in the form of total revenue $TR_\theta(t)$ for firm $t \in T$. An exact computation of payoffs would compute all bilateral trade flows between nodes, which are determined according to the gravity trade model discussed previously. Trade flows, which can be interpreted as exports from the perspective of an individual city within the U.S., take the least-cost route between source and destination cities. Each firm owns a set of links $\{ij_1, ij_2, \dots\}$, its subnetwork $\mathcal{S}_t$, that completely characterizes the firm along with its type $\theta \in \{RR, CN\}$. For link ij, add up the link-level trade flow X_{ij} multiplied by the price of each link p times its length in Euclidean distance d_{ij} to determine total revenue $TR_\theta(t) = p \sum_{ij \in \mathcal{S}_t} X_{ij} d_{ij}$ for each firm t.[67] Computationally, this is slow and takes significant computer time. With 595 nodes, it is simply too slow to run counterfactuals this way since trade flows need to be recomputed thousands of times before the counterfactual network is declared stable.

As an alternative to the computationally intractable exact solution to trade flows, I approximate by randomly sampling bilateral trade flows. Specifically I sample 0.1% of total flows. This method is reasonable since the fraction of bilateral flows sampled is constant throughout both observed and counterfactual scenarios. Since link construction costs are determined inside the model, random sampling from the payoff or total revenue function does not affect model outcomes since the model scales down costs appropriately through adjustment in type cost c_θ for each type to deliver an observed network that satisfies the SLNS criterion. Thus the fraction of flows sampled does not affect the end result except in adding noise to individual stable network search algorithm runs because of the random sampling procedure. This added noise is not severe in practice and disappears as the number of simulation replications increases.

6.2 Cost Heterogeneity Between Types: Type Cost

I use observed railroad and canal networks to bound cost parameters for building and operating railroad and canal links. Define c_{RR} and c_{CN} as the **type cost** associated with an individual link for railroads and canals, respectively, which can be interpreted as the annualized present discounted value of link construction and operating costs when multiplied by the appropriate terrain cost for each type for link ij. Thus the total cost of building link ij between cities i and j can be expressed as $TC_\theta(i,j) = c_\theta \tau_\theta(i,j)$, where c_θ is the type cost bounded by observed railroad and canal networks and $\tau_\theta(i,j)$ is the terrain cost defined previously. Note that c_θ does not depend on the specific identity of link ij while $\tau_\theta(i,j)$ does due to the heterogeneity provided by geography. Thus all the within-type heterogeneity due to geography is captured by the terrain cost and the role of the type cost is to reconcile firm payoffs with costs for each type and allow comparability in total link costs across types.

[67] Note that this payoff function has no closed form due to least-cost routing and must be simulated.

This provides bounds on c_{RR} for railroads and c_{CN} for canals, thus type costs are type-specific but no variation is present within each type or across firms. Assume that the observed 1861 U.S. transportation network is single link Nash stable, specifically railroad and canal networks individually satisfy the SLNS criterion defined previously. This implies that observed individual firm subnetworks cannot be improved upon in terms of profitability via unilateral single link deviations executed by the firm. An example of a firm's subgraph is provided in Figure 31 (Appendix, part A) for the South Carolina Railroad, which consists of six links between cities in South Carolina and Georgia. The firm's subgraph is connected and disjoint relative to the union of the subgraphs of railroad and canal competitors, consistent with the restrictions I impose on firm subgraphs in the model. Unilateral single link deviations can be placed into two categories: removal of owned links and addition of new links.

Consider first the unilateral removal of single links present in the observed network by individual firms. To be specific this is a proposed alteration of a firm's subnetwork $\mathcal{S}_t$, which includes all owned links. For any link that exists in the network, the firm that owns and operates it must not improve its profitability through removal of the link. Intuitively, this is because the firm chose to build the link and had the option to remove it in the observed world, thus the link must be individually payoff-improving. This is expressed as the following inequality for set of observed firms of type θ T_θ

$$\forall \theta \in \{RR, CN\},\ \ \forall t \in T_\theta,\ \ \forall ij \in \mathcal{S}_t \qquad -\Delta TR_\theta(t) \geq -\Delta TC_\theta(t) = c_\theta \tau_\theta(i,j)$$

for railroads and canals, respectively, for all observed links (i,j) in factual subnetworks $\mathcal{S}_t$ over all railroad and canal firms. In words this inequality says that the loss in revenue must weakly outweigh the gain in terms of reduction in subnetwork cost when an observed link is removed. In practice, I take an average instead of the minimum over this expression for all links in the factual network to provide an upper bound on c_{RR} and c_{CN} since $\tau_\theta(i,j)$ has already been determined.[68] This is done to abstract away from heterogeneity at the firm level within firms of the same type.[69] An example is provided in Figure 33, where for the South Carolina Railroad the observed rail link between Augusta and Branchville is severed; the traffic loss must have exceeded the cost savings for the firm if its subnetwork satisfied the SLNS criterion, which holds by assumption.[70] Repeating this argument across all observed firms provides an upper bound on c_θ for each type since link costs could not have been too high to force the firm to remove links it has built in the factual world.

[68] Again the level of $\tau_\theta(i,j)$ is arbitrary and any scaling of this function which does not affect relative costs within each type, for example scaling by some factor k such that the new terrain cost is $k\tau_\theta(i,j)$, is absorbed into c_θ by construction.

[69] Although it would certainly be interesting to consider heterogeneity between firms of the same type, it is in my view unnecessary to answer the empirical question at hand and is left to future work.

[70] One can think of this is an internal consistency requirement which says that since counterfactual networks are determined according to the SLNS solution concept, so was the observed network.

Now consider the unilateral addition of single links that are not found in the observed network. For any link that does not exist in the factual network, the firm must not improve its profitability by adding the link to its subnetwork. This is expressed as the reverse of the previous inequality for link removal

$$\forall \theta \in \{RR, CN\}, \ \ \forall t \in T_\theta, \ \ \forall ij \notin \mathcal{S}_t \qquad \Delta TR_\theta(t) \leq \Delta TC_\theta(t) = c_\theta \tau_\theta(i,j)$$

for railroads and canals $\theta \in \{RR, CN\}$, respectively, for all absent links ij not found in subnetwork $\mathcal{S}_t$ over all firms. Additionally any new link ij must satisfy the usual conditions that it is not in any firm's observed subnetwork and starts from a node i covered by the firm's current set of links. In words, the costs in terms of link construction and operation must exceed the benefits in additional traffic and revenue for links that are not observed. Again, I take an average instead of the maximum over this expression for all links to provide a lower bound on c_{RR} and c_{CN}. An example is given in Figure 32, where the South Carolina Railroad attempts to extend their rail line from Columbia to Alston; the cost of the link must have outweighed the benefits in terms of attracting additional traffic to the firm's subnetwork since it was not observed in the factual network. Again, repeating this procedure over all firms of both types provides a lower bound on c_θ since link costs must have been high enough to prevent firms from building links that are not observed in the factual world.[71]

This procedure provides bounds on link type costs c_{RR} and c_{CN} for railroads and canals, respectively. Table 6 (Appendix, part B) shows the resulting estimates for type cost bounds. The inequalities are listed in the first column, separated by type for railroads and canals. Unilateral single link removal provides an upper bound on link cost in row 1; from unilateral single link addition, row 2 provides a lower bound on the link cost. The bound itself in the second column in is arbitrary model units, which is not interpretable outside of the model itself. Thus for comparison purposes I provide alternate real-world interpretations of the bound numbers in more useful units in the next four columns, assuming that a link of average cost is proposed for construction for each type. In terms of annualized costs, the links themselves were relatively cheap in terms of observed output in 1861: the average railroad link cost ranges from 0.03% to 0.18% of output in annualized terms, and for canals the range is 0.03% to 0.06%. The main conclusion from this table is that, on average, the model says that railroads are more expensive to build and operate than canals. In thousands of 2012 dollars per mile, the annualized link cost bounds are $[75, 396]$ for railroads and $[69, 140]$ for canals. These numbers are reasonable when one takes into account that they are interpretable as the

[71] For computational feasibility, I randomly sample links to add via this procedure from the set of potential links for each type. Whereas one can exhaust the set of possible link removal deviations in factual networks since only 1000 total links of both types were built, checking all link addition options is not feasible since with 595 nodes in the network the number of links in the set of potential links to add is too large to fully exhaust. I sample a sufficient number of link addition options such that the resulting lower bound on costs is sufficiently stable to additional sampling.

annualized construction and operating cost for the average link, where the average link is quite long at 550 miles in length, far above average link length in the data.

Combining these bounds and taking the midpoint of the resulting interval, one obtains a point estimate for each type cost, which is subsequently used in the network formation counterfactuals. Point estimates are presented in Table 7. The point estimates in the second column are midpoints of the intervals in the first column, sorted by type for railroads and canals. I find that the average railroad link is more than twice as expensive to build and operate in annualized terms as the average canal link. However, these averages can be deceiving since so many links are possible and only a small fraction of them are built in either the observed network or the counterfactuals. Also one would expect that built links in the counterfactuals are a selected sample from the set of potential links, specifically built links should be substantially cheaper than average. Thus one cannot judge from relative link costs alone in terms of the relative output loss when one of the transport modes is removed, and it remains to be seen in the counterfactual simulations of SLNS networks whether output is higher with railroads only or canals only. It turns out that in both cases output falls by nearly 50%, and due to uncertainty over the set of equilibria one cannot say in which case the output loss is higher.

6.3 Stable Network Search Algorithm

I now outline the **stable network search algorithm** which is designed to find equilibria of the transport network formation game, or counterfactual networks that satisfy the SLNS criterion as members of the set of Decentralized Transport Network Equilibria or DTNE individually. First, initialize variables at iteration $k = 0$. Specify the maximum distance to consider for links in single link addition unilateral deviations by firms, 1/4 times the mean of bilateral distance 550 mi., and the number of addition and removal alternatives to consider per iteration, three addition and one removal. Terrain costs $\tau_\theta(i,j)$ were computed previously and firm type cost parameters c_θ are taken from the first stage where I constructed and estimated cost bounds. Again, this yields total link cost $TC_\theta(i,j) = c_\theta \tau_\theta(i,j)$ for each type $\theta \in \{RR, CN\}$. Next, specify the number of firms to initially seed with at least one link each: 500 firms for both railroads only and canals only counterfactuals. The number of initial firms is chosen to intentionally overshoot the number of firms in stable equilibrium networks since exit is allowed: a firm exits by choosing to remove its last link and receive a payoff normalized to zero.[72] Assign each firm a random initial link in the set of potential links, which is a random sample of the complete set of possible links with the selection criterion that all links must have length less than the maximum distance threshold. This allows construction of the initial counterfactual network, firm subgraphs, and list of active firms. I also require that the initial network is connected such

[72]Firm exit is equivalent to the firm continuing to operate and earning zero profits in perpetuity.

that no individual cities are isolated from the others. To achieve this, I randomly add links to the first set of 500 until such connectivity is obtained, thus firms start with at least one link.[73]

In the canals only counterfactual without railroads, which speaks to Fogel's question on the marginal growth contribution of railroads, the railroad network is set to empty and the canal network is randomly drawn initially and subject to the stable network search algorithm. In the railroads only counterfactual with no canals, the canal network is permanently empty and the railroad network is determined by the algorithm in an analogous way. In both counterfactuals, canals only and railroads only, the combined waterway, coastal shipping, and road network G is taken as given from the data.[74]

At iteration k, the simulation algorithm proceeds as follows. Start by selecting a random firm from the firm list. Compute the distance from all nodes in that firm's subgraph to all other nodes in the set of potential destination cities; drop those city pairs with distance above the maximum distance threshold. Place the remaining pairs into a set of possible links to build. Take a random sample from this set of size specified previously, three possible links to add, then check for duplicate links with the current state of the counterfactual network and remove them.[75] Next, evaluate payoffs over all possible link construction alternatives for the firm. Here, the space of alternatives is comprised of at most three link construction alternatives, one link removal alternative,[76] and one alternative where the firm maintains its current network by doing nothing. The firm then chooses the action that maximizes its profits.[77] Once the decision rule has been determined, update the counterfactual network, firm subgraphs, and list of firms if necessary from iteration k objects to iteration $k+1$ objects to prepare for the next iteration. Repeat until network stability is achieved in terms of the SLNS criterion.[78] In the context of the algorithm, this means that all firms choose to do nothing when they receive a decision opportunity and desire no revision in their subnetworks.

7 Results

7.1 Canals Only Counterfactual

Figures 34-36 (Appendix, part A) show one possible evolution of the counterfactual canal network over algorithm iterations after starting with 500 firms, randomly seeded with one link each and any additional

[73] Firm ownership is also randomly assigned to these additional links necessary to obtain an initially connected network.

[74] These networks, endowed by nature, are also time-invariant, just like the empty railroad network in the canals only counterfactual world where the railroad was never invented.

[75] No multigraphs allowed, i.e. no building directly on competitor links to siphon their traffic.

[76] The link removal alternative is randomly chosen from the firm's set of links (i.e. the firm's subnetwork).

[77] Recall that the firm's action space is building a link, removing a link, or doing nothing. Network stability occurs when all firms choose to do nothing when given the chance to act.

[78] According to some stopping criterion, such as sufficiently limited movement in total output or the number of links for some period in terms of algorithm iterations. One can think of the resulting stable networks as approximately stable since fully verifying the SLNS criterion is not computationally feasible: there are too many possible unilateral single link addition deviations to check for each firm in reasonable running time for the algorithm.

links to guarantee that the initial multimodal transport network is connected. Again this is the counterfactual world where the railroad was never invented considered by Fogel, leaving the United States with only canals to build up its transport network. The majority of link removal and firm exit occurs in the first 500 iterations due to the overshooting nature of the algorithm which seeds the economy with too many transport firms and links initially.[79] The network is approximately stable by 2000 iterations with minimal movement in total output or number of links. It is important to stress that these figures are only for one run of the stable network search algorithm and are not necessarily representative of the class of SLNS canal networks with railroads removed.

In the example algorithm run in Figures 34-36, one sees that canals are built primarily to fill in the gaps in transportation network coverage between natural waterways. Compared to the observed network, canals alone do a poor job of covering inland areas in the South and Midwest. Most cities are still covered by the multimodal transport network in the final stable network, with isolated cities primarily in Illinois, Indiana, North Carolina, and Pennsylvania. However, the resulting output loss is still large because of an increase in average transport costs: the average path length between any two locations increases in the counterfactual canal network, even if they are connected, resulting in high costs. As one would expect, most of the counterfactual canals are built along adjacent rivers or bodies of water due to low link construction costs. Output is concentrated in river corridors, specifically the Mississippi-Ohio River to Lake Erie and Susquehanna River to Lake Ontario.

7.2 Railroads Only Counterfactual

Figures 37-39 present an example counterfactual railroad network after starting with 500 firms in the railroads only, no canals counterfactual. This is the opposite counterfactual world where canals were never introduced to the U.S., but the railroad was still invented. Again, the majority of link removal and exit occurs in the first 500 iterations due to intentional overshooting of the number of transport firms and the network is stable by 2000 iterations. Note that there are fewer links and firms in the railroads only counterfactual relative to the canals only counterfactual. Railroads are built inland to cover holes in the waterway network, with dense clusters in Illinois, Ohio, and Pennsylvania. Most output occurs in the Midwest and coastal Northeast, with a number of isolated cities in the South and inland Midwest. As one would expect from spatial variation in terrain costs for railroads, link construction tends to avoid minor waterways and coastal areas in the counterfactuals.

[79] This is done to avoid the coordination problem that no firm wants to build the first link in the empty network when output is low according to the gravity trade protocol. Thus the overshooting initialization is conservative in the sense that the output loss could be worse when starting the algorithm from the empty network.

7.3 Comparing Counterfactuals

Now turn towards comparing the canals only and railroads only counterfactuals on a number of model outcomes. Table 8 (Appendix, part B) provides the main results with both median point estimates and 90% coverage intervals to account for multiplicity of equilibria in the form of stable networks.

Counterfactual output series are plotted in Figure 40; they have been rescaled with factual output normalized to one. This plot, which provides counterfactual output losses, answers the primary research question as to how much output falls when either railroads or canals were never invented. When network stability is obtained according to the SLNS criterion, relative to the factual world, output is 45% lower in the canals only counterfactual and 49% lower in the railroads only counterfactual. However, I cannot say in which case the output loss is larger due to uncertainty over the multiplicity of equilibria: the 90% coverage intervals overlap when network stability is obtained near iteration 2000. In both cases the output loss is large compared to the previous literature, including what Fogel found and results obtained in the subsequent literature by Fishlow (1965) and Donaldson and Hornbeck (2013). Geography induces a strong complementarity between railroads and canals, since construction of one type of transportation link increases the total volume of trade and may make owners of the other type better off since heterogeneity in terrain keeps the two modes at arm's length, out of direct competition with each other. It appears that empirically railroads and canals are strategic complements, not strategic substitutes, so the output loss is severe when one of the competing transport modes is removed from the technology set of economy.

Figure 41 shows the variance of output for both counterfactual networks; it is normalized such that the variance of output in the factual is one. This plot is qualitatively similar to the previous one for output: the variance of output falls by 69% in both cases, canals only and railroads only counterfactuals. There is nearly equal average output dispersion in both counterfactuals, which implies less inequality across cities in variance terms. However, with only one of the transportation technologies available, relative inequality across cities in terms of the maximum-minimum gap would have been sharply higher because as Figures 34-36 for canals only and Figures 37-39 for railroads only show, many cities are cut off from the transport network entirely in the counterfactuals.[80] The model is uninformative for economic inequality across individuals, but says that inequality rises across cities in terms of the gap between maximum and minimum city output.[81]

Figure 42 plots the number of links in the counterfactual network against iterations of the stable network search algorithm. By assumption, the network starts with 500 firms and at least 500 links, with sufficient

[80] Although their relative populations (for the isolated cities) are low as a result.

[81] Recall that one can think of the output distribution reflecting the population distribution over cities with a linear production function for city output in a single factor, labor. Thus according to this story households migrate automatically in a manner consistent with the counterfactual output distribution such that income per capita is equalized across cities. Therefore there is no dispersion in per-capita income in either the observed or counterfactual world, less output dispersion as measured by the variance across cities in the counterfactual, and more inequality in the counterfactual in terms of the maximum-minimum output differential between cities.

links to guarantee that the initial multimodal network is connected; this is why the number of links starts above the number of firms initially. Both networks contract and then moderately grow at roughly the same rate and the gap in the number of links persists until the networks reach stability, with the canals only counterfactual having more links on average than the railroads only counterfactual.

See Figure 43 for counterfactual total firm revenue; in the plot, total firm revenue in the factual transportation network is normalized to one for each type individually. Relative to the factual world, canal firms earn much more in revenue, with railroad firms about as well off collectively in the counterfactual as in the observed world.

8 Conclusion

Motivated by the seminal work of Robert Fogel on U.S. railroads, I reformulate Fogel's original counterfactual history question on 19th century U.S. economic growth without railroads by treating the transport network as an endogenous equilibrium object. I quantify the effect of the railroad on U.S. growth from its introduction in 1830 to 1861. Specifically, I estimate the output loss in a counterfactual world without the technology to build railroads, but retaining the ability to construct the next-best alternative of canals. My main contribution is to endogenize the counterfactual canal network through a decentralized network formation game played by profit-maximizing transport firms. I perform a similar exercise in a world without canals.

I find that railroads and canals are strategic complements, not strategic substitutes. Therefore, the output loss can be quite acute when one or the other is missing from the economy. See Table 8 (Appendix, part B) for the main results. In the set of Single Link Nash Stable or SLNS networks, relative to the 1861 factual world, I find that output is 45% lower in the canals only counterfactual and 49% lower in the railroads only counterfactual on average. I also find that, relative to the 1861 factual world, in the counterfactual with canals only the variance of output decreases by 69%, the number of links built by canal firms increases by 508%, the number of active canal firms increases by 43%, and total canal firm revenue increases by 716%. In the counterfactual with railroads only relative to the observed world, I find that the variance of output also decreases by 69%, but the number of links built by railroad firms decreases by 45%, the number of active railroad firms decreases by 83%, and total railroad firm revenue increases by 7%. With only one of the transportation technologies available, inequality in the U.S. across cities would have been lower in variance terms but sharply higher in terms of the maximum-minimum gap. Such a stark output loss is due to two main mechanisms: inefficiency of the decentralized equilibrium due to network externalities and complementarities due to spatial heterogeneity in costs across the two transport modes. The complementarity between railroads and canals is driven by geography: the presence of surface water in a locality lower costs for canals but raises them for railroads. Strategic pricing of links by firms and endogenous settlement of cities in terms of making their locations a choice variable for households are two possible avenues for future work.

References

Ades, Alberto F. and Edward L. Glaeser, "Evidence on Growth, Increasing Returns, and the Extent of the Market," *Quarterly Journal of Economics*, 1999, *114*, 1025–1045.

Aghion, Philippe and Peter Howitt, "A Model of Growth Through Creative Destruction," *Econometrica*, 1992, *60*, 323–351.

Anderson, James and Eric van Wincoop, "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 2003, *93*, 170–192.

Atack, Jeremy, "On the Use of Geographic Information Systems in Economic History: The American Transportation Revolution Revisited," *Journal of Economic History*, 2013, *73*, 313–338.

_ **and Peter Passell**, *A New Economic View of American History: From Colonial Times to 1940*, W.W. Norton & Company, 1994.

_ **and Robert A. Margo**, "The Impact of Access to Rail Transportation on Agricultural Improvement," *Journal of Transport and Land Use*, 2011, *4*, 5–18.

_ **, Fred Bateman, Michael Haines, and Robert A. Margo**, "Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850-1860," *Social Science History*, 2010, *34*, 171–197.

Aumann, Robert J. and Roger B. Myerson, "Endogenous Formation of Links between Players and of Coalitions: An Application of the Shapley Value," in Alvin E. Roth, ed., *The Shapley Value: Essays in Honor of Lloyd S. Shapley*, Cambridge University Press, 1988.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, "On the Road: Access to Transportation Infrastructure and Economic Growth in China," 2012. Mimeo, NBER Working Paper 17897.

Baum-Snow, Nathaniel, "Did Highways Cause Suburbanization?," *Quarterly Journal of Economics*, 2007, *122*, 775–805.

Becker, Gary S. and Kevin M. Murphy, "The Division of Labor, Coordination Costs, and Knowledge," *Quarterly Journal of Economics*, 1992, *107*, 1137–1160.

Carter, Susan, Scott Gartner, Michael Haines, Alan Olmstead, Richard Sutch, and Gavin Wright, *Historical Statistics of the United States: Millennial Edition*, Cambridge University Press, 2006.

Chaney, Thomas, "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 2008, *98*, 1707–1721.

Cooper, Russell and Andrew John, "Coordinating Coordination Failures in Keynesian Models," *Quarterly Journal of Economics*, 1988, *103*, 441–463.

David, Paul, "Transport Innovation and Economic Growth: Professor Fogel on and off the Rails," *Economic History Review*, 1969, *22*, 506–524.

Dell, Melissa, "Trafficking Networks and the Mexican Drug War," 2012. Mimeo, Harvard.

Donaldson, David, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," 2012. Mimeo, MIT.

_ **and Richard Hornbeck**, "Railroads and American Economic Growth: A "Market Access" Approach," 2013. Mimeo, Harvard.

Duranton, Gilles and Matthew A. Turner, "Urban Growth and Transportation," *Review of Economic Studies*, 2012, *79*, 1407–1440.

_ **, Peter M. Morrow, and Matthew A. Turner**, "Roads and Trade: Evidence from the US." Forthcoming, Review of Economic Studies.

Durlauf, Steven N., "Nonergodic Economic Growth," *Review of Economic Studies*, 1993, *60*, 349–366.

Elliott, Matthew and Benjamin Golub, "A Network Approach to Public Goods," 2013. Mimeo, Caltech.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System." Mimeo, UC-Berkeley.

Fernald, John. G., "Roads to Prosperity? Assessing the Link Between Public Capital and Productivity," *American Economic Review*, 1999, *89*, 619–638.

Feyrer, James, "Distance, Trade, and Income: The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment," 2011. Mimeo, Dartmouth.

Fishlow, Albert, *American Railroads and the Transformation of the Ante-Bellum Economy*, Harvard University Press, 1965.

Fogel, Robert W., *Railroads and American Economic Growth: Essays in Econometric History*, Johns Hopkins Press, 1964.

_ , "Notes on the Social Saving Controversy," *Journal of Economic History*, 1979, *39*, 1–54.

Galeotti, Andrea et al., "Network Games," *Review of Economic Studies*, 2010, *77*, 218–244.

Galor, Oded and David N. Weil, "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond," *American Economic Review*, 2000, *90*, 806–828.

Goodrich, Carter, *Government Promotion of American Canals and Railroads, 1800-1890*, Columbia University Press, 1960.

Gramlich, Edward M., "Infrastructure Investment: A Review Essay," *Journal of Economic Literature*, 1994, *32*, 1176–1196.

Grubler, Arnulf, *The Rise and Fall of Infrastructures: Dynamics of Evolution and Technological Change in Transport*, Physica-Verlag, 1990.

Holmes, Thomas, "The Diffusion of Wal-Mart and Economies of Density," *Econometrica*, 2011, *79*, 253–302.

Jackson, Matthew O., *Social and Economic Networks*, Princeton University Press, 2008.

_ , *An Overview of Social Networks and Economic Applications* Handbook of Social Economics, North Holland,

_ **and Asher Wolinsky**, "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 1996, *71*, 44–74.

Jenks, Leland H., "Railroads as an Economic Force in American Development," *Journal of Economic History*, 1944, *4*, 1–20.

Kelly, Morgan, "The Dynamics of Smithian Growth," *Quarterly Journal of Economics*, 1997, *112*, 939–964.

_ , "Linkages, Thresholds, and Development," *Journal of Economic Growth*, 2001, *6*, 39–53.

Kremer, Michael, "The O-Ring Theory of Economic Development," *Quarterly Journal of Economics*, 1993, *108*, 551–575.

Krugman, Paul, "Increasing Returns and Economic Geography," *Journal of Political Economy*, 1991, *99*, 483–499.

_ , "What's New about the New Economic Geography?," *Oxford Review of Economic Policy*, 1998, *14*, 7–17.

Lebergott, Stanley, "United States Transport Advance and Externalities," *Journal of Economic History*, 1966, *26*, 437–461.

Lucas, Robert, "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 1988, *22*, 3–42.

_ , "Making a Miracle," *Econometrica*, 1993, *61*, 251–272.

Matsuyama, Kiminori, "Increasing Returns, Industrialization, and Indeterminacy of Equilibrium," *Quarterly Journal of Economics*, 1991, *106*, 617–650.

McClelland, Peter D., "Railroads, American Growth, and the New Economic History: A Critique," *Journal of Economic History*, 1968, *28*, 102–123.

_ , "Social Rates of Return on American Railroads in the Nineteenth Century," *Economic History Review*, 1972, *25*, 471–488.

Murphy, Kevin, Andrei Shleifer, and Robert Vishny, "Industrialization and the Big Push," *Journal of Political Economy*, 1989, *97*, 1003–1026.

Myerson, Roger B., "Graphs and Cooperation in Games," *Mathematics of Operations Research*, 1977, *2*, 225–229.

Nerlove, Marc, "Railroads and American Economic Growth," *Journal of Economic History*, 1966, *26*, 107–115.

Ngai, L. Rachel, "Barriers and the Transition to Modern Growth," *Journal of Monetary Economics*, 2004, *51*, 1353–1383.

North, Douglass C., *The Economic Growth of the United States: 1790-1860*, W. W. Norton & Company, 1966.

Perez-Cervantes, Fernando, "Railroads and Economic Growth: A Trade Policy Approach," 2012. Mimeo, University of Chicago.

Poor, Henry V., *Railroads and Canals of the United States of America*, John H. Schultz & Co., 1860.

Redding, Stephen J. and Daniel M. Sturm, "The Costs of Remoteness: Evidence from German Division and Reunification," *American Economic Review*, 2008, *98*, 1766–1797.

Romer, Paul M., "Endogenous Technological Change," *Journal of Political Economy*, 1990, *98*, S71–S102.

Rostow, W. W., "The Take-Off Into Self-Sustained Growth," *Economic Journal*, 1956, *66*, 25–48.

Segerstrom, Paul S., "Endogenous Growth without Scale Effects," *American Economic Review*, 1998, *88*, 1290–1310.

Sheng, Shuyang, "Identification and Estimation of Network Formation Games," 2012. Mimeo, University of Southern California.

Shy, Oz, "A Short Survey of Network Economics," *Review of Industrial Organization*, 2011, *38*, 119–149.

Silva, J.M.C. Santos and Silvana Tenreyro, "The Log of Gravity," *Review of Economics and Statistics*, 2006, *88*, 641–658.

Stigler, George J., "The Division of Labor is Limited by the Extent of the Market," *Journal of Political Economy*, 1951, *59*, 185–193.

Stokey, Nancy, "Learning by Doing and the Introduction of New Goods," *Journal of Political Economy*, 1988, *96*, 701–717.

Taylor, George R. and Irene D. Neu, *The American Railroad Network: 1861-1890*, Harvard University Press, 1956.

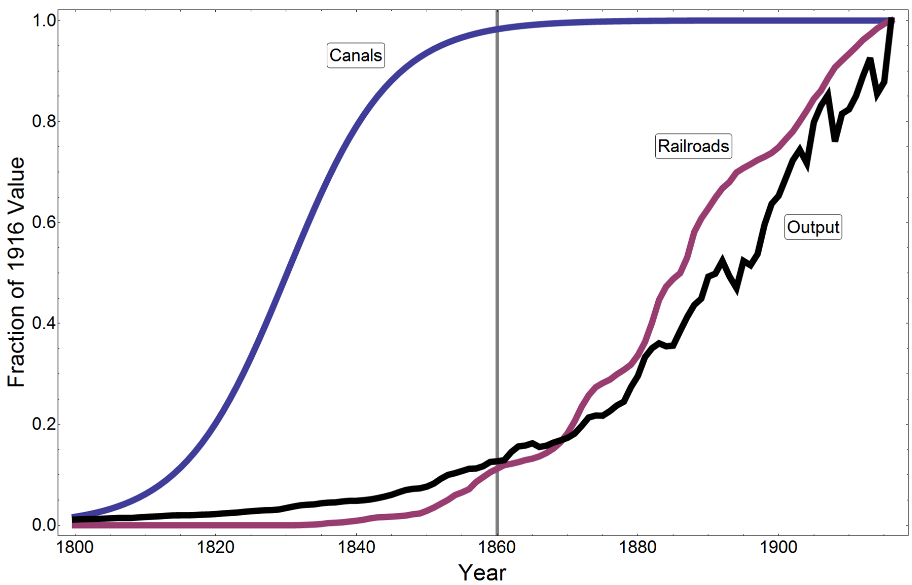
Williamson, Jeffery G., *Late Nineteenth Century American Development: A General Equilibrium History*, Cambridge University Press, 1974.

Yang, Xiaokai and Jeff Borland, "A Microeconomic Mechanism for Economic Growth," *Journal of Political Economy*, 1991, *99*, 460–482.

Appendix

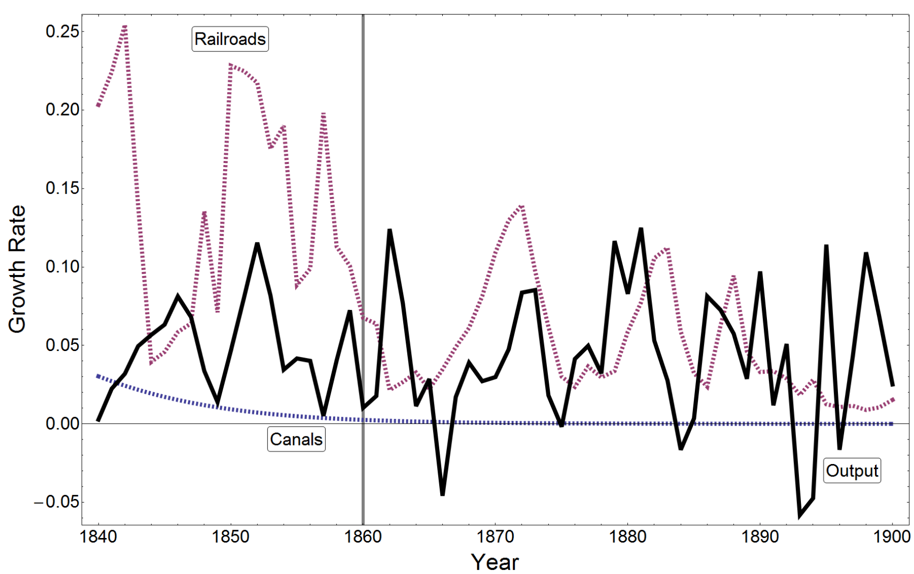
A Figures

Figure 1: Transportation and Output: Canals and Railroads



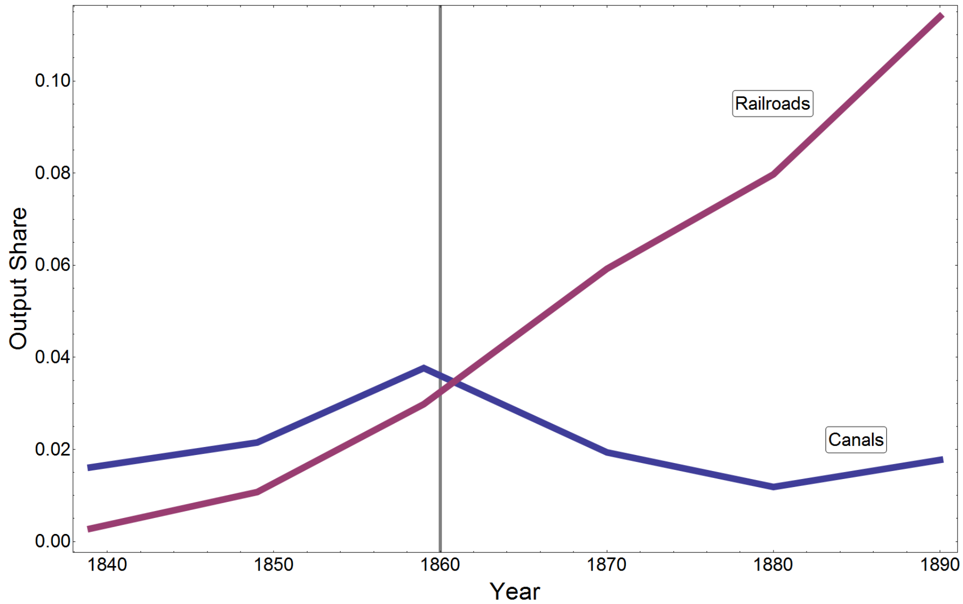
Note: Source is Grubler (1990) and Historical Statistics of the United States: Millennial Edition. Series value in 1916 normalized to one. "Canals" is total canal mileage, "Railroads" is total railroad mileage.

Figure 2: Transportation and Output Growth: Canals and Railroads



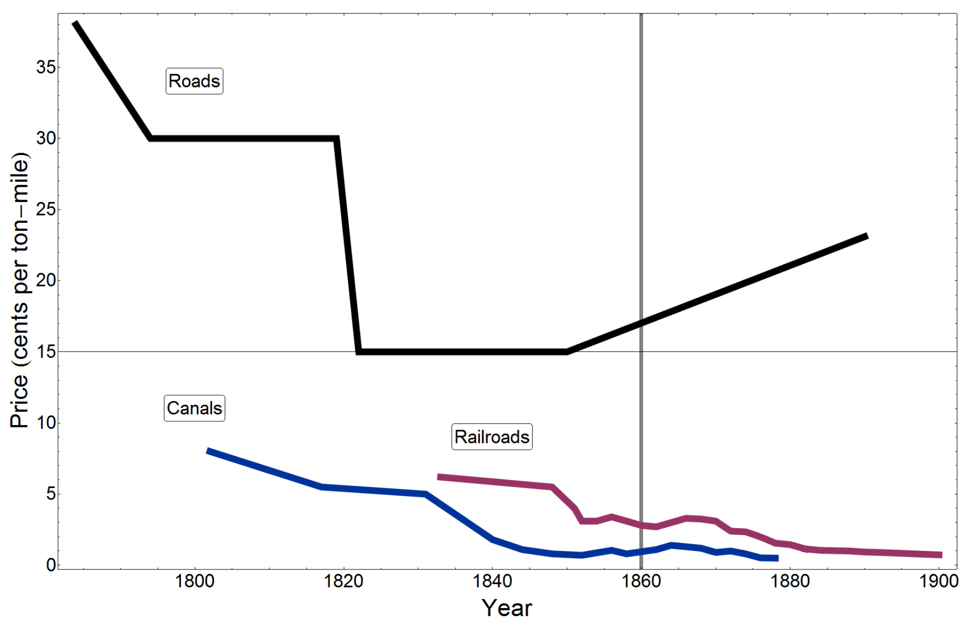
Note: Source is Grubler (1990) and Historical Statistics of the United States: Millennial Edition. Growth rate 0.1 is 10%. "Canals" is total canal mileage, "Railroads" is total railroad mileage.

Figure 3: Output Share: Canals and Railroads



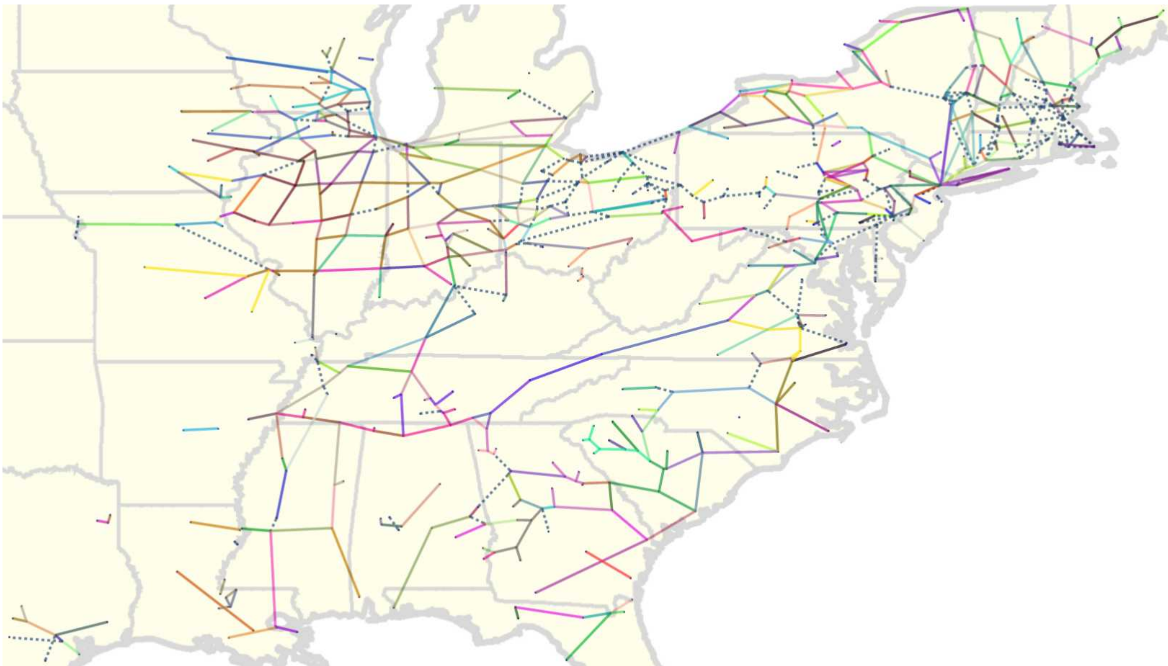
Note: Source is Fishlow (1965, 1966), Davis and Easterlin (1972). Output share 0.1 is 10% of output in that year. "Canals" is canal industry and "Railroads" is railroad industry.

Figure 4: Prices: Roads, Canals, and Railroads



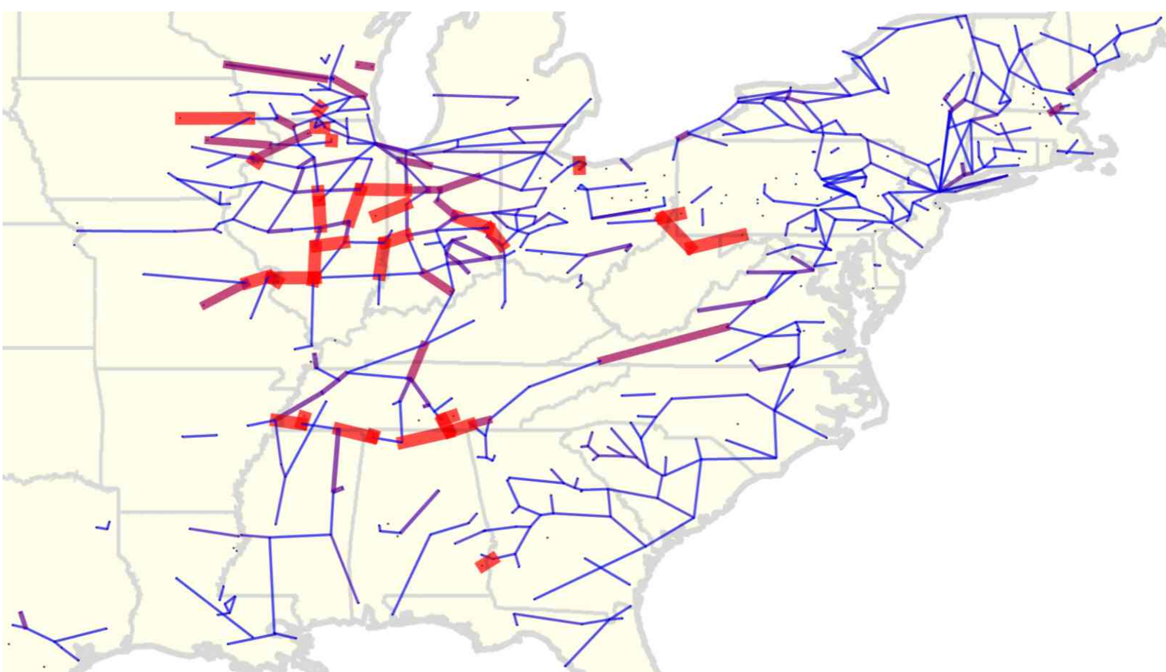
Note: Source is North (1965). Prices (freight rates) quoted in cents per ton-mile shipped. "Canals" is average canal rate and "Railroads" is average railroad rate; same for "Roads" with wagons.

Figure 5: 1861 U.S. Railroad Network, Firm Distribution



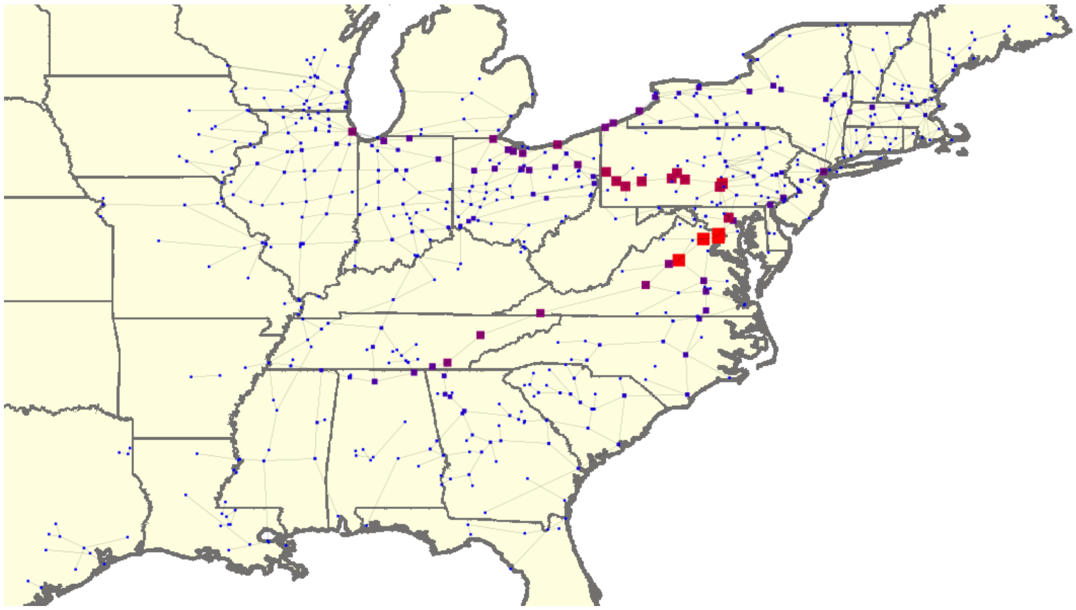
Note: Figure shows network of links due to constructed railroads. Railroads from Taylor and Neu (1956). Cities denoted by dots on map. Each individual railroad firm's subnetwork is shaded with a different random color. Dashed blue links cannot be attributed to a specific firm using information from Taylor and Neu (1956) alone (thus these links are not attached to any particular firm's subgraph for the empirical work, but they do serve to reduce transport costs).

Figure 6: 1861 U.S. Railroad Firms, Theoretical Revenue



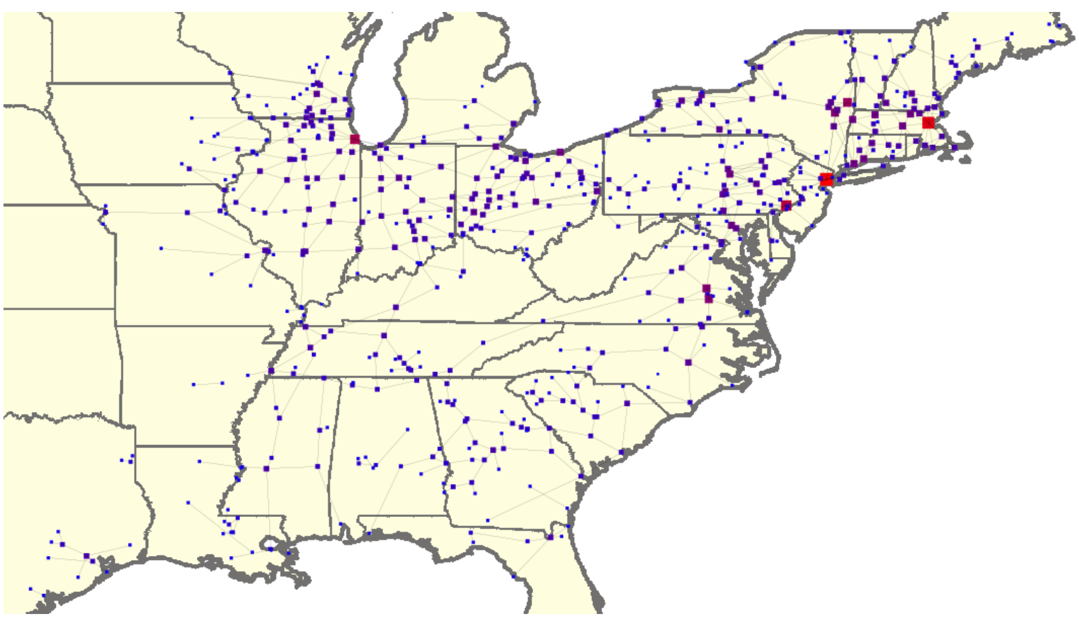
Note: Figure shows network of links due to constructed railroads. Railroads from Taylor and Neu (1956). Cities denoted by dots on map. Links shaded according to relative revenues, thicker red (thinner blue) is higher (lower) revenue. Revenue is derived from gravity trade model, thus theoretical.

Figure 7: 1861 U.S. Railroad Network, Weighted by Betweenness Centrality



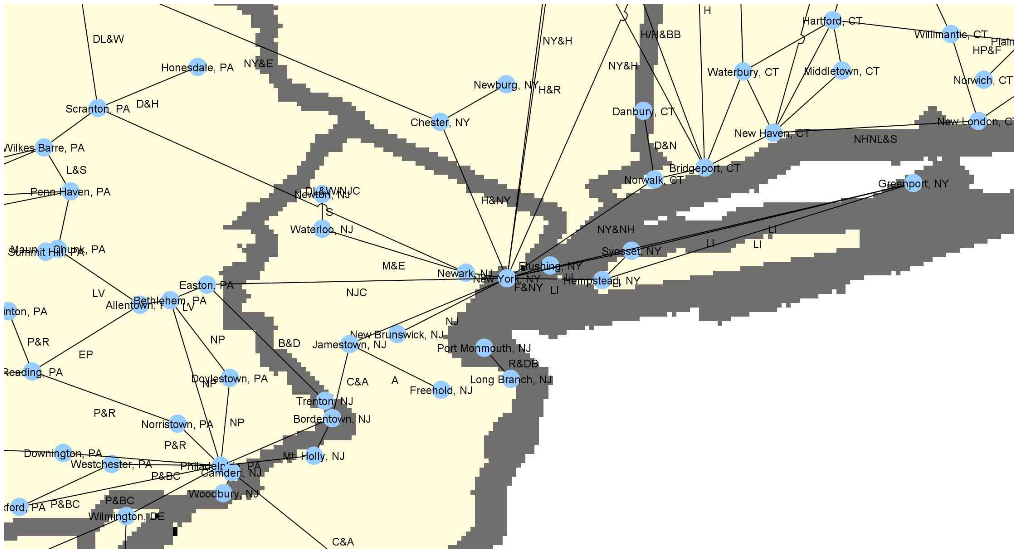
Note: Figure shows network of links due to constructed railroads. Railroads from Taylor and Neu (1956). Cities denoted by dots on map. Cities weighted by betweenness centrality, from small blue (low centrality) to large red (high). Betweenness centrality defined as fraction of all bilateral shortest paths between cities that pass through any particular city. Note that the most central cities are on "trunk lines" that pass between the three regions: Northeast, South, and Midwest. The major cities (New York, Boston, Chicago) are not particularly central according to this measure.

Figure 8: 1861 U.S. Railroad Network, Weighted by Degree



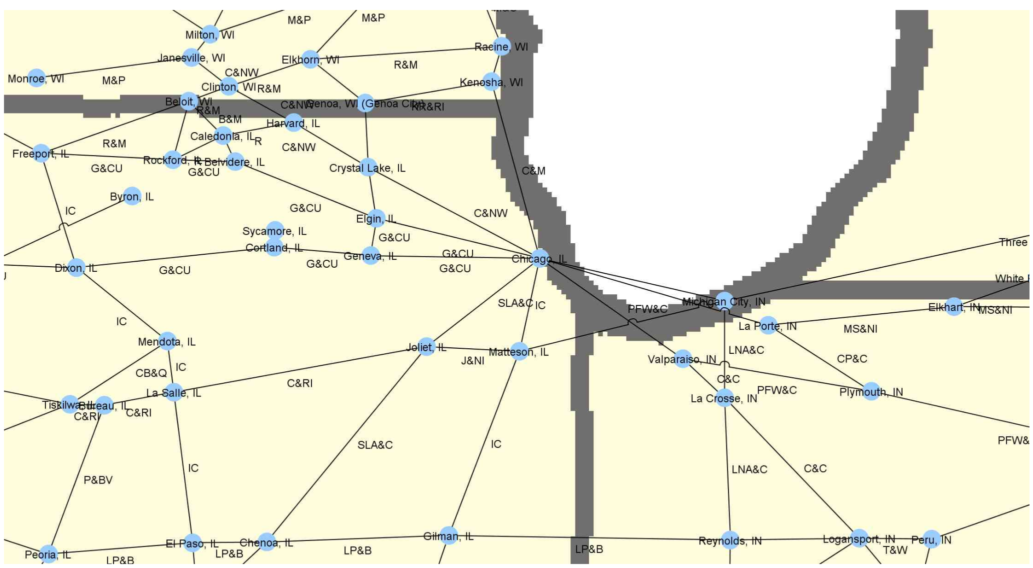
Note: Figure shows network of links due to constructed railroads. Railroads from Taylor and Neu (1956). Cities denoted by dots on map. Cities weighted by degree, from small blue (low degree) to large red (high). Degree defined as number of links attached to any particular city. The major cities (New York, Boston, Chicago) have highest degree since they are hubs of local hub and spoke networks.

Figure 9: 1861 U.S. Railroad Network, New York City Hub



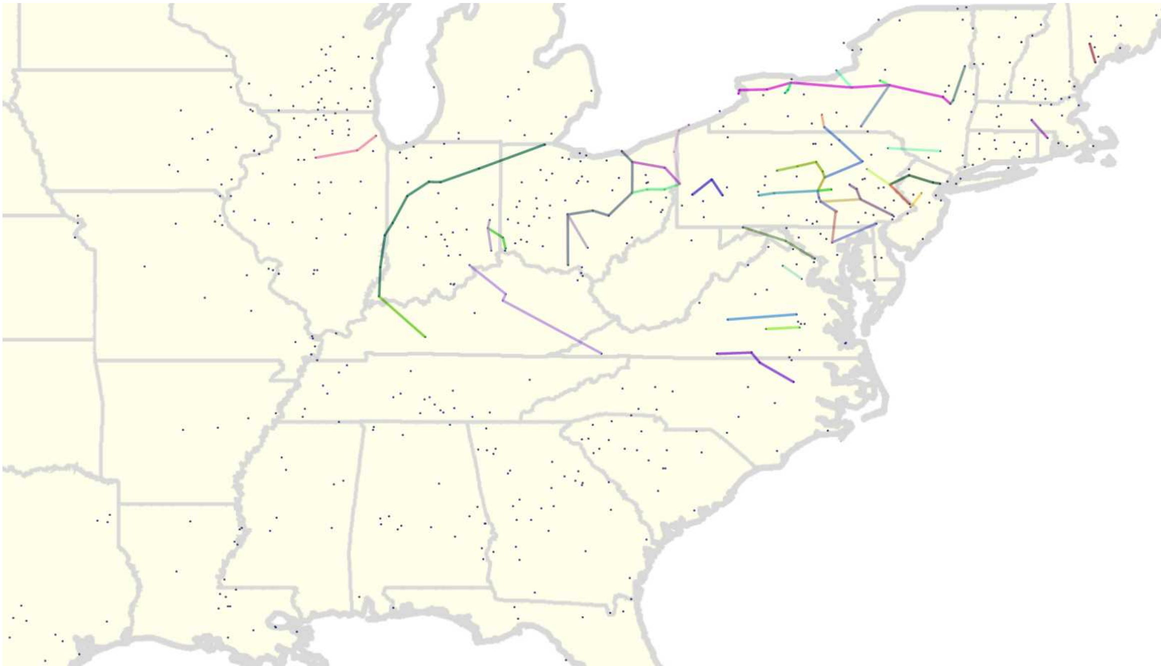
Note: Figure shows New York City hub of observed 1861 U.S. railroad network. Railroads from Taylor and Neu (1956). Cities denoted by dots on map, labeled by name and state (ex. Scranton, PA). Links are railroad connections, labeled by firm ownership (ex. NY&H is the New York & Harlem Railroad). Strongly resembles local hub and spoke system. Such connectivity rationalized by high population and geographic location of New York City (Atlantic Ocean, Hudson River).

Figure 10: 1861 U.S. Railroad Network, Chicago Hub



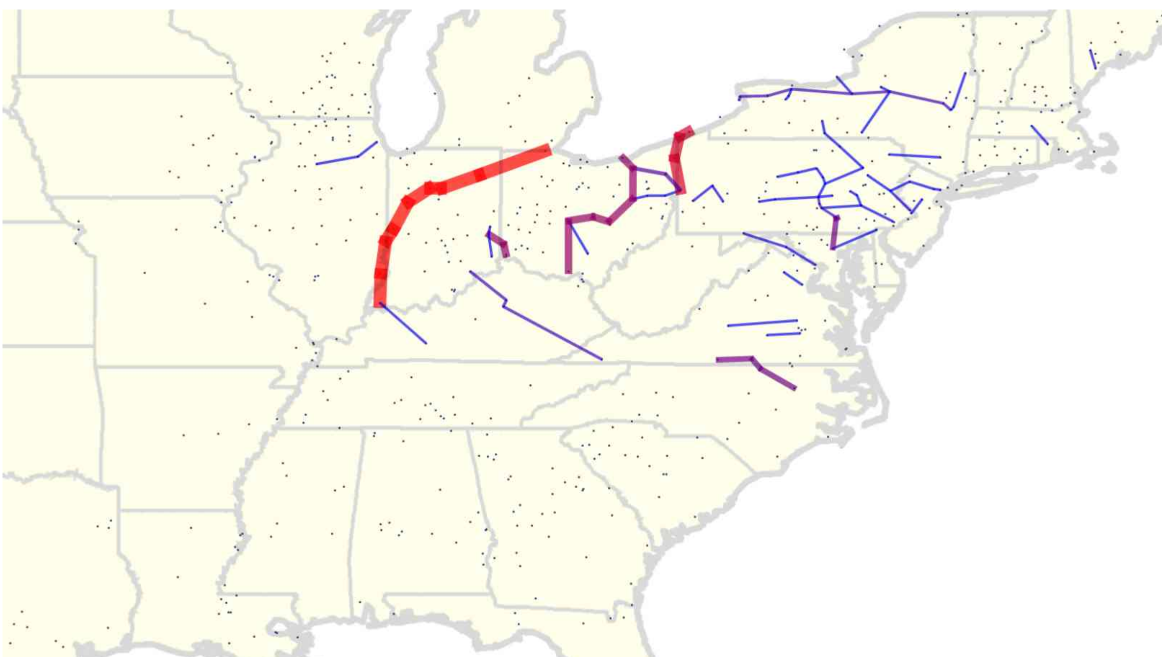
Note: Figure shows Chicago hub of observed 1861 U.S. railroad network. Railroads from Taylor and Neu (1956). Cities denoted by dots on map, labeled by name and state (ex. Elgin, IL). Links are railroad connections, labeled by firm ownership (ex. IC is the Illinois Central Railroad). Strongly resembles local hub and spoke system. Such connectivity rationalized by high population and geographic location of Chicago (Lake Michigan, Chicago River).

Figure 11: 1861 U.S. Canal Network, Firm Distribution



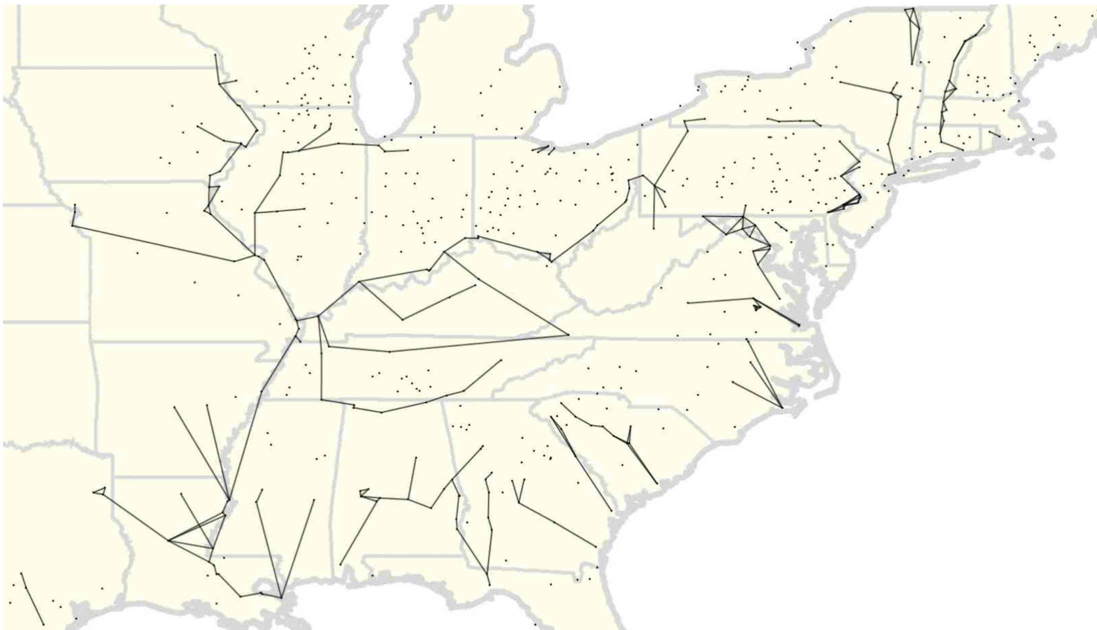
Note: Figure shows network of links due to constructed canals. List of canals from table "List of U.S. Canals (operational in 1861)" in Appendix. Cities denoted by dots on map. Each individual canal firm's subnetwork is shaded with a different random color.

Figure 12: 1861 U.S. Canal Firms, Theoretical Revenue



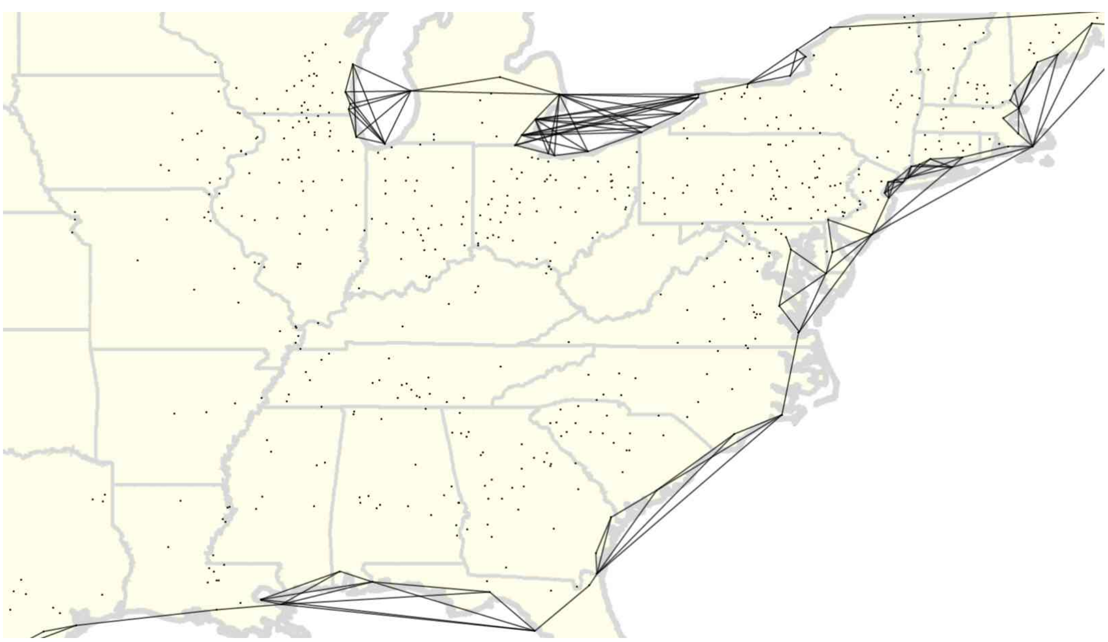
Note: Figure shows network of links due to constructed canals. List of canals from table "List of U.S. Canals (operational in 1861)" in Appendix. Cities denoted by dots on map. Links shaded according to relative revenues, thicker red (thinner blue) is higher (lower) revenue. Revenue is derived from gravity trade model, thus theoretical.

Figure 13: U.S. Waterway Network



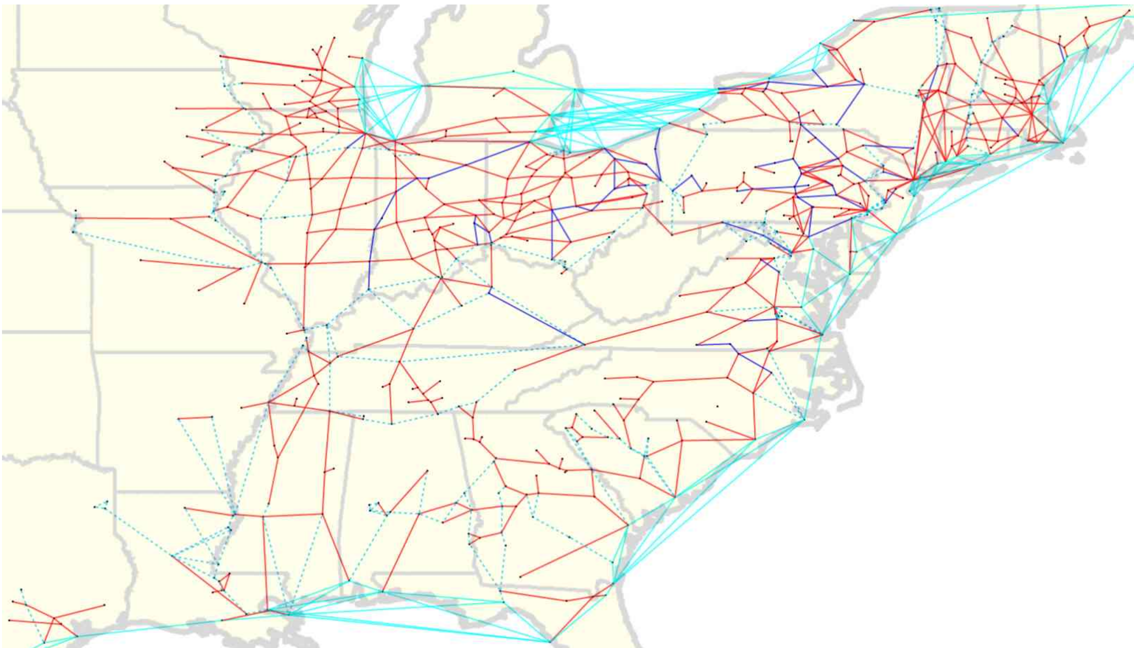
Note: Figure shows network of links due to natural waterways. List of waterways from table "List of Selected Major U.S. Waterways" in Appendix. Cities denoted by dots on map.

Figure 14: U.S. Great Lakes and Coastal Network



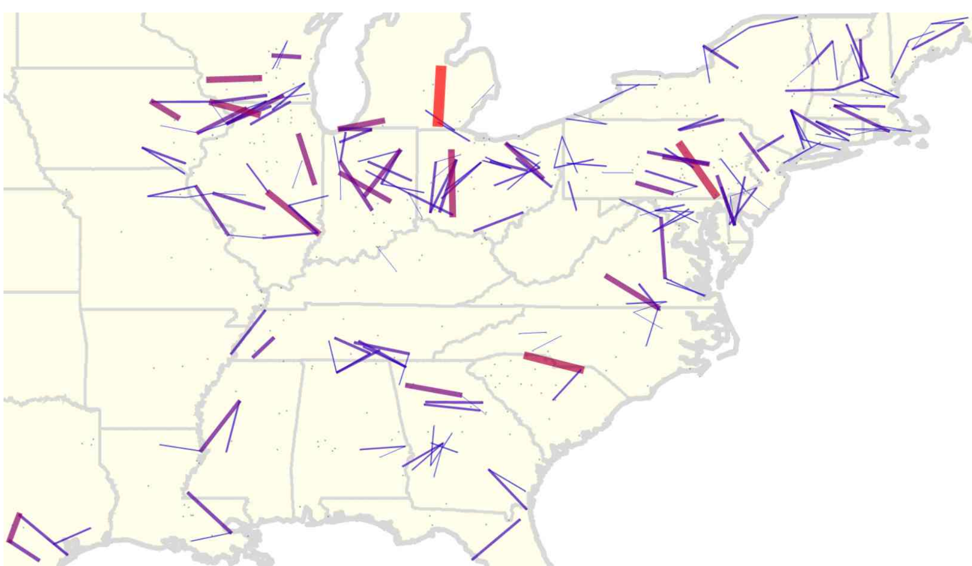
Note: Figure shows network of links due to coastal shipping and Great Lakes. Artificial network, proxy for shipping opportunities. Cities denoted by dots on map. Can go from Great Lakes to Atlantic coast due to presence of St. Lawrence Seaway (assumed to transfer over to counterfactual as part of coastal shipping links).

Figure 15: U.S. Combined Multimodal Transport Network



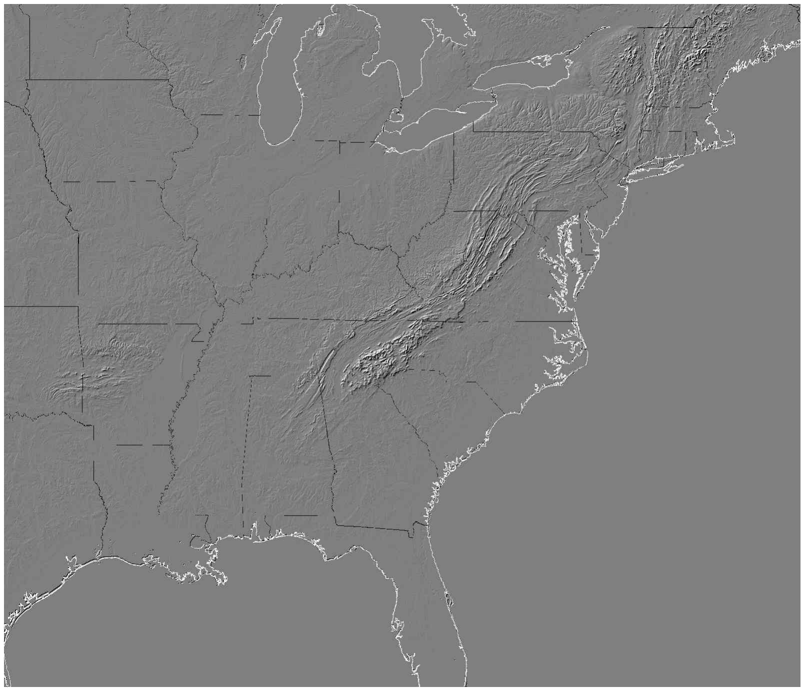
Note: Figure shows combined multimodal transport network as union of all transport links available in observed network. Notation: red links are railroads, solid dark blue links are canals, dashed light blue links are natural waterways, and solid teal links represent coastal shipping (and Great Lakes). Cities denoted by dots on map. Can go from Great Lakes to Atlantic coast due to presence of St. Lawrence Seaway (assumed to transfer over to counterfactual).

Figure 16: U.S. Bilateral Transport Costs



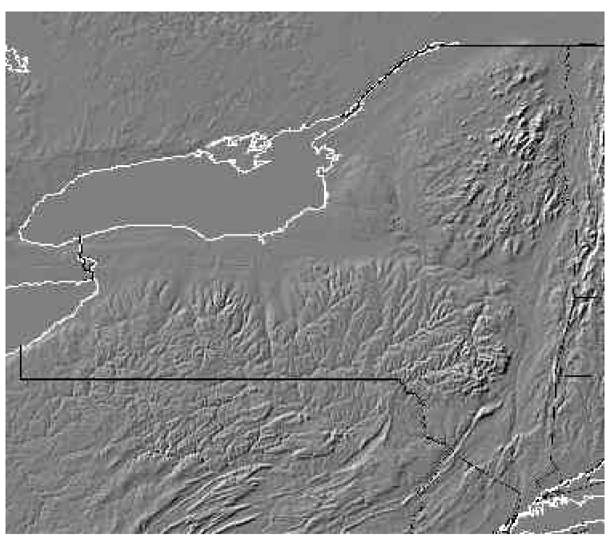
Note: Figure plots bilateral transport cost (or trade cost) for a set of randomly selected pairs of cities. Transport cost is denoted by c_{ij} in the paper and is interpreted as the bilateral cost of moving goods from city i to city j in terms of the fraction of the value of the goods lost in transport (iceberg trade cost). Cities denoted by dots on map. Links shaded according to relative transport cost, thicker red (thinner blue) is higher (lower) cost. Transport cost is highest in the Midwest (Wisconsin, Illinois, Michigan) but relatively uniform otherwise since all cities are tied into the transport network by either railroad or canal in the data.

Figure 17: U.S. Terrain Elevation Map, Overview



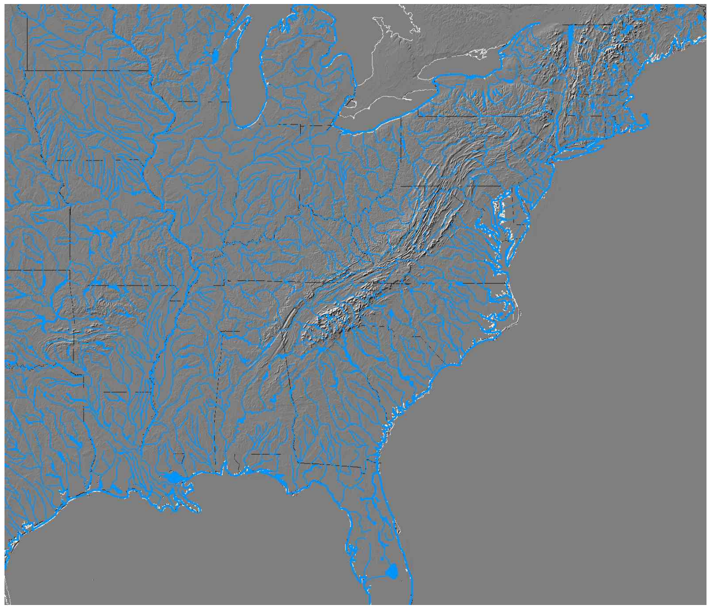
Note: Figure shows terrain map of U.S. shaded by elevation. Change in color gradient of map denotes hills, mountains according to elevation. Coastal areas outlined in white, for illustrative purposes. State boundaries denoted by black lines, also for illustration.

Figure 18: U.S. Terrain Elevation Map, New York Detail



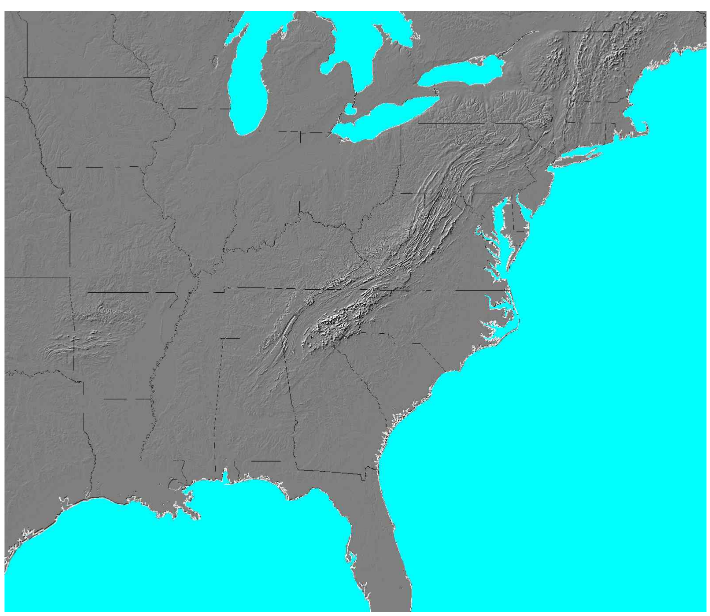
Note: Figure shows terrain map of U.S. shaded by elevation. Coastal areas outlined in white, for illustrative purposes. State boundaries denoted by black lines, also for illustration. Detail map of state of New York. Note relatively uniform (in terms of terrain gradient) path taken by Erie Canal in Mohawk Valley (East-West channel, middle of state).

Figure 19: U.S. Minor Waterways



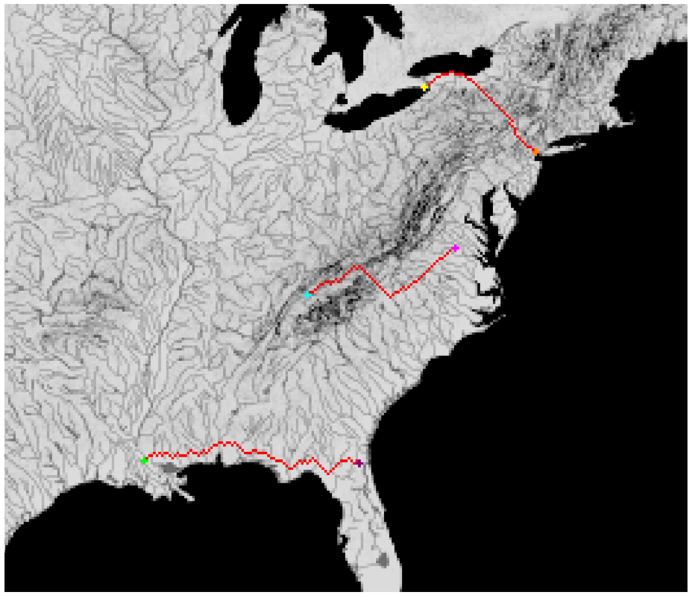
Note: Figure shows terrain map of U.S. with state boundaries. Areas denoting natural minor waterways (smaller lakes, rivers, streams, etc.) shaded in light blue. These waterways cannot be used for transport directly since they are not immediately navigable, but rather make canals less expensive to build and railroads more expensive.

Figure 20: U.S. Great Lakes and Atlantic Coast



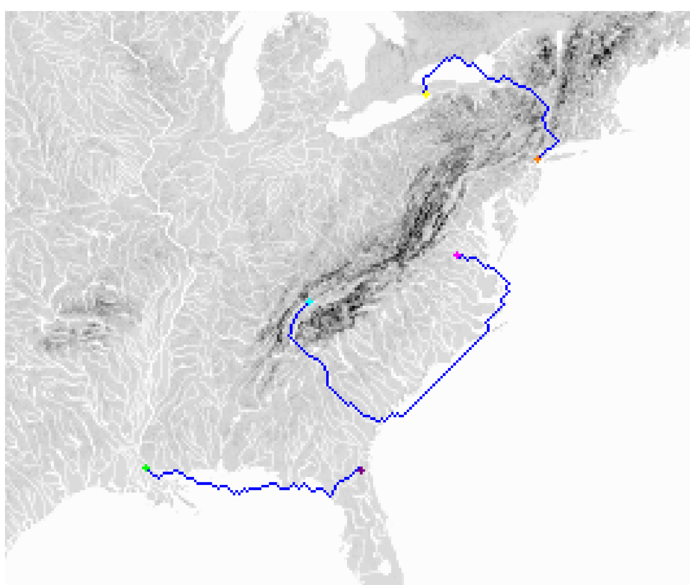
Note: Figure shows terrain map of U.S. with state boundaries. Areas denoted coastal, including Great Lakes, shaded in teal. Can go from Great Lakes to Atlantic coast due to presence of St. Lawrence Seaway (assumed to transfer over to counterfactual as part of coastal shipping links), thus coastal system is connected (from Great Lakes to Atlantic Ocean).

Figure 21: Selected Least-cost Railroad Links



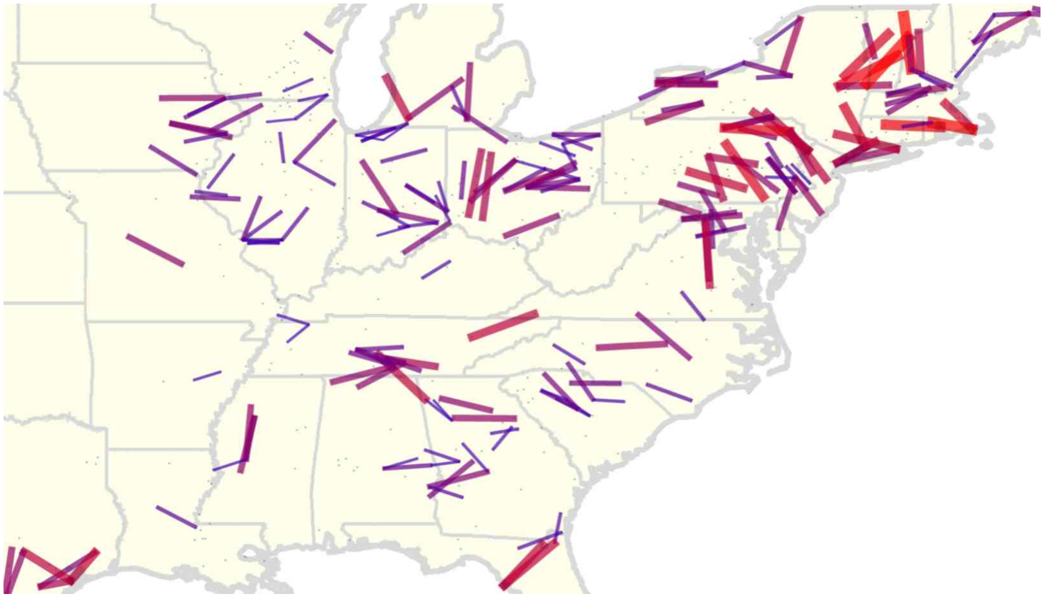
Note: Figure shows a selection of least-cost railroad link paths in terrain: three candidate links (from top to bottom) between New York City-Buffalo, Richmond-Knoxville, and Jacksonville-Baton Rouge. Shading of cells indicates high (dark) or low (light) relative cost penalty for railroad construction, thus coastal areas are most costly for railroads.

Figure 22: Selected Least-cost Canal Links



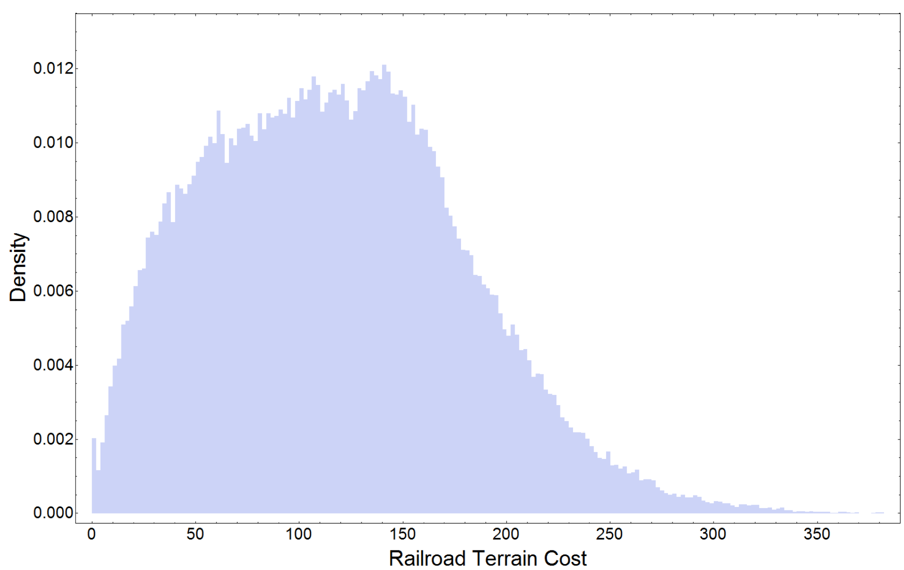
Note: Figure shows a selection of least-cost canal link paths in terrain: three candidate links (from top to bottom) between New York City-Buffalo, Richmond-Knoxville, and Jacksonville-Baton Rouge. Shading of cells indicates high (dark) or low (light) relative cost penalty for canal construction, thus coastal areas are least costly for canals.

Figure 23: Railroad Terrain Cost



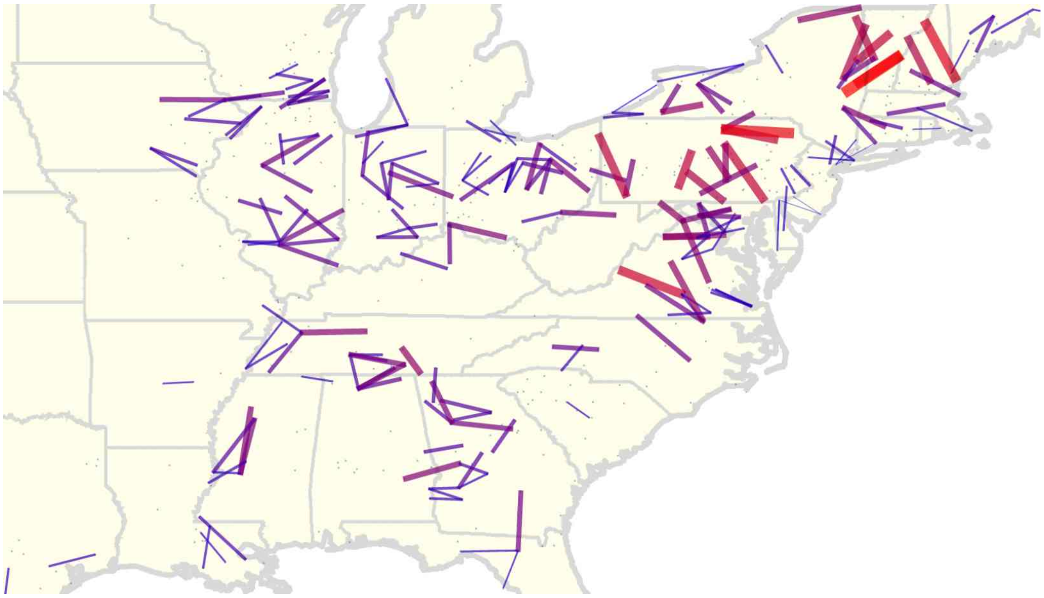
Note: Figure plots bilateral railroad terrain cost for a set of randomly selected pairs of cities. Railroad terrain cost is denoted as $\tau_{RR}(i,j)$ in the paper and is interpreted as the bilateral cost of building a link from city i to city j when multiplied by c_{RR} (type cost for railroads). Cities denoted by dots on map. Links shaded according to relative terrain cost, thicker red (thinner blue) is higher (lower) cost. Note that cost is not uniform with sharp regional variation. Highest penalty found in Northeast due to rough terrain.

Figure 24: Empirical Distribution of Railroad Terrain Cost



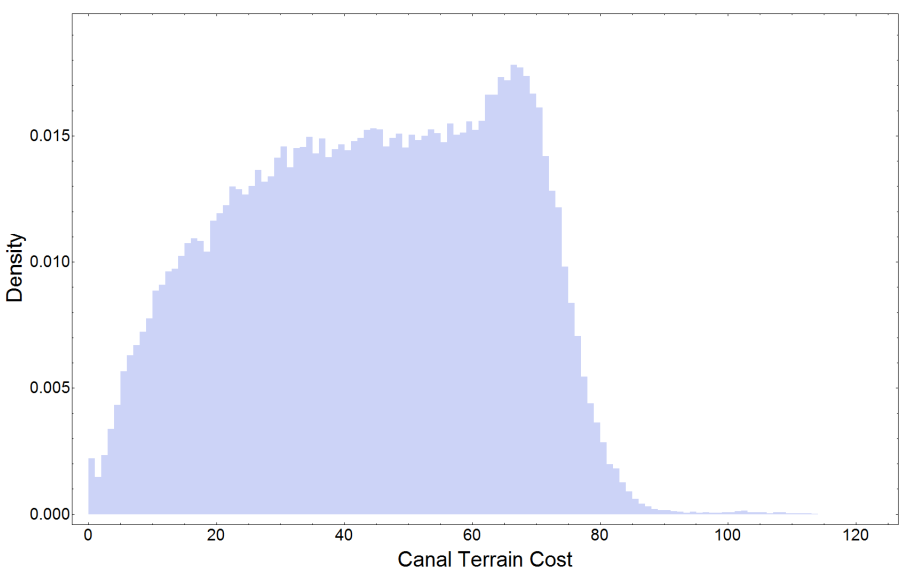
Note: Distribution above is for all possible railroad links that could have been built between pairs of cities. Railroad terrain cost is denoted as $\tau_{RR}(i,j)$ in the paper and is interpreted as the bilateral cost of building a link from city i to city j when multiplied by c_{RR} (type cost for railroads). Note that distribution is peaked at moderate cost with large right tail.

Figure 25: Canal Terrain Cost



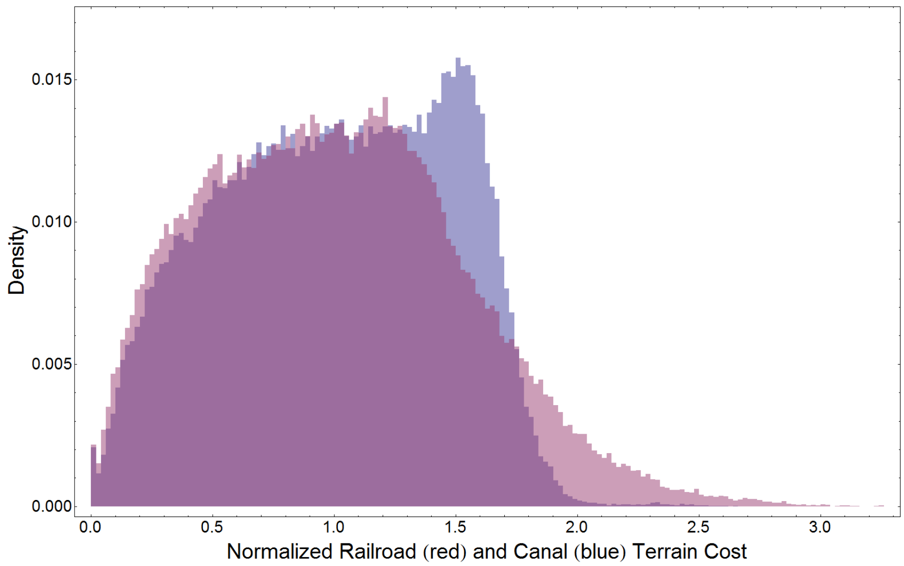
Note: Figure plots bilateral canal terrain cost for a set of randomly selected pairs of cities. Canal terrain cost is denoted as $\tau_{CN}(i, j)$ in the paper and is interpreted as the bilateral cost of building a link from city i to city j when multiplied by c_{CN} (type cost for canals). Cities denoted by dots on map. Links shaded according to relative terrain cost, thicker red (thinner blue) is higher (lower) cost. Note that cost is relatively uniform in South and Midwest. Highest cost for Northeastern canals in rough terrain.

Figure 26: Empirical Distribution of Canal Terrain Cost



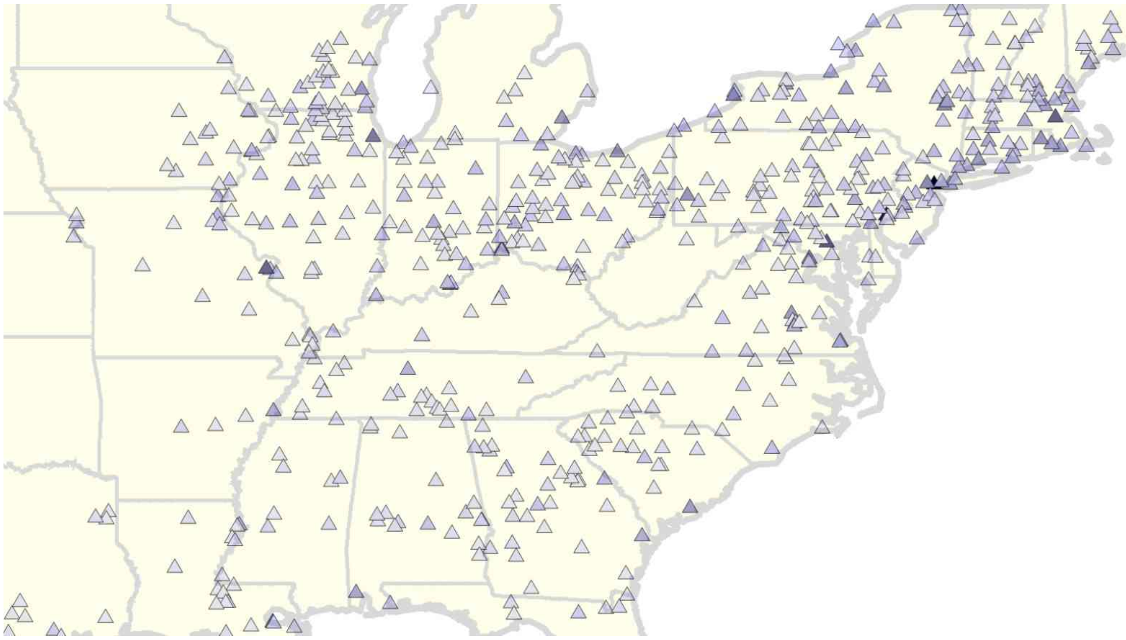
Note: Distribution above is for all possible canal links that could have been built between pairs of cities. Canal terrain cost is denoted as $\tau_{CN}(i, j)$ in the paper and is interpreted as the bilateral cost of building a link from city i to city j when multiplied by c_{CN} (type cost for canals). Note that distribution is relatively uniform with a small right tail.

Figure 27: Empirical Distribution of Normalized Terrain Cost



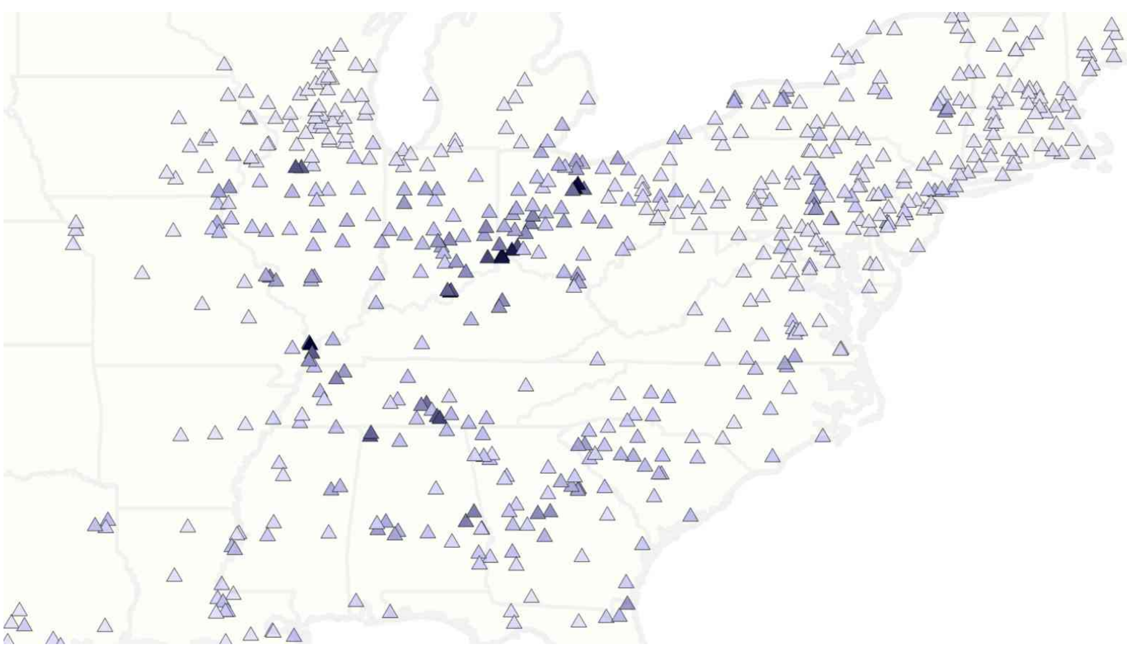
Note: Distributions normalized such that mean set equal to one for each distribution individually. Railroad marked in red, canal marked in blue. Railroad terrain cost is denoted as $\tau_{RR}(i,j)$ in the paper and is interpreted as the bilateral cost of building a link from city i to city j when multiplied by c_{RR} (type cost for railroads). Canal terrain cost is denoted as $\tau_{CN}(i,j)$ in the paper and is interpreted as the bilateral cost of building a link from city i to city j when multiplied by c_{CN} (type cost for canals). Purpose of normalization is to note that canal penalty is relatively more uniform while railroads are more skewed towards relatively more expensive links; railroad terrain cost distribution has a larger right tail.

Figure 28: Observed 1861 U.S. Output Distribution



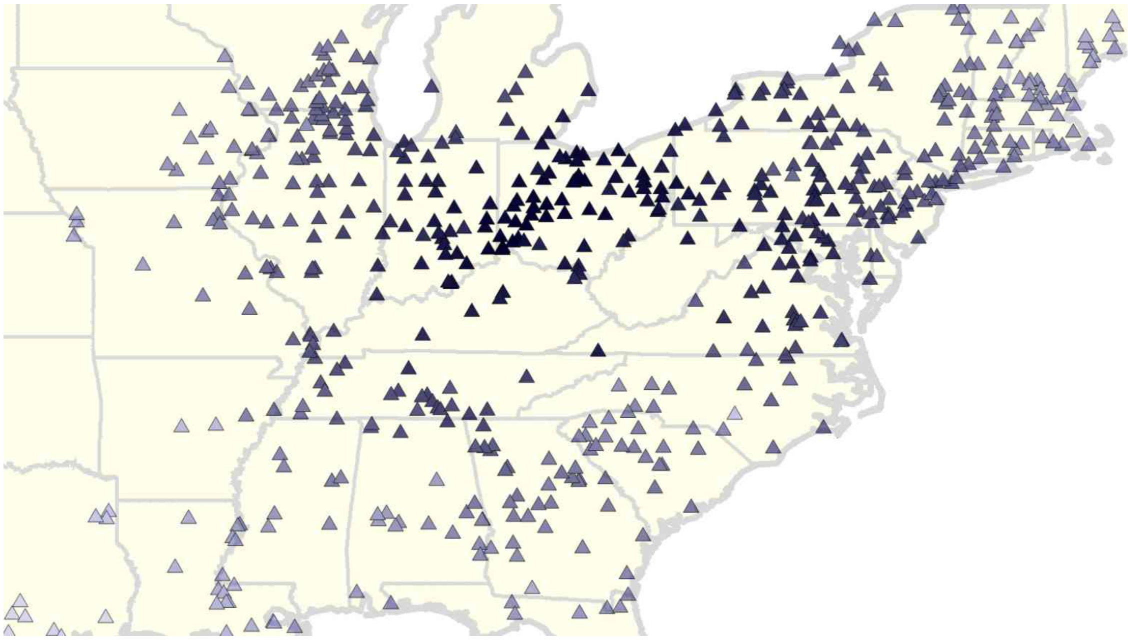
Note: Data from 1860 U.S. census on population distribution across cities (triangles on map). Darker shading indicates relatively more output at a city, not level differences across figures.

Figure 29: Exact Theoretical Output Distribution



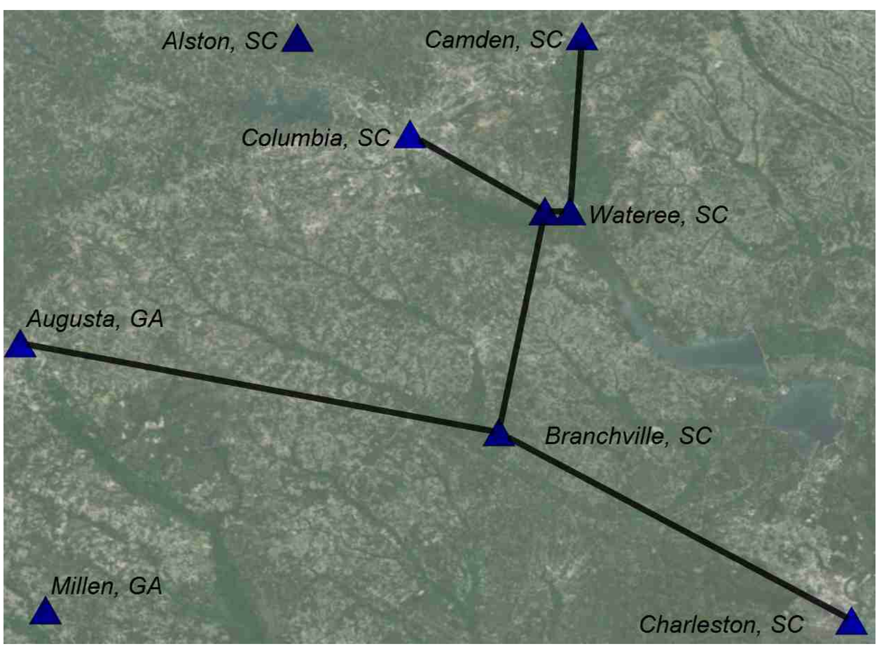
Note: Output distribution from exact solution to gravity trade model. Cities are triangles on map. Darker shading indicates relatively more output at a city, not level differences across figures.

Figure 30: Approximate Theoretical Output Distribution



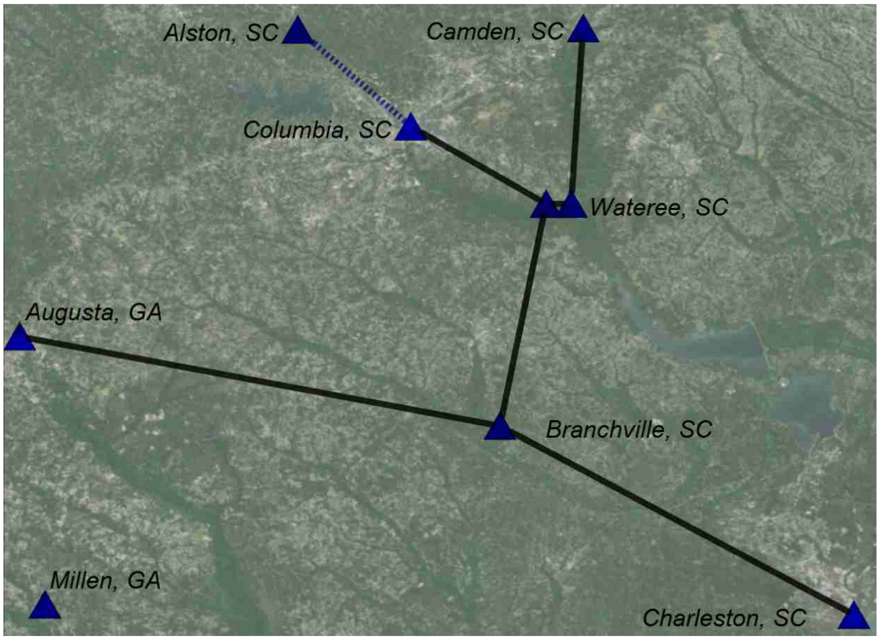
Note: Output distribution from approximate solution to gravity trade model. Cities are triangles on map. Approximation replaces distribution of bilateral distances with average bilateral distance from a particular city to all others in gravity model. Darker shading indicates relatively more output at a city, not level differences across figures.

Figure 31: Observed 1861 U.S. Railroad Network, South Carolina Railroad



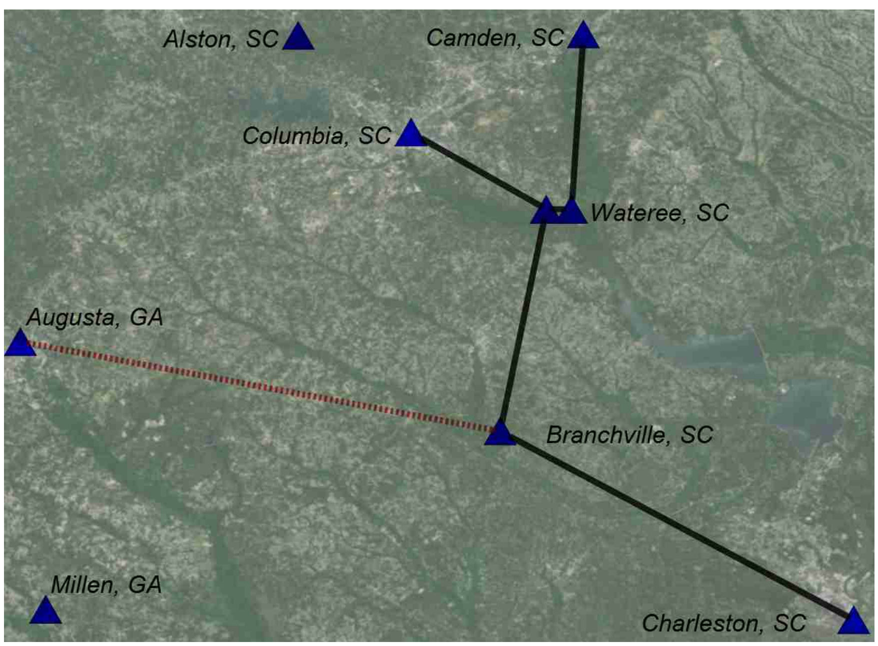
Note: Figure shows observed set of links for South Carolina Railroad in 1861. Cities (triangles on map) labeled with city names. Note that firm's subnetwork is disjoint (relative to other firms) and connected.

Figure 32: Type Cost Bounds, Unilateral Single Link Addition



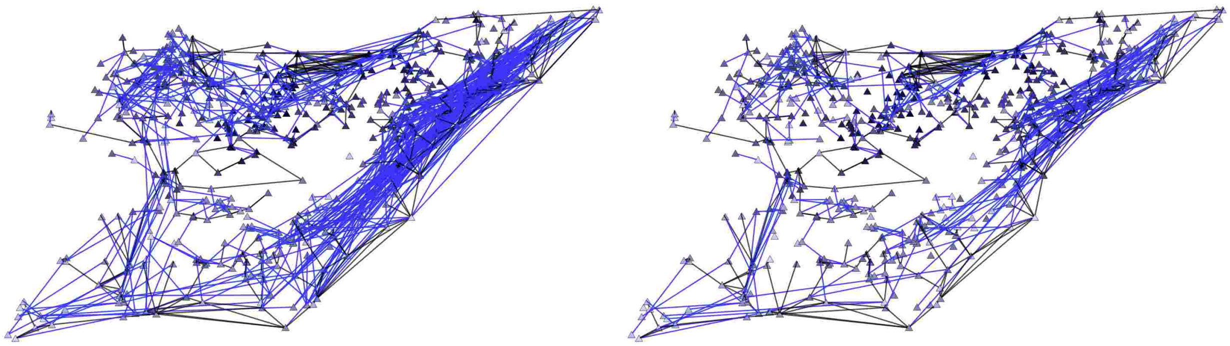
Note: Figure shows observed set of links for South Carolina Railroad in 1861 (solid lines) and one new proposed link (blue dashed line) between Alston and Columbia. Cities (triangles on map) labeled with city names.

Figure 33: Type Cost Bounds, Unilateral Single Link Removal



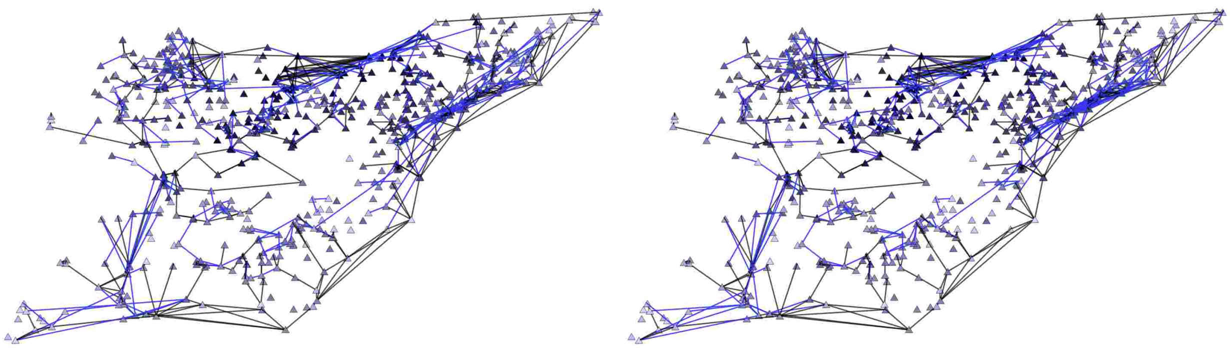
Note: Figure shows observed set of links for South Carolina Railroad in 1861 (solid lines) with one link proposed for removal (red dashed line) between Augusta and Branchville. Cities (triangles on map) labeled with city names.

Figure 34: Counterfactual Canal Network Example: 0 and 500 Iterations



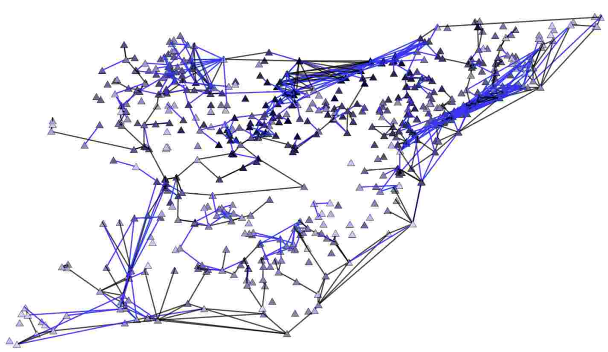
Note: Figure shows one possible result of stable network search algorithm after 0 (left, initial condition) and 500 (right) iterations. Blue links are canals. Cities (triangles on map) shaded according to relative output in final output distribution. Canals only (no railroads) counterfactual.

Figure 35: Counterfactual Canal Network Example: 1000 and 1500 Iterations



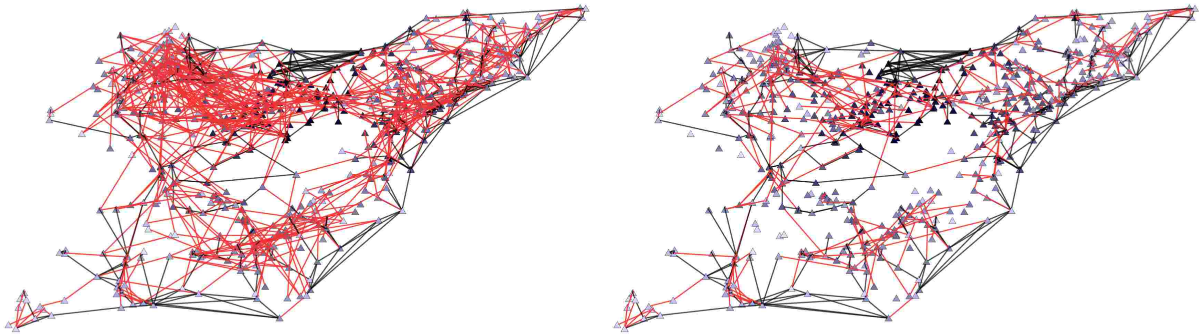
Note: Figure shows one possible result of stable network search algorithm after 1000 (left) and 1500 (right) iterations. Blue links are canals. Cities (triangles on map) shaded according to relative output in final output distribution. Canals only (no railroads) counterfactual.

Figure 36: Counterfactual Canal Network Example: 2000 Iterations



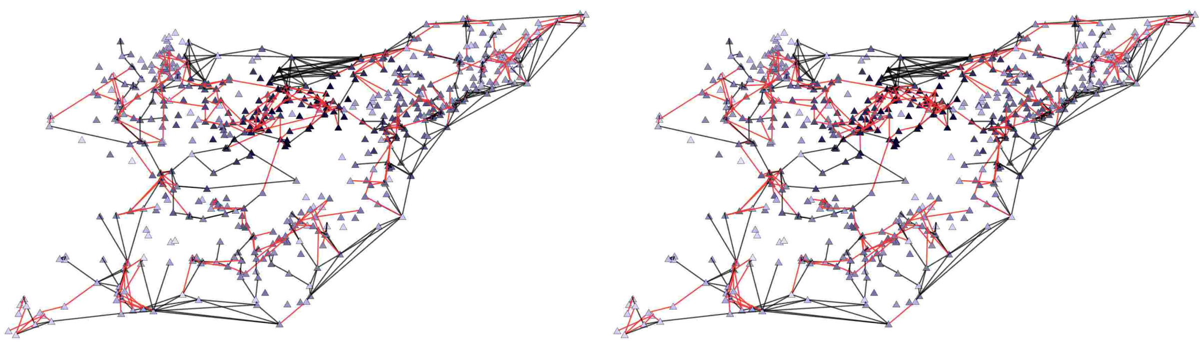
Note: Figure shows one possible result of stable network search algorithm after 2000 iterations. Blue links are canals. Cities (triangles on map) shaded according to relative output in final output distribution. Network is considered stable (SLNS in the paper) at this point, thus final output distribution derived from this set of networks. Canals only (no railroads) counterfactual.

Figure 37: Counterfactual Railroad Network Example: 0 and 500 Iterations



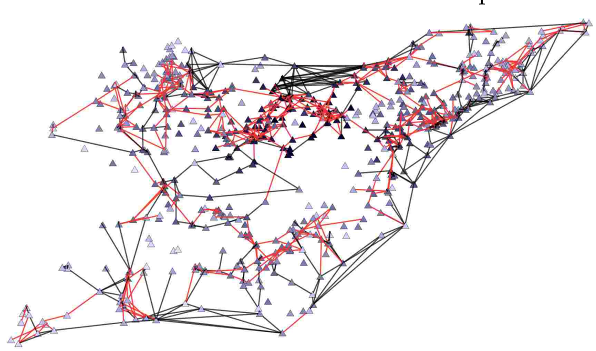
Note: Figure shows one possible result of stable network search algorithm after 0 (left, initial condition) and 500 (right) iterations. Red links are railroads. Cities (triangles on map) shaded according to relative output in final output distribution. Railroads only (no canals) counterfactual.

Figure 38: Counterfactual Railroad Network Example: 1000 and 1500 Iterations



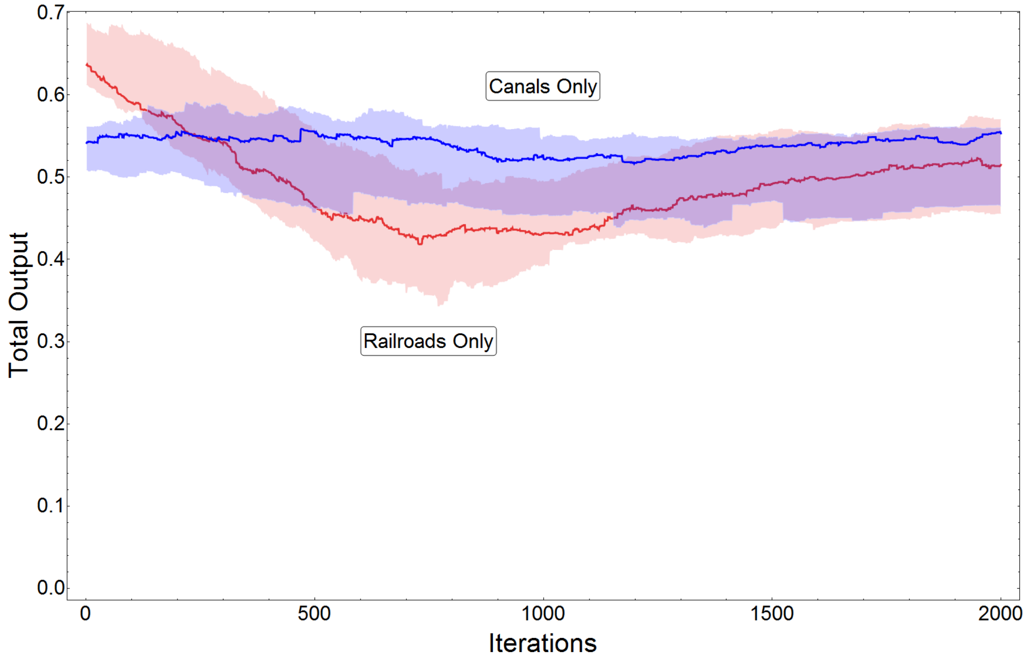
Note: Figure shows one possible result of stable network search algorithm after 1000 (left) and 1500 (right) iterations. Red links are railroads. Cities (triangles on map) shaded according to relative output in final output distribution. Railroads only (no canals) counterfactual.

Figure 39: Counterfactual Railroad Network Example: 2000 Iterations



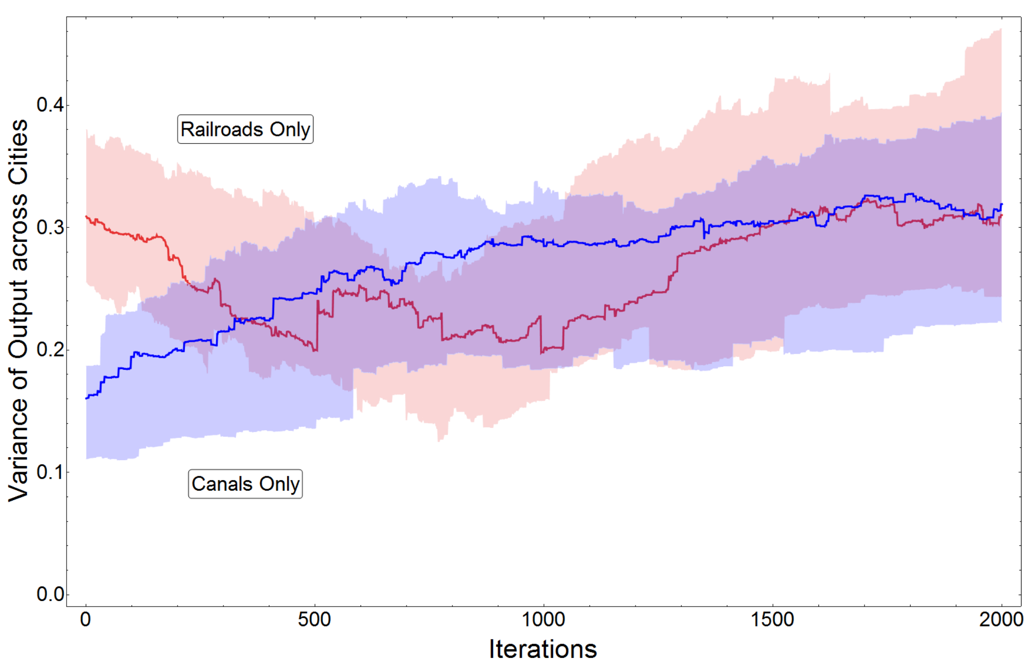
Note: Figure shows one possible result of stable network search algorithm after 2000 iterations. Red links are railroads. Cities (triangles on map) shaded according to relative output in final output distribution. Network is considered stable (SLNS in the paper) at this point, thus final output distribution derived from this set of networks. Railroads only (no canals) counterfactual.

Figure 40: Comparing Counterfactuals: Output



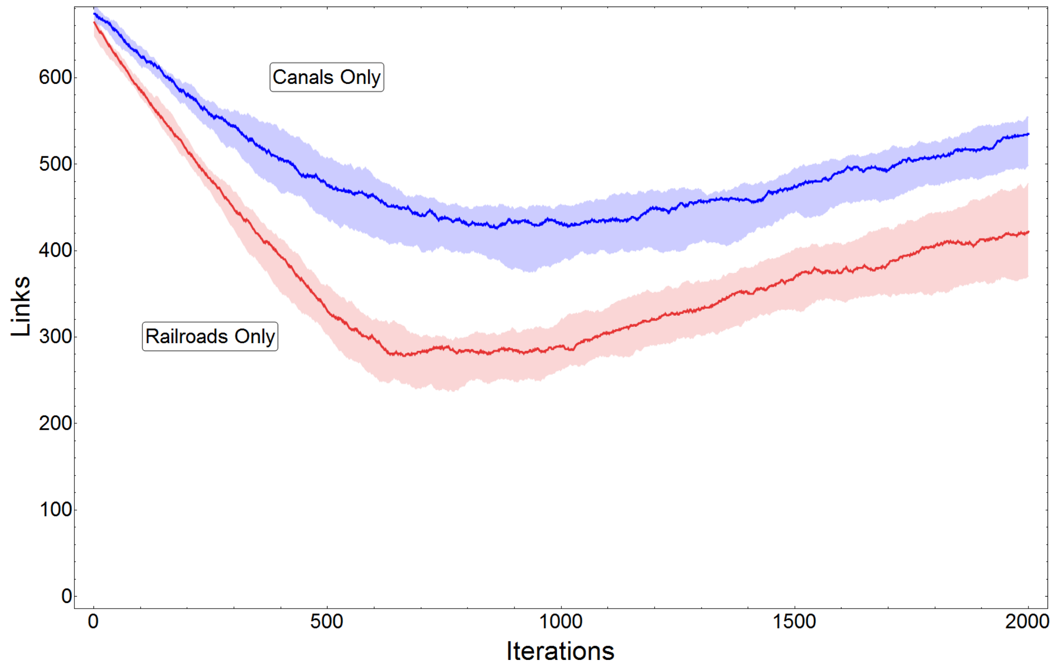
Note: Observed output normalized to one. Bands are 90% coverage bounds due to multiple equilibria. Canals only counterfactual in blue and railroads only counterfactual in red. Iterations from stable network search algorithm.

Figure 41: Comparing Counterfactuals: Variance of Output



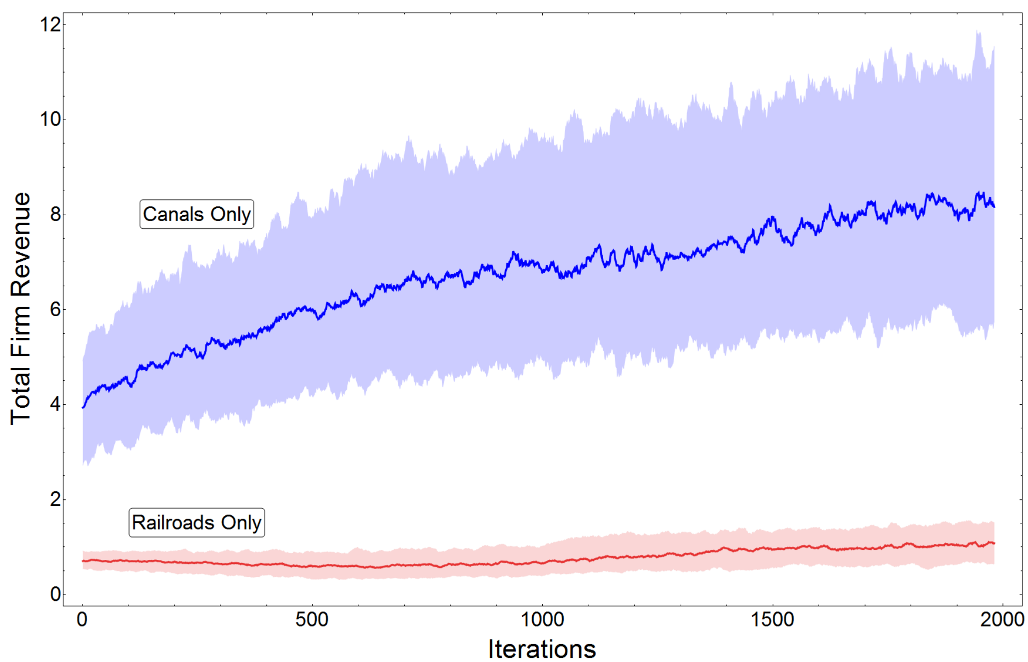
Note: Observed variance of output across cities normalized to one. Bands are 90% coverage bounds due to multiple equilibria. Canals only counterfactual in blue and railroads only counterfactual in red. Iterations from stable network search algorithm.

Figure 42: Comparing Counterfactuals: Links



Note: Plot shows number of links in counterfactual networks. Bands are 90% coverage bounds due to multiple equilibria. Canals only counterfactual in blue and railroads only counterfactual in red. Iterations from stable network search algorithm.

Figure 43: Comparing Counterfactuals: Firm Revenue



Note: Observed total firm revenue for all firms in each sector (canals, railroads) normalized to one (for each sector separately). Bands are 90% coverage bounds due to multiple equilibria. Canals only counterfactual in blue and railroads only counterfactual in red. Iterations from stable network search algorithm.

B Tables

Table 1: List of U.S. Canals Operational in 1861

Name	State(s)
Bald Eagle & Spring Creek Navigation	PA
Beaver & Erie Canal	PA
Black River Canal	NY
Blackstone Canal	MA, RI
Champlain Canal	NY
Chemung Canal	NY
Chenango Canal	NY
Chesapeake & Delaware Canal	DE, MD
Chesapeake & Ohio Canal	DC, WV, MD
Cincinnati & Whitewater Canal	IN, OH
Cumberland & Oxford Canal	ME
Delaware & Hudson Canal	NY, PA
Delaware & Raritan Canal	NJ
Delaware Division	NJ, PA
Eastern Division	PA
Erie Canal	NY
Genesee Valley Canal	NY
Green River Canal	KY
Hocking Canal	OH
Illinois & Michigan Canal	IL
James River & Kanawha Canal	VA
Juniata Division	PA
Kentucky River Canal	KY
Lehigh Canals	PA
Morris Canal	NJ, NY, PA
North Division	PA
Ohio and Erie Canal	OH
Oswego Canal	NY
Pennsylvania & Ohio Canal	OH, PA
Rappahannock Navigation	VA
Roanoke Navigation	NC, VA
Sandy & Beaver Canal	OH, PA
Schuylkill Navigation	PA
Susquehanna Division	PA
Susquehanna and Tidewater Canal	MD, PA
Union Canal	PA
Upper Appomattox Navigation	VA
Wabash & Erie Canal	IN, OH
West Branch	PA
Western Division	PA
Whitewater Canal	IN
Wisconsinco Canal	PA

Note: Source is American Canal Society. Only canals operational in 1861 are included in this list (many canals stopped operating in the decades after 1861).

Table 2: List of Selected Major U.S. Waterways

Name
Alabama River
Allegheny River
Arkansas River
Brazos River
Chattahoochee River
Cobham Bay
Connecticut River
Cumberland River
Delaware River
Green River
Hudson River
Illinois River
Kentucky River
Lake Champlain
Mississippi River
Missouri River
Neuse River
Ohio River
Potomac River
Red River
Roanoke River
Sandusky River
Sangamon River
Santee River
Savannah River
Susquehanna River
Tar River
Tennessee River
White River
Wisconsin River

Note: Source is U.S. National Atlas. Waterways selected according to national prominence (historical traffic volume) and navigability, counted as usable transport links in the empirical work.

Table 3: Observed Transport Network Prices

Transport mode	1860 cost (cents/ton-mile)	1890 cost (cents/ton-mile)	Assumed cost (normalized)
Roads	15	23.1	1
Railroads	1.96	0.63	0.098
Canals	0.99	0.49	0.098
Natural waterways	0.37 - 0.70	0.49	0.098
Coastal shipping		0.49	0.098

Note: Source is Taylor (1951) for 1860 and Fogel (1964) for 1890, reproduced in Donaldson and Hornbeck (2013). Prices are usage fees expressed in cents per ton-mile of cargo hauled; 1860 cents for 1860 costs and 1890 cents for 1890 costs. Assumed cost is what is used in this paper, which is interpreted as fraction of cargo taken as a usage fee per mile (which depends on the value of the cargo hauled and the length of the trip). This is done to abstract away from quality differences in railroad and canal service (railroad was faster, more reliable, etc., thus priced higher) since what matters most is the gap between the road usage fee and everything else (this gap is an order of magnitude larger than observed gap between railroads and canals).

Table 4: Summary Statistics: Neighbor Terrain Cost

	$\mu(\eta_\theta(\ell_{AB}))$	$\sigma(\eta_\theta(\ell_{AB}))$
Railroads ($\theta = RR$)	5.12	4.01
Canals ($\theta = CN$)	0.75	0.7

Note: Mean $\mu(\eta_\theta(\ell_{AB}))$ and standard deviation $\sigma(\eta_\theta(\ell_{AB}))$ of neighbor terrain cost taken from cell penalty function $\Phi_{RR}(\ell_{AB})$ for railroads and $\Phi_{CN}(\ell_{AB})$ for canals. Neighbor terrain cost for railroads is denoted as $\eta_{RR}(\ell_{AB})$ in the paper and is interpreted as the cost of building a link from cell A to cell B when multiplied by c_{RR} (type cost for railroads) provided that cells A and B are neighbors (adjacent). Neighbor terrain cost for canals is denoted as $\eta_{CN}(\ell_{AB})$ in the paper and is interpreted as the cost of building a link from cell A to cell B when multiplied by c_{CN} (type cost for canals) provided that cells A and B are neighbors. A cell is a small rectangular block of terrain defined in the paper.

Table 5: Summary Statistics: Terrain Cost

	$\mu(\tau_\theta(i,j))$	$\sigma(\tau_\theta(i,j))$
Railroads ($\theta = RR$)	117	61
Canals ($\theta = CN$)	44	20

Note: Mean $\mu(\tau_\theta(i,j))$ and standard deviation $\sigma(\tau_\theta(i,j))$ of terrain cost taken from $\tau_{RR}(i,j)$ for railroads and $\tau_{CN}(i,j)$ for canals. Railroad terrain cost is denoted as $\tau_{RR}(i,j)$ in the paper and is interpreted as the cost of building a link from city i to city j when multiplied by c_{RR} (type cost for railroads). Canal terrain cost is denoted as $\tau_{CN}(i,j)$ in the paper and is interpreted as the cost of building a link from city i to city j when multiplied by c_{CN} (type cost for canals). Mean bilateral Euclidean distance is 550 mi. between pairs of cities.

Table 6: Cost Bounds: Inequalities

	Bound	[1]	[2]	[3]	[4]
$c_{RR} \leq$	0.0003	1540	0.03	2.8	75
$c_{RR} \geq$	0.0017	8171	0.18	14.7	396
$c_{CN} \leq$	0.0008	1431	0.03	2.6	69
$c_{CN} \geq$	0.0016	2887	0.06	5.2	140

[1]: Average Total Cost (thousands of 1861 $US)
[2]: Fraction of 1861 U.S. Output (%)
[3]: Thousands of 1861 $US/mi.
[4]: Thousands of 2012 $US/mi.

Note: Notation: c_{RR} railroad link type cost, c_{CN} canal link type cost; RR railroad, CN canal type of link. Column "Bound" (1) interpreted as the appropriate bound for the inequality in each row (ex. $c_{RR} \leq 0.0003$ for first row). Column "Average Total Cost" (2) interpreted as average total cost of building a link and paying the appropriate costs for the inequality in each row, expressed in 1861 U.S. dollars. Column "Fraction of 1861 U.S. Output (%)" (3) takes average total cost and divides by nominal output to give a sense of the resources required for link construction in 1861, expressed as a percentage. Column (4) "Thousands of 1861 $US/mi." expresses bound as nominal cost per mile in 1861 U.S. dollars for the average link. Column "Thousands of 2012 $US/mi." (5) expresses bound as nominal cost per mile in 2012 U.S. dollars.

Table 7: Cost Bounds: Intervals and Point Estimates

	Interval	Point Estimate	[1]	[2]	[3]	[4]
c_{RR}	[0.0003,0.0017]	0.001	4856	0.10	8.7	235
c_{CN}	[0.0008,0.0016]	0.0012	2159	0.05	3.9	105

[1]: Average Total Cost (thousands of 1861 $US)
[2]: Fraction of 1861 U.S. Output (%)
[3]: Thousands of 1861 $US/mi.
[4]: Thousands of 2012 $US/mi.

Note: Notation: c_{RR} railroad link type cost, c_{CN} canal link type cost; RR railroad, CN canal type of link. Column "Interval" interpreted as the interval bound for the parameter in each row (ex. $c_{RR} \in [0.0003, 0.0017]$ for first row). Column "Point Estimate" takes midpoint of interval bound (or average of left and right interval endpoints). Column "PE, Average Total Cost" (1) interpreted as average total cost of building a link and paying the appropriate cost for the parameter in each row, expressed in 1861 U.S. dollars, for the point estimate. Column "PE, Fraction of 1861 U.S. Output (%)" (2) takes average total cost and divides by nominal output to give a sense of the resources required for link construction in 1861, expressed as a percentage, for the point estimate. Column (3) "PE, Thousands of 1861 $US/mi." expresses point estimate as nominal cost per mile in 1861 U.S. dollars for the average link. Column "Thousands of 2012 $US/mi." (4) expresses point estimate as nominal cost per mile in 2012 U.S. dollars.

Table 8: Main Findings

Model Outcome	Counterfactual	90% Coverage Interval	Median	Median Relative to Observed
Output	Canals only	[0.46,0.55]	0.55	45% loss
	Railroads only	[0.45,0.57]	0.51	49% loss
Variance of Output	Canals only	[0.22,0.39]	0.31	69% loss
	Railroads only	[0.24,0.46]	0.31	69% loss
Links	Canals only	[499,554]	535	508% gain
	Railroads only	[371,477]	422	45% loss
Firms	Canals only	[51.9,67.4]	60	43% gain
	Railroads only	[17,33.4]	26	83% loss
Firm Revenue	Canals only	[5.73,11.54]	8.16	716% gain
	Railroads only	[0.64,1.51]	1.07	7% gain

Note: "Canals only" refers to canals only counterfactual, specifically properties of the set of single link Nash stable canal networks in the absence of railroads; "Railroads only" refers to the railroads only counterfactual. All observed quantities normalized to one for ease of comparison, except for number of links and firms. Notation: "Output" is total output across all cities, "Variance of Output" is the variance of output across individual cities, "Links" is the number of links in the counterfactual network, "Firms" is the number of transport firms in the counterfactual network, and "Firm Revenue" is total firm revenue across all firms of a given type (railroad, canal). Counterfactual results are presented as the median across equilibria (single link Nash stable or SLNS transport networks) with associated 90% coverage intervals to address uncertainty due to multiple equilibria. A uniform prior over equilibria in the set of stable networks is assumed.

C Additional Motivation

At the end of this section I discuss a stylized example to show how a stable network search algorithm that individually permutes links may fail to find the globally optimal network from the social planner's perspective. First, I examine the available empirical evidence on 19th century U.S. economic growth and railroads. Second, I attempt to generate a simple counterfactual to answer the question using a reduced-form vector autoregression (VAR) model. Lastly, I briefly discuss the issues involved in taking a modified version of the canonical growth model to the data in an attempt to estimate the growth contribution of railroads. The VAR and canonical growth models are used as motivation for taking a more structural game theoretic network formation approach to answering Fogel's classic counterfactual history question on railroads and growth.

C.1 Historical Macroeconomic Data Facts

C.1.1 Standard Macroeconomic Aggregates

Figure 45 in the Appendix (part C) plots output over time for the United States in the 19th century. The light solid line is the data and the heavy dashed line is the Hodrick-Prescott filtered trend.[82] As one would expect, output fluctuates in an apparently random manner around its long-run trend. The average growth rate over the period 1800-1900 is 4%. Focusing on the time horizon for Fogel's counterfactual, 1830-1890, the average output growth rate is 4.4%.

The hypothesis posited by historians such as Jenks (1944) claims that railroads induced a burst of rapid growth in output during this period. Examining the data, this claim has limited support. Splitting the data at 1830, roughly when railroads were introduced into the U.S. economy, the average growth rate is 3.5% before 1830 (1800-1829) and 4.5% after (1830-1900). This implies a growth rate differential of 1% per year, which is suggestive of the "marginal growth contribution of railroads" in an informal way. Canals were present throughout the entire century and railroads were introduced only in 1830, so looking just at the data one could claim that without railroads, output growth in the U.S. would have been 1% lower for 60 years. This implies an output loss in 1890 due to no railroads of 46%. Of course, such a direct answer to the question is unsatisfactory because other variables correlated with output were also varying during the period.

The cyclical component of output is derived from the same Hodrick-Prescott filter. Insofar as the trend is the long-run trajectory and the cycle is an anomaly relative to trend, there is no systematic evidence here for viewing the introduction of railroads as a positive output growth shock. The cyclical component oscillates above and below zero according to the business cycle, which seems qualitatively unchanged by railroads. There is a 5-year spike which peaks a few years after 1830, but its duration is too short to claim that it is due to the supposedly long-lasting effects of the introduction of railroads. Note from 45 that the volatility of the cyclical component is increasing over time. Perhaps railroads destabilized output? However, such a proposition is not directly related to the growth story as usually told in the history literature.

Figure 46 is a plot of the U.S. physical capital stock over the 19th century. Both canal and railroad plant and equipment are included in this series. The average growth rate of the capital stock over this period is 4.25%. So, with a standard value for capital's share of national income of 1/3, capital can explain 35% of output growth on average over the 19th century. Unless railroads contaminated the total factor productivity

[82]Such a value of $\lambda = 100$ is appropriate for annual data.

series, this places an upper bound of on the marginal growth contribution of railroads of 1.4% per year, which cumulates to an output loss of 58% in 1890.

See Figure 47 for a plot of the U.S. population over the 19th century. On average, the population of the United States grew 3% by per year during this period. Population growth is a reasonable proxy for growth in the labor force and subsequent growth in labor supply or hours worked. Thus, with labor's share equal to 2/3, growth in labor can account for 49% of output growth on average for the period. This leaves 16% of output growth due to productivity growth over the period. So, attributing all of capital growth and productivity growth to railroads, the marginal growth contribution is certainly bounded above by 2% per year, or an output loss of 72% in 1890. Through an argument that immigrants were attracted to the U.S. due to railroads, augmenting labor supply, one could try to inflate this upper bound.

C.1.2 Transport Networks

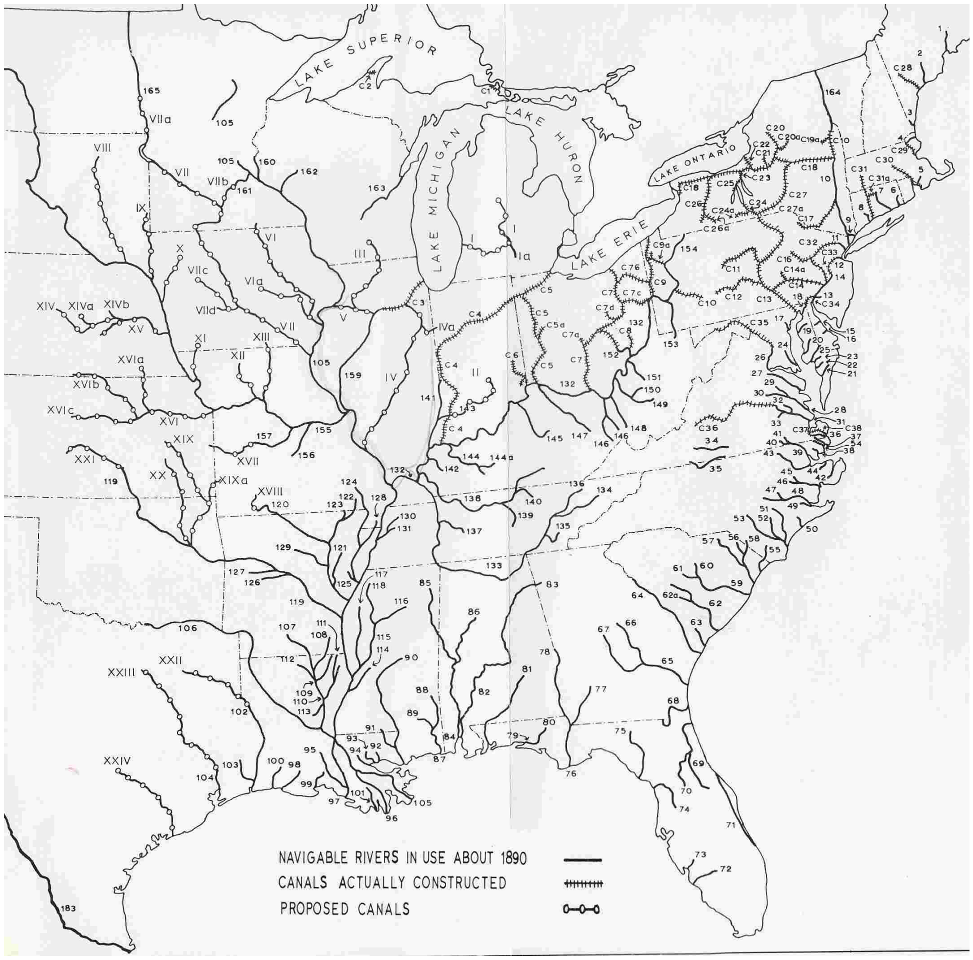
Figure 48 shows the growth of the United States railroad network in terms of total mileage of track in operation (Appendix, part C). The first commercial railroad in the U.S., the Baltimore & Ohio Railroad, was chartered in 1829 and had 13 miles of track in operation in 1831. Railroad building booms occurred in the 1850s and late 1860s post-Civil War, but here they appear as relatively minor deviations from long-run growth. Although not shown in the plot, the railroad network reached its peak mileage of 300,000 in 1929 and has declined since.[83]

The evolution of total mileage for the U.S. canal network is presented in Figure 49. Again, when framed in terms of long-run growth, behavior of the canal network appears predictable. Adjusting for differences in the introduction date of the technology, canal growth was more rapid than that of railroads, with canals saturating after 60 years but railroads taking roughly 80 years to do so.

Figure 50 plots the result of fitted logistic growth curves for railroads and canals: growth rates for the two transport networks from 1790 to 1916. The data have been adjusted in two ways. First, to account for when the canal data goes missing, when the level of canal mileage goes from 4000 to zero for the missing values, one very negative growth rate is set to zero. Second, to account for the introduction of the two technologies into the economy (i.e. when the data starts), the first very positive growth rate is set equal to 1/2 times the second growth rate (which uses two values of actual data that are far from zero). Again, canals reached the saturation point faster than railroads. Railroads were introduced at about the halfway point for saturation of canals.

[83] Both passenger and freight traffic for railroads also peaked in the 1920s.

Figure 44: Fogel's Alternative Canal Network (1890 U.S.)



Note: Source is Fogel (1964). Date is 1890. Solid lines denote navigable rivers (designated with numbers), cross-dashed lines denote canals actually constructed (designated with C followed by numbers), lines with circles denote new canals proposed by Fogel that were not actually built in the observed 1890 canal network (designated with Roman numerals). Note that this paper does not use any of Fogel's original data on waterways or canals built in 1890 (since the date examined in this paper is 1861).

C.2 Preliminary Counterfactuals in Reduced-form VAR Model

Now use the data presented in the previous section to estimate two reduced-form vector autoregressions: the first in levels and the second in growth rates. Define output as Y_t, capital K_t, labor L_t, miles of railroad RR_t, and miles of canals CN_t at year t. The first VAR I will consider is estimated with vector of data $X_t = [Y_t, K_t, L_t, RR_t, CN_t]'$ in levels from 1830 to 1890 with three lags. For pairwise correlations of the data used in the VAR, see Table 9 in the Appendix, part C. For a relative volatility table with the standard deviation of output normalized to one, see Table 10. Railroads are much more volatile than canals or the physical capital stock in general, but less volatile than output.

The estimated equation of this VAR model is

$$X_t = b_1 X_{t-1} + b_2 X_{t-2} + b_3 X_{t-3} + \varepsilon_t$$

where ε_t is a vector of reduced-form (regression) errors. Since ε_t is a regression error, it does not have a structural interpretation and it is not usable to form interpretable impulse response functions. However, we can run Granger causality tests and a simple counterfactual where RR_t is set equal to zero for all $t \geq 1840$ in the model estimated using the factual data. The VAR is then run until 1890 using $RR_t = 0$ with all other variables unchanged except output, which is fed back into the model. Granger causality tests indicate that railroads Granger cause capital and canals; canals Granger cause output and railroads ($\alpha = 0.01$ level). Thus the feedback between canals and railroads goes in both directions.

A problem immediately presents itself when running a simple counterfactual with three lags in the VAR: due to the sign of the estimated coefficients on RR_{t-1}, RR_{t-2}, and RR_{t-3}, a nonsensical result is generated in the no railroads case where 1890 output increases (relative to the factual) by a factor of 3.5. In the no canals case, output falls by 55%. So the model with three lags gets the sign wrong: it must be the case that output falls without railroads, the question is how much does it fall empirically. One possible resolution to this problem is to remove lags from the VAR until the counterfactual result makes sense. However, when capital is included with only one lag, the estimated coefficient on railroads is negative, which will again generate this result that output increases in the no railroads counterfactual.

When capital is removed from the model with one lag, the estimated coefficients on both railroads and canals are positive. This model generates the results shown in Figure 51. Starting from 1839 observed output, the VAR is run with $RR_t = 0$ (solid line) and $CN_t = 0$ (dashed line) until 1890. The output loss is then the difference between observed output in 1890 and counterfactual output in 1890 generated by the VAR model. Thus 1890 output falls by 80% in the no railroads counterfactual and 29% in the no canals counterfactual. However, there is yet another problem: output goes negative in the no canals counterfactual but recovers to finish positive. Such difficulty (output increasing in the counterfactual, output going negative, etc.) motivates the need for a structural network formation model to reformulate Fogel's original question on railroads and growth.

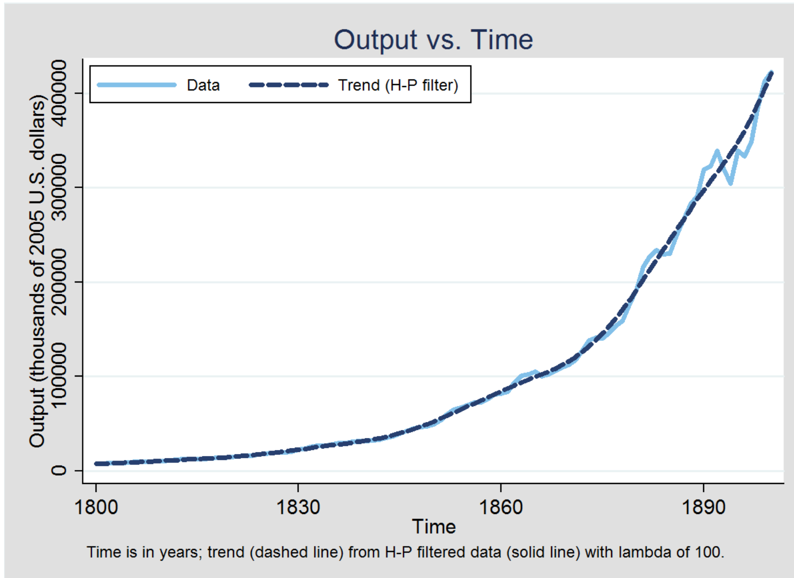
The second VAR is estimated with $\Delta X_t / X_t = [\Delta Y_t / Y_t, \Delta L_t / L_t, \Delta RR_t / RR_t, \Delta CN_t / CN_t]'$ in growth rates from 1830 to 1890 with one lag (capital excluded). This specification of the VAR was chosen to match the final specification in levels. The estimated equation of this model is

$$\frac{\Delta X_t}{X_t} = b \frac{\Delta X_{t-1}}{X_{t-1}} + \varepsilon_t$$

with results reported in Figure 52. The counterfactual experiment is analogous to the one run for the VAR model specified in levels. However, since the VAR is now in differences, it is necessary to compound the growth rate differences shown in Figure 52 in order to compute the output loss in 1890. Doing so, 1890 output increases by a factor of 7.22 in the no railroads counterfactual (solid line) and falls by 6% in the no canals counterfactual (dashed line). This occurs because the estimated coefficient on railroad mileage growth is negative, which is again problematic. One would expect that a structural model should at least get the sign correct.[84] Thus to conclude, I find that writing down the simplest possible reduced-form VAR model does not provide a satisfactory resolution to the question of railroads and growth.

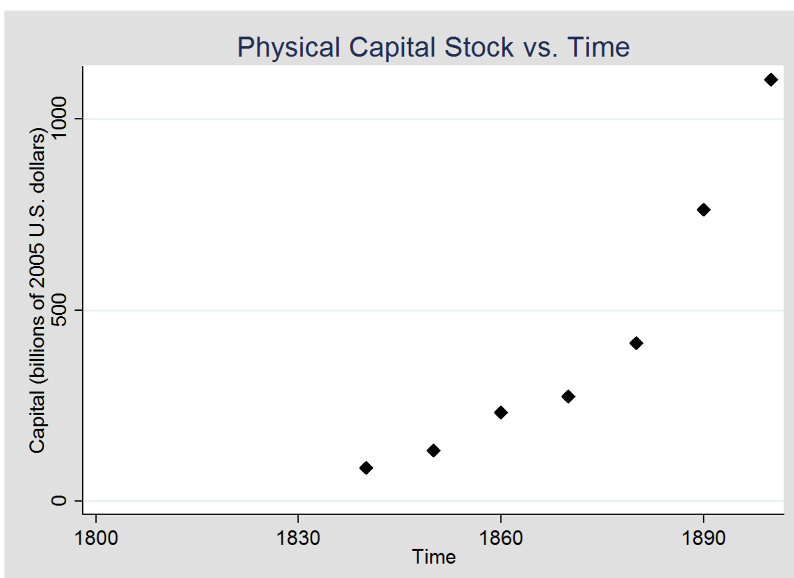
[84] Barring some counterintuitive channel, output should always fall in the counterfactuals since productive technology is being removed from the economy.

Figure 45: U.S. Output (Y)



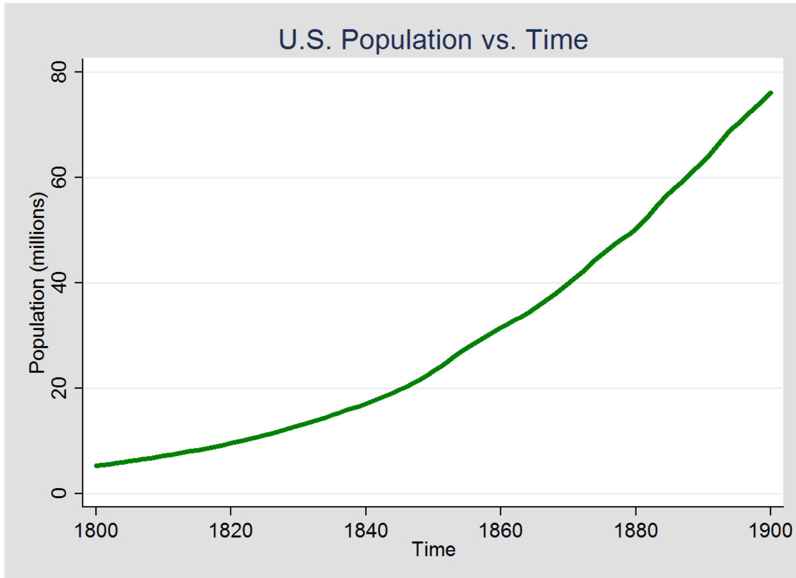
Note: Figure plots data on U.S. output (GNP, solid light blue, i.e. Y) along with Hodrick-Prescott (H-P) filtered trend ($\lambda = 100$, dashed dark blue). Source is Historical Statistics of the United States: Millennial Edition.

Figure 46: U.S. Physical Capital (K)



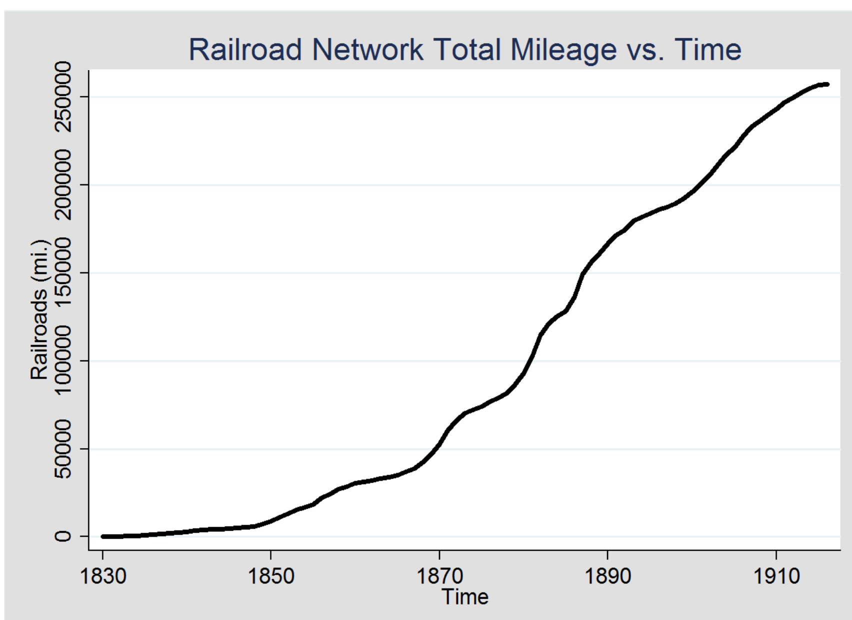
Note: Figure plots data on U.S physical capital stock (i.e. K). Source is Engerman and Gallman (1986).

Figure 47: U.S. Population (L)



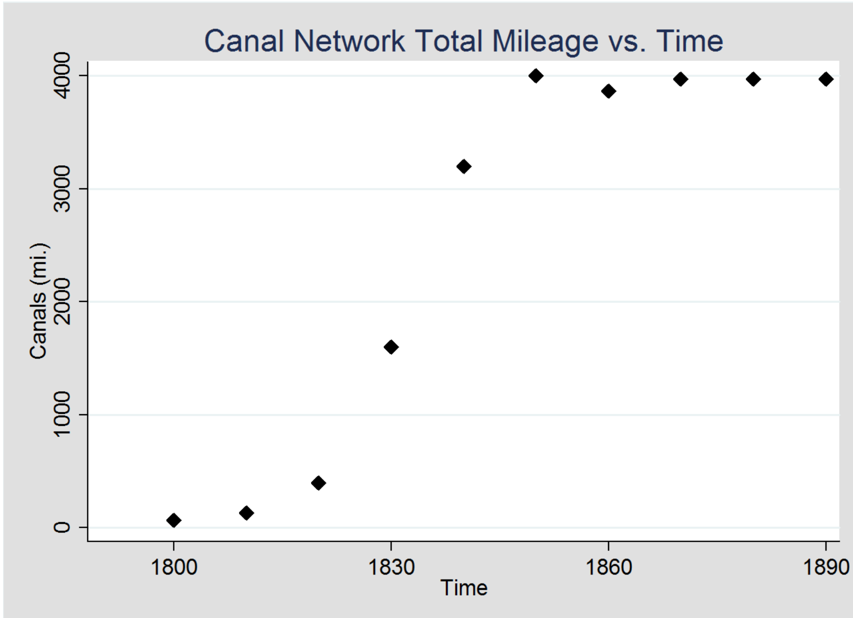
Note: Figure shows data on total U.S. population (i.e. L). Source is Historical Statistics of the United States: Millennial Edition.

Figure 48: U.S. Railroads (RR)



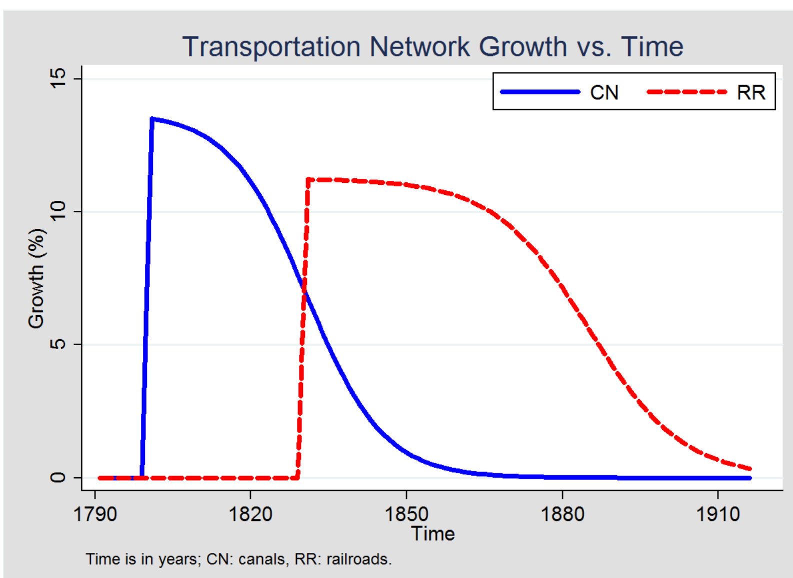
Note: Figure shows data on total mileage of U.S. railroad network (i.e. RR). Source is Historical Statistics of the United States: Millennial Edition.

Figure 49: U.S. Canals (CN)



Note: Figure shows data on total mileage of U.S. canal network (i.e. CN). Source is Grubler (1990). Data available each decade.

Figure 50: Growth of U.S. Transport Network



Note: Figure presents transport network growth rates for canals (CN, solid blue line) and railroads (RR, dashed red line). Source is Historical Statistics of the United States: Millennial Edition and Grubler (1990). Canal data interpolated from decade frequency. Note that railroads were introduced to the U.S. about 20 years after the introduction of canals. Growth rates not defined for first year that links were built thus omitted here.

Table 9: Pairwise Correlations

	Y	K	L	RR	CN
Y	1				
K	0.99	1			
L	0.97	0.98	1		
RR	0.98	0.98	0.94	1	
CN	0.60	0.63	0.73	0.51	1

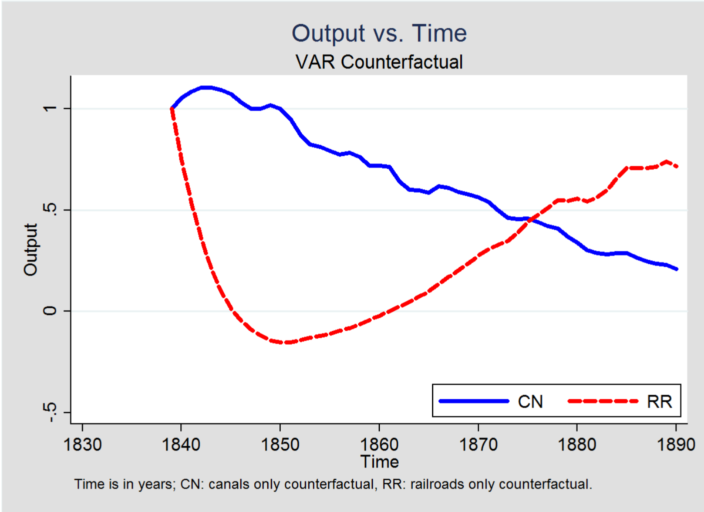
Note: All variables in levels from 1830 to 1890. Notation: Y is output (GNP), K is capital stock, L is labor (total population), RR is total miles of railroad lines, CN is total canal mileage.

Table 10: Relative Volatility

	σ
Y	1
K	0.002
L	0.189
RR	0.626
CN	0.006

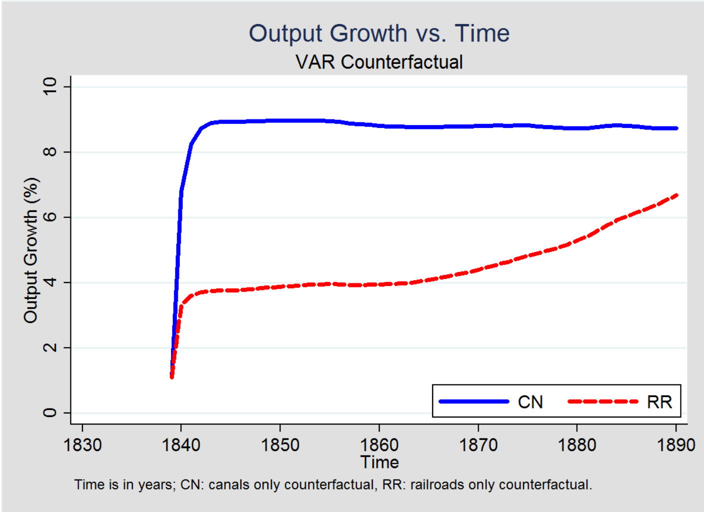
Note: All variables in levels from 1830 to 1890, volatility of output normalized to one for comparison purposes. Notation: Y is output (GNP), K is capital stock, L is labor (total population), RR is total miles of railroad lines, CN is canal mileage.

Figure 51: VAR Counterfactual for Output in Levels



Note: Figure presents resulting path for output when $RR = 0$ for all dates (CN, canals only counterfactual, solid blue line) or $CN = 0$ for all dates (RR, railroads only counterfactual, red dashed line). VAR model specified in levels, first specification in the text, from 1840 to 1890. Canal data interpolated from decade frequency. Note that path for output goes negative temporarily for railroads only counterfactual since VAR model does not impose nonnegativity constraint.

Figure 52: VAR Counterfactual for Output in Growth Rates



Note: Figure presents resulting path for output when $RR = 0$ for all dates (CN, canals only counterfactual, solid blue line) or $CN = 0$ for all dates (RR, railroads only counterfactual, red dashed line). VAR model specified in growth rates, second specification in the text, from 1840 to 1890. Canal data interpolated from decade frequency. Note that removing canals is substantially more harmful for growth than removing railroads.

C.3 Preliminary Counterfactuals in Growth Model

C.3.1 Baseline Model

Consider a version of the canonical growth model with three factors of production: land L_t, in fixed supply, conventional physical capital K_t, reproducible, and transportation capital K_t^T, also reproducible.[85] Assume an infinite horizon $t = 0, 1\ldots$ with initial conditions $K_0 = \overline{K}_0$, $K_0^T = \overline{K}_0^T$; physical capital depreciates at rate $\delta \in (0,1)$ and transport capital depreciates at rate $\delta^T \in (0,1)$. The production function

$$Y_t = f(L_t, K_t)$$

satisfies the usual conditions: $f(.,.)$ is increasing and concave in each of its arguments, with positive cross-partials, satisfying the Inada conditions in land and capital. Transport capital does not enter the production function directly, but rather helps in the assimilation of additional land to use in production according to

$$L_t = g(K_t^T)$$

where $g(.)$ is increasing, concave, and satisfies the Inada condition in transport capital.[86] Household preferences over consumption streams $\{C_t\}_{t=0}^{\infty}$ are ordered by

$$U = \sum_{t=0}^{\infty} \beta^t u(C_t)$$

where $\beta \in (0,1)$ is the discount factor and $u(.)$ is the period utility function, which is increasing, concave, and satisfies the Inada condition in consumption. The way the production technology has been defined, there are no externalities or frictions, so the welfare theorems hold and it is both necessary and sufficient to consider the solution to the planner's problem to characterize the decentralized competitive equilibrium.

Let X_t denote investment in physical capital, X_t^T investment in transport capital, and C_t consumption in period t. The resource constraint is

$$Y_t = C_t + X_t + X_t^T$$

in a closed economy without fiscal or monetary policy. The planner solves dynamic programming problem

$$V(K, K^T) = \max_{K', (K^T)'} \left\{ u(C) + \beta V(K', (K^T)') \right\}$$

$$\begin{aligned} \text{s.t.} \quad K' &= (1-\delta)K + X \\ (K^T)' &= (1-\delta^T)K^T + X^T \\ Y &= f(L, K) = C + X + X^T \\ L &= g(K^T) \end{aligned}$$

with first-order conditions

$$\begin{aligned} [K'] &: \quad u'(C) = \beta V_K(K', (K^T)') \\ [(K^T)'] &: \quad u'(C) = \beta V_{K^T}(K', (K^T)') \end{aligned}$$

and envelope conditions

[85] Transportation capital is a subset of physical capital as usually defined. So, to be clear, capital K_t here excludes transport capital K_t^T to avoid double counting.

[86] This formulation of the model maps into an observationally equivalent reduced form where transport links enter the production function directly.

$$\begin{aligned}
[K] &: \quad V_K(K, K^T) = u'(C)\left(1 + f_K(L, K) - \delta\right) \\
[K^T] &: \quad V_{K^T}(K, K^T) = u'(C)\left(1 + f_L(L, K)g'(K^T) - \delta^T\right)
\end{aligned}$$

which, when combined, yield the usual consumption Euler equations

$$\begin{aligned}
u'(C) &= \beta u'(C')\left(1 + f_K(L', K') - \delta\right) \\
u'(C) &= \beta u'(C')\left(1 + f_L(L', K')g'((K^T)') - \delta^T\right)
\end{aligned}$$

in the two possible investment opportunities: physical capital and transport capital. Given the initial conditions, these equations can be solved forward numerically to determine the time paths for consumption, physical capital, transport capital, and output. It is immediate that

$$f_K(L, K) - \delta = f_L(L, K)g'(K^T) - \delta^T$$

so the after depreciation rates of return on the two investment technologies are equated in all periods; investors are indifferent between investing in physical capital and transport capital.

Without a production externality for transport capital, the resulting competitive equilibrium is efficient and what matters for the rate of transport capital growth is parameters and functional forms for $u(.)$, $f(.,.)$, and $g(.)$. With a production externality,

$$Y = \bar{L}f(L, K)$$

where $\bar{L}$ is the aggregate stock of land taken as given by firms and households. Therefore, the private return to investment in transport capital is

$$R^T_{priv.} = 1 + \bar{L}f_L(L, K)g'(K^T) - \delta^T$$

but the social return is

$$R^T_{soc.} = R^T_{priv.} + g'(K^T)f(L, K)$$

with inequality $R^T_{priv.} < R^T_{soc.}$ and the usual underinvestment result obtains.

C.3.2 Adding railroads and canals

Split transport capital K^T into two types: railroad capital K^{RR} and canal capital K^{CN}. Assume that railroad capital depreciates at rate δ^{RR} and canals depreciate at rate δ^{CN}. The function for the evolution of land becomes

$$L_t = g(K^{RR}_t, K^{CN}_t)$$

where it is reasonable to continue to assume that $g(.,.)$ is increasing and concave in both arguments, but the sign or magnitude of the cross partial term is an empirical question. For example, under the case of large complementarities between the two types of transport capital,

$$\frac{\partial^2 g}{\partial K^{RR} \partial K^{CN}} \gg 0$$

and, if the railroad or canal technology were missing from the economy, the resulting output loss would be large. Also, being able to run meaningful counterfactuals without railroads or canals implies that the acquisition of land still takes place at some strictly positive level with $K^{RR} = 0$ or $K^{CN} = 0$. With $g(.,.)$ symmetric in the two types and $\delta^{RR} = \delta^{CN}$, the output loss associated with taking railroads or canals away must be the same. Thus the debate in the literature about the marginal growth contribution of railroads is fundamentally about the empirical properties of the $g(.,.)$ function. However, ex ante, I do not have strong priors on the functional form or properties of this function, and it may be difficult to identify in practice.

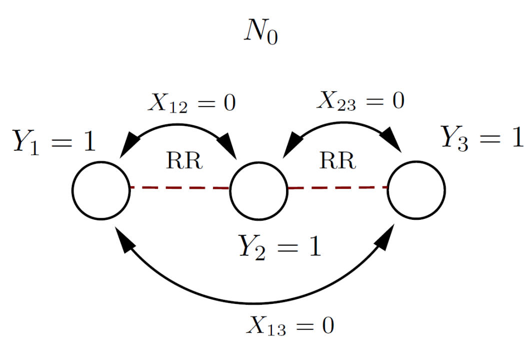
Although there are certainly many other possible sources of variation in output in the data, for the purposes of this paper, railroads and canals will be the only forcing process for output in order to isolate their effect.

C.4 Illustrative Example: Social Planner's Problem

In general inefficiency is present in the decentralized stable transport network due to network effects. Transport firms do not fully internalize the beneficial external effects of link construction on trade since any individual firm does not own all transport links. Additionally firms do not take into account the negative effect of link construction on their competitors, specifically trade diversion and siphoning off of competitor traffic due to completion of a new link. The direct solution to this inefficiency is mergers and acquisitions among transport firms; interesting, but beyond the scope of this paper due to computational feasibility.[87] However, with multiple modes of transport and each mode owned by a separate monopolist, firms will still not fully internalize the externality. Thus a benevolent social planner can potentially improve economic outcomes.

The difficulty in solving the social planner's problem, or any equilibrium decentralized stable network problem in general, is due to the need to check robustness of the network to deviations that consist of individual links or sets of links. To illustrate this issue, consider a stylized example similar to the one used previously to discuss complementarities between railroads and canals. Since the inefficiency exists independent of the complements issue, consider a world where only railroads are available for construction. The figure below illustrates the initial condition of such a world. Since the planner does not have to keep track of the distribution of firms, consider the case where a single state-owned company builds and operates the transport network. Link construction costs are paid lump-sum out of total output with no distributional issues across cities (nodes).

Figure: Empty Network



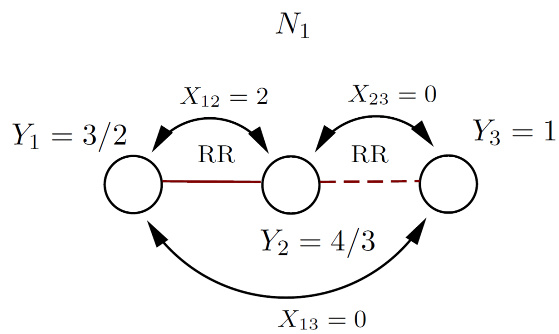
The initial condition is the empty network with no links but two potential railroad links (dashed lines, RR), denoted N_0. As before, with no connectivity output is 1 for all cities and no trade takes place: $X_{ij} = 0$ for

[87] It is difficult to evaluate the value to the firm of all possible mergers between firms, a combinatorial problem, since it turns out that the payoff function has no closed-form solution and must be simulated. Each competing firm is unique.

all i, j. Again the stylized trade protocol $X_{ij} = Y_i Y_j$ is of gravity type and city output Y_i increases when trade costs decline. Network N_0 has total output 3 and no links, thus no link construction costs.

Suppose that the planner runs a particular search algorithm to find the optimal network configuration, where optimal is defined as the transport network that maximizes total output across cities minus total link construction costs (i.e. total consumption).[88] This algorithm starts with the empty network and changes (or permutes) links one at a time to find the global optimum. Starting from the empty network, no links exist to take away, so the planner must try to add a link. First the algorithm tries to add link 1 as seen in the figure below.

Figure: Adding Link 1

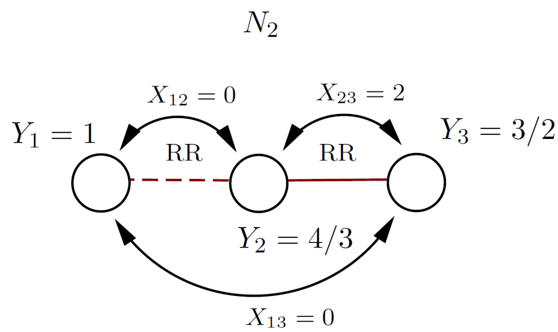


Denote this alternative network as N_1, where the planner attempts to build one of the possible links but not the other. Assume that both links cost one unit of output to build. A modest increase in output and trade flows occurs due to the lowering of trade costs by link construction, but upon further examination this is insufficient to cover the costs incurred by the planner. Total output increases to $23/6$ but with an investment cost of 1 the planner is left with only $23/6 - 1 = 17/6 < 3$ in total consumption after subtracting investment from output. Since the empty network N_0 results in total consumption 3, the planner rejects moving from N_0 to N_1 and continues searching for the global optimum in the space of transport networks.

The search algorithm now tries to add link 2, but the result is identical to that of the attempt to add link 1 and is similarly rejected. When link 2 is added and the planner moves to network N_2 from N_0, total consumption again falls to $17/6 < 3$ and the planner rejects the move. Since network N_2 mirrors N_1, the outcomes are symmetric; it appears to the algorithm that the losses in increased investment costs outweigh the gains from more trade and increased output. If the planner's algorithm starts at N_0, it is now stuck there since both possible moves to N_1 and N_2 are rejected if the planner is restricted to individual link permutations.

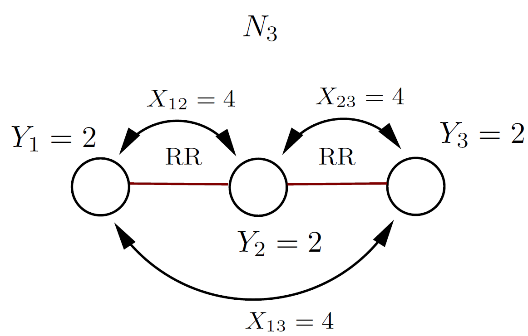
[88]This example ignores distributional issues across cities, as if the planner has the ability to make costless lump-sum transfers of the consumption good between cities. Thus all that matters is total consumption and not its allocation.

Figure: Adding Link 2



Because the set of potential links is so small, brute force search suggests that the last remaining network state to check is the complete network with both links present simultaneously. The complete network N_3, which requires that both links are built at once moving from N_0 to N_3, is presented below.

Figure: Complete Network



The complete network maximizes total output by minimizing trade costs, but this configuration does not necessarily maximize total consumption and thus welfare. Total output is 6 and consumption is 4 with investment costs of 2.

The relevant model outcomes for brute force search are total output Y, investment I in transport capital in form of network links, and consumption C. The planner wants to maximize C by choosing the network. Model outcomes across all four possible network configurations N_0 to N_3 are presented in the following table.

Table: Model Outcomes by Network

	Y	I	C
N_0	3	0	3
N_1	23/6	1	17/6
N_2	23/6	1	17/6
N_3	6	2	4

Since the planner's objective is to maximize C by choosing the network, complete network N_3 is optimal. However, starting from N_0 as the initial condition and permuting individual links in the spirit of an algorithm that searches for unilateral deviations, the algorithm rejects moving from N_0 to N_1 or N_2, thus such a search does not find global optimum N_3 from N_0. This is due to the initial seeding of links in the form of empty network N_0 and only allowing for single link deviations. The empirical example one should have in mind here is the Transcontinental Railroad, where a coordination problem exists to build all links at once in a long chain to complete the connection to the West.[89]

If the search algorithm allowed for general deviations in the form of sets of links, the planner would easily find N_3 from N_0. In such a stylized setting, it is easy to exhaust all possible deviations composed of bundles of links and even feasible to perform brute force search by checking the payoff function across all possible states. However, in the data with hundreds of cities and hundreds of thousands of potential links, the number of possible networks to check scales up rapidly and brute force search becomes computationally infeasible, as does checking all possible deviations involving sets of links. The curse of dimensionality associated with this problem is such that too many multi-link deviations exist to check in reasonable running time for a candidate search algorithm.[90]

Although I use a search algorithm that permutes links individually to restore computational feasibility, I address this issue by drawing more long routes that would not be potential links if only nearest neighbors were considered for potential link candidates. I also randomly seed initial links, specifically overshooting by placing too many links (and firms, in the decentralization) initially, so the likelihood of seeding network state N_3 directly as the initial condition for search (and finding it as the global optimum) increases.

[89]Government officials understood at the time that private incentives were insufficient to support construction of the Transcontinental Railroad. This coordination problem was solved by the government through subsidization of link construction, which unfortunately devolved into the Credit Mobilier corruption scandal where a subsidized railroad company (the Union Pacific Railroad) systematically overbilled the government through an intermediary.

[90]Lurking in the background here is the larger issue of how cities or links are defined or which potential links are allowed. It turns out that aggregating the cities by combining them is not a productive solution because trade model outcomes are not invariant to the number of cities, thus varying the number of cities from the observed to the counterfactual introduces an unrecoverable scaling factor in the firm payoff function. For computational tractability, I restrict the set of potential links to a set of links below a particular distance threshold (i.e. links cannot be built between cities that are too far apart, but this threshold goes far beyond nearest neighbors).