Australasian Transport Research Forum 2022 Proceedings
28-30 September, Adelaide, Australia
Publication website: http://www.atrf.info

Co-evolution of Sydney's public transport access and demand

Hema Rayaprolu, Bahman Lahoorpoor, and David Levinson
School of Civil Engineering, The University of Sydney, Australia
Email for correspondence: hema.rayaprolu@sydney.edu.au

1. Introduction

Throughout the history of Sydney's public transport, the dynamics of supply and demand underwent a myriad of changes. Today, we can investigate their co-evolution in detail thanks to powerful data processing and computation tools. Taking this opportunity, we tracked the evolution of Sydney's public transport network and patronage over the entire period of its existence (1855 onward) and examined the causal dynamics between supply and demand. When we have two time-series, we can study the causal direction between them by testing whether lagged information on one significantly improves the predictions of the other and vice-versa. Such tests were developed by Granger (1969) and are widely known as Granger causality tests. In this paper, we explore Granger causality between Sydney's public transport network and patronage. The subsequent sections describe the time-series network and patronage data, elaborate on the models investigated, and discuss our findings.

2. Data and Method

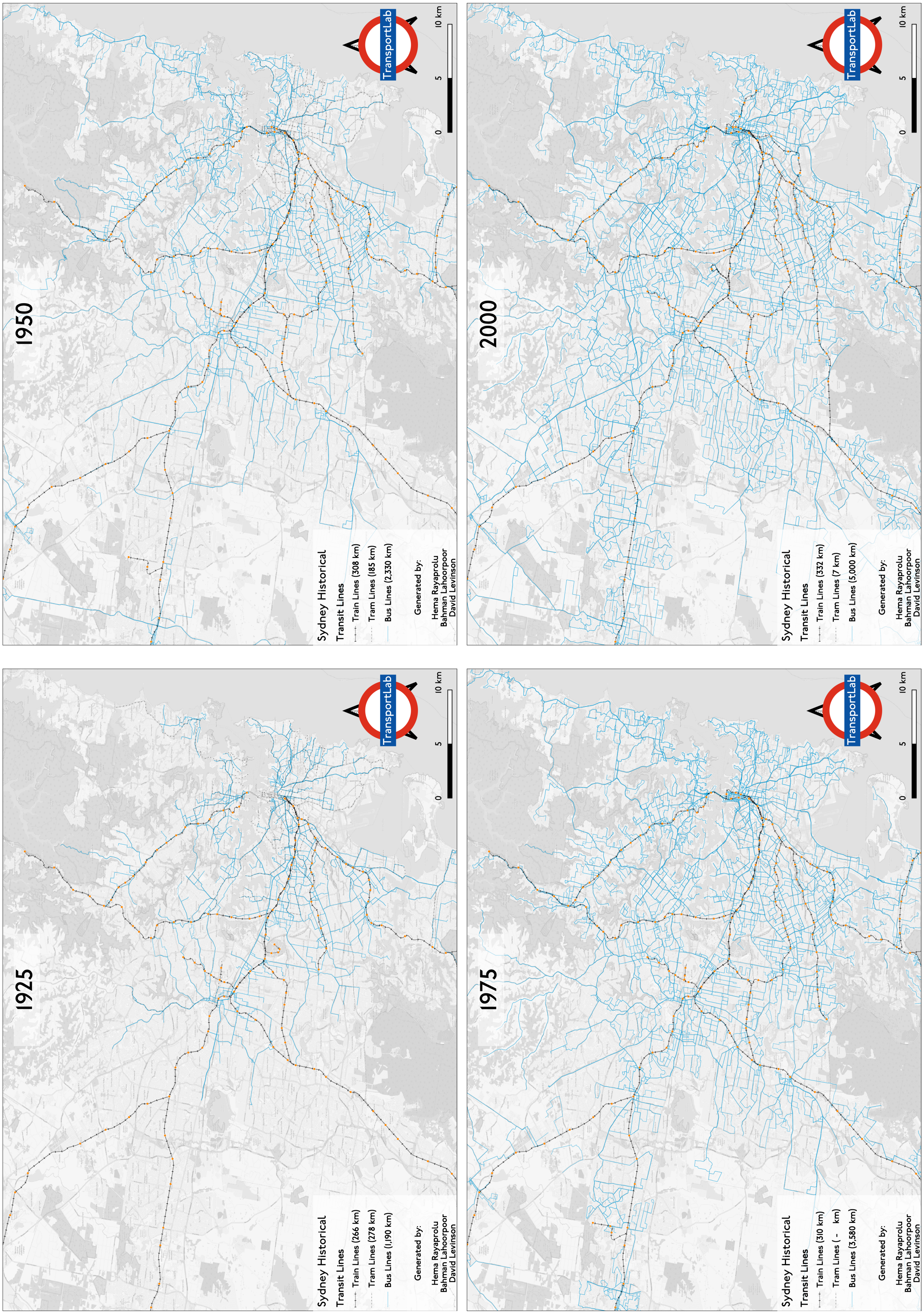
To examine the co-evolution, we collected and processed historical network and demand data. While historical public transport ridership numbers were readily available, historical networks had to be digitised in order to quantify the extent and service offered.

2.1. Historical public transport network and access

We digitised the historical train (1855 onwards), tram (1861 onwards) and bus networks (1925 onwards) based on archived maps and text on their routes, and generated General Transit Feed Specification (GTFS) files for each year from 1855 to 2015. Snapshots of the network in 1925, 1950, 1975 and 2000 are shown in Figure 1. More details on the digitisation and GTFS generation can be found in Rayaprolu and Levinson (2021) and Lahoorpoor and Levinson (2022). We did not generate networks for years beyond 2015 because GTFS have been archived and are available from Transport for NSW (2021) and Open Mobility Data (2021). However, the generated train service levels could not be matched with those published as historical information about intercity, express and skip-stop services and timetables was unavailable. Consequently, our analysis is restricted to the period between 1855 and 2014.

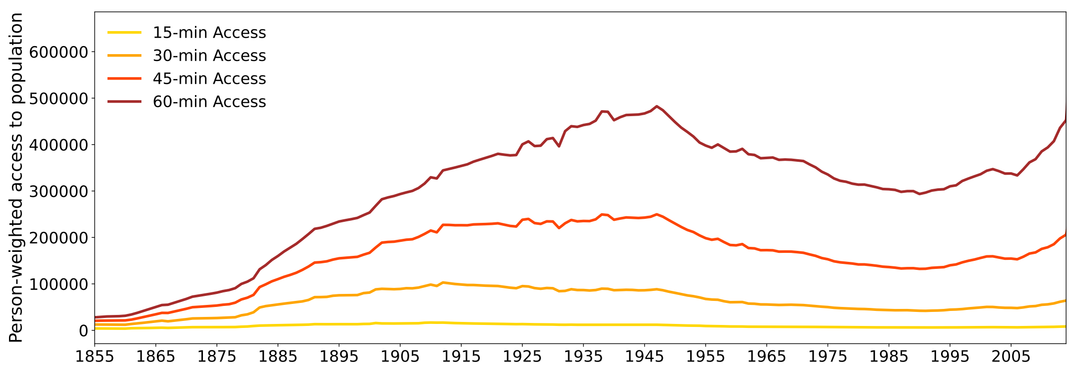
We quantified the networks generated for each year in terms of 'access' as it integrates land use with transport. We define access as the cumulative number of opportunities (people, jobs, shops, schools, hospitals, etc.) that can be reached by using public transport within a certain amount of time. Jobs are traditionally used to measure access as commuting to and from work tends to form the majority of daily trips. However, as spatially disaggregated employment data was unavailable

Figure 1: Snapshots of the Sydney's public transport networks in 1925, 1950, 1975 and 2000 as digitised by the authors



historically, we measured access to people instead. To do this, we collated population counts from censuses starting 1851, standardized geographical boundaries, assigned population to mesh blocks, and interpolated population for years between census years (Lahoorpoor and Levinson, 2021). With population counts by mesh block and GTFS for each year, we estimated the population that could be reached within 15, 30, 45 and 60 minutes by public transport for each year. To do this, we queried travel time isochrones for each mesh block at 15, 30, 45 and 60 minute cutoffs using Open Trip Planner (OTP) (OpenTripPlanner, 2021). OTP builds a network graph based on the supplied street network and GTFS, and computes point-to-point travel times and isochrones for a given point, time cutoff, mode, day and time. We supplied OTP the GTFS generated, and queried isochrones for each mesh block centroid for 8:00 am on a Wednesday. The street network remained unchanged (downloaded from OpenStreetMap (OpenStreetMap contributors, 2021) in July 2020) as historical street networks are unavailable. The isochrones obtained were overlapped with population in the corresponding year to estimate access. Population within a mesh block was assigned to its centroid for computational ease. The access values computed for mesh blocks were then aggregated as a person-weighted average for the entire Greater Sydney region. The evolution of the so-obtained 15, 30, 45 and 60 minute public transport access is shown in Figure 2.

Figure 2: Evolution of Sydney's public transport access: 1855 to 2014



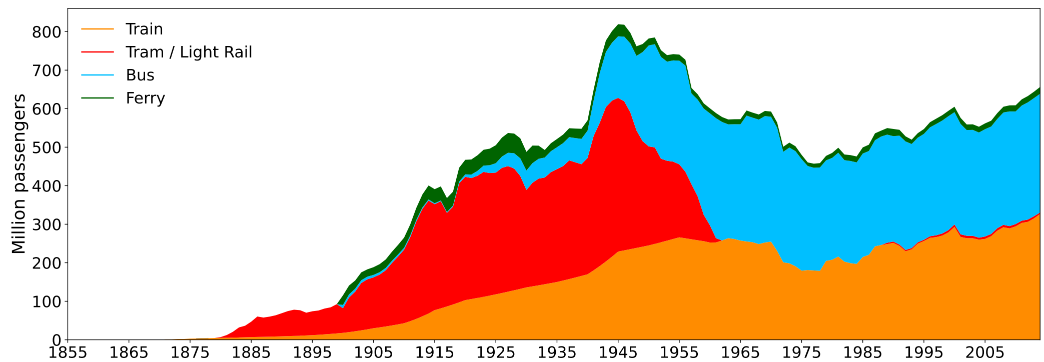
2.2. Historical public transport patronage

Annual public transport ridership has been reported by (Department of Infrastructure, Transport, Regional Development and Communications, 2014) from 1871 onward in terms of annual number of passengers using the different modes – train, tram or light rail, bus and ferry. The evolution of annual public transport ridership by mode in the Greater Sydney region is presented in Figure 3. As demand data is only available from 1871 onward, our study is restricted to the period between 1871 and 2014.

2.3. Modelling causality

With time-series data on public transport network (access) and ridership (demand), we estimated models for demand and access with lagged variables of each other to test whether lagged information on access improves predictions of demand and vice-versa. We tested different lags and found a lag of one year to work best. The model specifications for the total period (1871 to 2014) are

Figure 3: Evolution of Sydney's public transport patronage: 1871 to 2014. Source: Department of Infrastructure, Transport, Regional Development and Communications, 2014



described in Equations 1 and 2.

$$D_t = \alpha_0 + \alpha_1 D_{t-1} + \alpha_2 \Delta A_{(t-1)-(t-2)} \tag{1}$$

$$A_t = \beta_0 + \beta_1 A_{t-1} + \beta_2 \Delta D_{(t-1)-(t-2)} \tag{2}$$

In the equations, t indicated the time period (year), D indicates demand measured as annual passengers, A indicates net public transport access measured as the difference between average public transport access to population (calculated as described in Section 2.1) and average walk access to population. Net public transport access was considered to filter out the impact of the access and egress mode on the measured public transport access. Intercepts and variable coefficients are represented by α_0, β_0, and α_1, α_2, β_1, β_2 respectively and their estimations are discussed in section 3.

In addition to testing causality, we attempted to model the co-evolution of access and demand with additional control variables by temporally segregating the data into four different time periods. The time periods considered were –Pre-WWI (1871 to 1913), Inter-war (1919 to 1938), Post-war demand decline (1946 to 1978), and Recent demand growth (1979 to 2014). In defining the time-periods, we dropped the years of WWI and WWII to exclude abnormalities during war times. For control variables, we explored overall population, population within 400 metres of public transport stops, vehicle ownership (obtained from Department of Infrastructure, Transport, Regional Development and Communications (2014)) and Gross Domestic Product (GDP) (obtained from Australian Bureau of Statistics (2022)). However, only vehicle ownership and Gross Domestic Product (GDP) were found to be significant.

3. Results and Discussion

The results of the causal models for the total period from 1871 to 2014 are presented in Table 1.

The left side of Table 1 shows the results of the model estimating demand in a time period as a function of itself in the previous time period and the change in access in the previous time period. And that on the right shows the results of the model estimating access in a time period as a function of itself in the previous time period and the change in demand in the previous time period. We estimated models with cumulative 15-min, 30-min, 45-min, 60-min access and partitioned 0 to 15-min, 15 to 30-min, 30 to 45-min and 45 to 60-min access. For the demand model, lagged change

Table 1: Estimates of demand and access causal models for the total study period (1871 to 2014)

$Y = ln(D)_t$					$Y = ln(A_{45})_t$				
X	Coeff.	S.E.	p	R^2	X	Coeff.	S.E.	p	R^2
I	0.848	0.110	0.000		I	0.449	0.066	0.000	
$ln(D)_{t-1}$	0.958	0.006	0.000	0.997	$ln(A_{45})_{t-1}$	0.965	0.005	0.000	0.997
$\Delta ln(A_{30})_{(t-1)-(t-2)}$	0.240	0.056	0.000		$\Delta ln(D)_{(t-1)-(t-2)}$	0.165	0.037	0.000	

D = Public transport demand, A_n = Net n-minute public transport access to population

in 30-min net access to population resulted in the best model. On the access side, 45-min net access to population was best explained by lagged change in demand. In both models, logarithmic transformations of demand and access gave the most significant results. The coefficients of both lagged change in access in the demand model and lagged change in demand in the access model are positive and significant, indicating an overall mutual causal relationship between public transport demand and access in the region.

To explore the dynamics of this relationship during different time-period within the total time series, we segregated the total period into four different time periods as described in Section 2.3 and estimated the co-evolution of demand and access. The results for the four time-periods are presented in Table 2.

In all four time periods, logarithmic transformations of demand and access resulted in the best models. Interestingly, lagged access explained demand better than lagged change in access in the periods post-WWI. In the post-war time periods, the co-evolution was best explained by including vehicle ownership and GDP as control variables. The access thresholds that resulted in the best models varied across models during all time periods. It is hard to speculate why a specific threshold would have a greater influence. A gravity-style access measure combining different would remove this ambiguity, however, such a measure would be difficult to estimate and interpret.

During the pre-WWI period, we observed a mutual relationship. Both lagged access and lagged demand were significant in predicting demand and access respectively. The relationship during the inter-war period was, however, not significant. Nevertheless, at a 90% significance level, access can be observed to cause demand. However, the reverse is not true. This could be explained by the economic difficulties characteristic of this period between the two world wars due to which demand could perhaps not be met sufficiently by investments in public transport.

Post WWII, there was a period of decline in public transport demand as seen in Figure 3. This decline period also coincides with the increasing uptake of private vehicle ownership. On including vehicles-registered-per-capita in the models for this decline period, we found a mutual relationship between demand and access, albeit less significant than the pre-war period. What is also interesting during this period is the influence vehicle ownership has on demand and access individually. The positive coefficient in the demand model indicates increase in public transport demand as vehicle ownership increases – perhaps an artefact of prosperity and sprawl leading to greater travel distances which could translate into increased annual public transport trips. While the relationship is reversed for access. Although private vehicles would not have a direct impact on public transport access, again considering vehicles as an indicator of sprawl and reduction in population density could explain the reduction in public transport access to population.

During the recent growth period post-1979, we again observed a weaker mutual relationship between demand and access to population. Controlling for GDP gave the best results. Increase in

Table 2: Estimates of temporally-segregated demand and access co-evolution models

Pre-WWI period: 1871 to 1913 (N=41)									
$Y = ln(D)_t$					$Y = ln(A_{60})_t$				
X	Coeff.	S.E.	p	R^2	X	Coeff.	S.E.	p	R^2
I	0.812	0.273	0.005	0.992	I	0.443	0.134	0.002	0.995
$ln(D)_{t-1}$	0.960	0.015	0.000		$ln(A_{60})_{t-1}$	0.966	0.012	0.000	
$\Delta ln(A_{30})_{(t-1)-(t-2)}$	0.247	0.101	0.020		$\Delta ln(D)_{(t-1)-(t-2)}$	0.151	0.065	0.025	
Inter-war period: 1919 to 1938 (N=20)									
$Y = ln(D)_t$					$Y = ln(A_{45})_t$				
X	Coeff.	S.E.	p	R^2	X	Coeff.	S.E.	p	R^2
I	5.549	1.838	0.008	0.813	I	-0.118	1.146	0.919	0.889
$ln(D)_{t-1}$	0.5144	0.090	0.000		$ln(A_{45})_{t-1}$	1.010	0.091	0.000	
$\Delta ln(A_{45})_{(t-1)}$	0.348	0.176	0.064		$\Delta ln(D)_{(t-1)-(t-2)}$	0.211	0.227	0.364	
Post-WWII-decline period: 1946 to 1978 (N=33)									
$Y = ln(D)_t$					$Y = ln(A_{30})_t$				
X	Coeff.	S.E.	p	R^2	X	Coeff.	S.E.	p	R^2
I	1.261	1.071	0.249	0.973	I	0.416	0.224	0.074	0.991
$ln(D)_{t-1}$	0.798	0.112	0.000		$ln(A_{30})_{t-1}$	0.961	0.021	0.000	
$\Delta ln(A_{45})_{(t-1)}$	0.236	0.114	0.047		$\Delta ln(D)_{(t-1)-(t-2)}$	0.232	0.129	0.083	
$\Delta ln(V)_{(t-1)-(t-2)}$	1.945	1.055	0.076		$\Delta ln(V)_{(t-1)-(t-2)}$	-1.588	0.829	0.065	
Recent-growth period: 1979 to 2014 (N=36)									
$Y = ln(D)_t$					$Y = ln(A_{30})_t$				
X	Coeff.	S.E.	p	R^2	X	Coeff.	S.E.	p	R^2
I	5.117	2.008	0.016	0.944	I	-0.854	0.344	0.018	0.982
$ln(D)_{t-1}$	0.698	0.121	0.000		$ln(A_{30})_{t-1}$	1.010	0.048	0.000	
$\Delta ln(A_{30})_{(t-1)}$	0.314	0.178	0.088		$\Delta ln(D)_{(t-1)-(t-2)}$	0.292	0.145	0.052	
$\Delta ln(G)_{(t-1)}$	0.088	0.045	0.059		$\Delta ln(G)_{(t-1)}$	0.069	0.028	0.021	

D = Public transport demand, A_n = Net n-minute public transport access to population, V = Vehicles-per-capita, G = GDP-per-capita

GDP was seen to increase both demand and access, as one would expect. The results during this period could be understood as a consequence of peak-sprawl and high-density developments along public transport corridors.

While we are able to draw broad conclusions about the causal relationships between public transport demand and access from this analysis, we would need spatially disaggregated information on demand and other control variables to make specific claims about their co-evolution. We are extending this study to panel analyses with spatially detailed patronage data for Sydney's rail network as a future step. Another direction we are exploring is to investigate access to jobs in addition to access to population at least in the recent growth period.

References

Australian Bureau of Statistics. 2022. *Australian National Accounts: National Income, Expenditure and Product*. `https://www.abs.gov.au/statistics/economy/national-`

accounts/australian-national-accounts-national-income-expenditure-and-product/latest-release.
Department of Infrastructure, Transport, Regional Development and Communications. 2014. *Long-term trends in urban public transport*. https://www.bitre.gov.au/publications/2014/is_060.
Granger, Clive WJ. 1969. "Investigating causal relations by econometric models and cross-spectral methods". In: *Econometrica: journal of the Econometric Society*, pp. 424–438.
Lahoorpoor, Bahman and David Levinson. 2021. *An empirical model of land use and public transit co-developments in Sydney*. Working Paper 06. The University of Sydney.
— 2022. "In Search of Lost Trams: Comparing 1925 and 2020 Transit Isochrones in Sydney". In: *Findings*, p. 33040.
Open Mobility Data. 2021. *Greater Sydney GTFS*. https://transitfeeds.com/p/transport-for-nsw/237. Accessed: 2021-07-14.
OpenStreetMap contributors. 2021. *Planet dump retrieved from https://planet.osm.org*. https://www.openstreetmap.org.
OpenTripPlanner. 2021. *Multimodal Trip Planning*. https://www.opentripplanner.org. Accessed: 2021-07-22.
Rayaprolu, Hema and David Levinson. 2021. *Evolution of Sydney's Bus Network: 1925 to 2020*. Presented at Australasian Transport Research Forum 2021.
Transport for NSW. 2021. *TfNSW Open Data Hub and Developed Portal*. https://opendata.transport.nsw.gov.au. Accessed: 2021-07-14.