

Scoring and Classifying Regions via Multimodal Transportation Networks

Aaron Bramson^{*1,2,3,4} Megumi Hori^{*1} Zha Bingran^{*1} Hirohisa Inamoto^{*1}

^{*1} GA Technologies Inc.

^{*2} RIKEN Center for Biosystems Dynamics Research

^{*3} Ghent University

^{*4} University of North Carolina at Charlotte

In order to better understand the role of transportation convenience in location preferences, as well as to uncover transportation system patterns that span multiple modes of transportation, we score geographic regions according to properties of their multimodal transportation networks. The various scores are then used to classify regions by their dominant mode of transportation, and rank/cluster regions by their transportation features. Specifically, we analyze the train, bus, and road networks of major cities and neighborhoods of Japan to classify them as being train-centric, bus-centric, or car-centric. We also generate scores based on various transportation features to rank cities by their access to public transportation and to categorize/cluster neighborhoods of major cities by their transportation and accessibility properties. We find that business hubs (having low populations) are conveniently reachable via public transportation but vary greatly in their automobile accessibility. Suburban regions have lower connectivity overall but are typically strongly connected to at least one business area. As increasingly rural areas rely more strongly on the road and bus networks, but the network features do not correlate highly with population density.

1. Introduction

Transportation networks can be considered multi-graphs or multilayer networks insofar as there are links of different types connecting nodes representing locations. However, they are also fundamentally geographically embedded which constrains the network structure and requires the inclusion of continuous distance and time weights in discrete network measures. This fusion of network and geographic metrics offers the opportunity to augment network similarity measures as well as fill crucial data gaps about transportation efficiency, accessibility, connectivity, and policies.

2. Data

The geographic foundation of our analysis is a 54,127m² (125m inner radius) hexagonal grid covering all of Japan. This is used to define locations as the centers of each hex using GoogleMap's coordinates of Tokyo Station (139.7649361E, 35.6812405N) as a fixed reference point. In order to compare cities and regions within cities, we define a region as all hexes with centroids within 20 km of a selected point. We chose a variety of points across the Tokyo, Kyoto, and Osaka Metropolitan areas to capture a diversity of situations (city centers, suburban bed towns, rural areas, etc.).

2.1 Network Data

We utilize four interwoven networks representing distinct modes of transportation: train/subway, bus/streetcar, road, and walking. The train/subway network represents stations as nodes and train routes as links. In this way, express trains that skip stations are captured by links directly connecting the stations used by that route. The bus network is similarly constructed

among bus stops. Our road network is constructed via OpenStreetMaps in which the nodes are intersections and links are road segments; both restricted to roads tagged as *tertiary* or above.

In addition to these networks we include a “walking network”. This walking network connects each node of the train network to (1) the closest location of our hex grid as well as (2) any location within 500m of each station. It also connects each bus station to the (1) closest location (2) any location within 200m, and (3) any train station within 200m (when both train and bus networks are included). The third type of link represents a transfer from train to bus. The walking network also connects the nodes of the road network to each location of the hex grid. Finally, we create walking links to convert the location hex grid into a regular $k=6$ lattice network to allow (slow) transit on foot where no other mode of transportation is available. This walking network is included in all analyses because it is necessary to connect each of the transportation networks to the geographic foundation.

For each link we include a weight equaling the traversal time. For the train and bus data this is set from respective schedules using the average traversal time for that link for that type of train/bus (e.g., local, express). For the road network we calculate the traversal time based on the length of the road segment and the official speed limit (i.e., not considering traffic congestion or actual speeds). For the walking network we assume an average speed of 4kph (15 minutes per km). This slower-than-average speed is used to account for congestion as well indirect walking routes.

In addition to the travel times, we also incorporate a transfer time where appropriate to account for both moving from one platform to another as well as the waiting time for the next train/bus/taxi/etc. Specifically, we add 5 minutes when switching between trains of different lines or types at the same station, and 3 minutes for switching modes: (train ↔ bus, train ↔ road, or bus ↔ road (walking time is already included in the walking link connecting stations, bus stops and intersections).

Contact: Aaron Bramson, GA Technologies, Roppongi Grand Tower 40F, Roppongi 3-2-1, Minato-ku, Tokyo, 106-6290, JAPAN, 03-6230-9180, bramson@ga-tech.co.jp

Although only a rough approximation of the interstitial time gap, it sufficiently summarizes the variance across locations, times of day, walking speeds, congestion conditions, etc. without adding unnecessary complication to the network model.

2.2 Demographic Data

In order to assess practical (versus potential) accessibility we incorporate data regarding the population distribution into our analysis. We take 250m² square grid population data obtained from [eStat2018] using grid coordinates from [geoSpatial2018]. Then we resample it to our hex grid using overlap proportions to interpolate the hex populations.

3. Methods

To compare neighborhoods within a city we collect the locations within a 5km radius of multiple secondary and tertiary city centers (these regions overlap). We isolate the transportation networks to within the region of analysis and apply scoring methods to the individual and combined transportation networks. Our most basic evaluation utilizes standard network measures such as diameter, eccentricity profiles, and betweenness profiles along with their time/distance weighted versions. Additionally, we will include specifically geographic and transportation-focused measures such as the profile of times to travel to each regional location, a profile of the number of people reachable within 5, 10, 15, 20 minutes, and the population weighted load on the transportation network to reach the region center.

3.1 Network Measures

To start we calculate several standard network measures (mean degree, mean betweenness, mean eigenvector centrality, mean eccentricity, diameter, clustering coefficient, alpha and beta indices, etc.) on of the following transportation networks: train+walk, bus+walk, road+walk, train+bus+walk, and train+bus+road+walk. We do this for each of several focal areas within Tokyo, Kyoto, and Osaka. This battery of tests allows us to examine both differences in transportation networks for each area and differences among areas for each transportation network.

For each transportation network we calculate the travel times using Dijkstra's algorithm: breadth-first summation of traversed edges' time weights. The core algorithm is augmented to handle transfer times at appropriate junctures. Isochrones are sets of locations binned by travel time, although most of our measures can and do utilize the real-valued traversal times.

3.2 Geotemporal Measures

As a basic measure of accessibility, we compute time-weighted number of hexes reachable from each hex: $\sum_j 1 / t_{ij}$ in which t_{ij} is the shortest time from hex i to hex j . Collecting the population data allows us to determine the *sociability score* of each location; that is, the number of people who can reach each location weighted by the time it takes to reach it. We simplify and generalize the measure from [Biazzo2018] to handle continuous travel time values and averaged edge traversal times. For each hex grid space i we calculate $\sum_j P_j / t_{ij}$ in which P_j is the population of grid space j and t_{ij} is again the shortest time from hex i to hex j . We furthermore include geotemporal versions of certain network measures, such as the time-weighted eccentricity



Figure 1. Isochrone map of Tokyo centered on Tokyo Station using the train+walk network. Darker colors indicate shorter travel times.

(longest, shortest-time path from the center to the periphery) and time-weighted betweenness.

3.3 Machine Learning Techniques

In addition to providing a profile of the multifaceted transportation system, the network and geotemporal measures above are also fuel for clustering and discriminant analysis. We apply an ensemble of available measures of whole-network similarity [Soundarajan2014] (*NetSimile* [Berlingerio2012], *Normalized LBD* [Richards2010], *Graphlets* [Pržulj2004]) as a basis for distance calculations in addition to standard vector-based methods. Using those distance measures we then apply an ensemble of available unsupervised learning techniques (K-means, spectral clustering, affinity propagation, agglomerative clustering, Gaussian mixture) on the regional profiles to score, cluster, and classify them.

4. Results and Conclusions

This is still a work in progress, but preliminary results reveal cities clustered into those which have a dense rail system (e.g. Tokyo), dense regions that instead rely on buses for public transportation (e.g. Kyoto), and regions with weak public transportation that require automobiles (e.g. small cities and suburbs). Although we expected these features to correlate well with population density; we instead find that other factors heavily influence the type and convenience of a transportation network; factors such as average income, percent of commercial properties, and age demographics.

Within cities we see a familiar pattern of easily accessible central regions with low populations and regions of higher population density further out, with populations again tapering down even further out. These suburban regions often have convenient public transport to the city centers, but locally require buses and/or cars for daily transportation. An analysis of

demographics on the presence of children and elderly within the household should also correlate well with a high score on car-centric transportation. These and other results create a multifaceted scoring of properties by their transportation and demographic features. Our current efforts aim to summarize and visualize these results in an intuitive and interactive way that will lead to greater insights and deeper questions.

While most applications of machine learning to transportation networks aim at traffic prediction, flow efficiency, and rerouting, we are particularly interested in identifying cities with underdeveloped public transportation systems and regions within cities with poor accessibility. Related to the latter point, we will uncover differences in regional accessibility by mode of transportation (e.g., areas that are only convenient if one has access to a car). Identifying under- and over-serviced areas can help in policy decisions including infrastructure planning and housing development. Finally, the fusion of geographic and network measures to score areas by the convenience of, and their reliance on, varying modes of transportation can inform decisions for location services (such as apartment hunting, ride sharing, and new store positioning).

4.1 Future Work

We will extend this analysis by including additional demographic and geographic data in the analysis. Our primary purpose here is scoring and clustering areas by transportation accessibility. Future work will examine the relationship between accessibility and socio-economic factors such as unemployment, income, home-ownership, household structure, age profile, crime, etc. We are also interested in identifying network community structure differences [Bohlin2014] among the transportation modes; that is, which geographic regions are considered to be parts of which neighborhoods when considering different networks. Finally, we wish to pursue question of robustness and efficiency via knockout and detour analyses. This can address response to accidents/failures, and further to identify required structural and throughput changes required to adapt to short-term passenger changes (e.g. the Olympics) and long-term demographic changes (e.g., aging population).

Finally, we are strongly interested in the impact of bicycle ride-sharing programs on transportation flow. Although these programs have long been popular in Europe and China, and bicycles usage is high across Japan, there is very little data or analysis on bicycle usage and its interaction with other transportation modes. The recent growing popularity of bicycle-sharing programs will provide additional data to foster more advanced impact studies.

References

[Berlingerio2012] Michele Berlingerio, Danai Koutra, Tina Eliassi-Rad, Christos Faloutsos. "NetSimile: A scalable approach to size-independent network similarity" *CoRR*, *Abs/1209.2684*, 2012.

[Biazzo2018] Indaco Biazzo, Bernardo Monechi, and Vittorio Loreto, Universal scores for accessibility and inequalities in urban areas, *arXiv:1810.03017v1*, 2018.

[Bohlin2014] Ludvig Bohlin, Daniel Edler, Andrea Lancichinetti, and Martin Rosvall. "Community detection and visualization of networks with the map equation framework." In *Measuring Scholarly Impact*. Springer, 2014, pp. 3-34.

[eStat2018] Official Statistics of Japan. www.e-stat.go.jp "統計データ/ 国勢調査/ 2015年/ 5次メッシュ(250mメッシュ)/ その1人口等基本集計に関する事項" (accessed December 12, 2018).

[geoSpatial2018] Association for Promotion of Infrastructure Geospatial Information Distribution. www.geospatial.jp "都道府県別 250mメッシュ" (accessed December 12, 2018).

[Richards2010] Whitman Richards and Owen Macindoe. "Decomposing Social Networks." *2010 IEEE Second International Conference on Social Computing 2010*, pp 114-119.

[Soundarajan2014] Sucheta Soundarajan, Tina Eliassi-Rad, and Brian Gallagher. "A guide to selecting a network similarity method." In *Proceedings of the 2014 SIAM International Conference on Data Mining*, pp. 1037-1045. Society for Industrial and Applied Mathematics, 2014.

[Pržulj2004] Nataša Pržulj, Derek G. Corneil, and Igor Jurisica. "Modeling interactome: scale-free or geometric?" *Bioinformatics* 20, no. 18 2004, 3508-3515.