A Holistic Evaluation of Transit Supply and Demand using Network Analysis: The TDI Framework

Anthony Sicilia
Department of Computer Science
University of Pittsburgh
anthonysicilia@pitt.edu

Alexandros Labrinidis
Department of Computer Science
University of Pittsburgh
labrinid@cs.pitt.edu

Konstantinos Pelechrinis
School of Computing & Information
University of Pittsburgh
kpele@pitt.edu

ABSTRACT

During recent years there have been several efforts from city and transportation planners, as well as, port authorities, to design multimodal transport systems, covering the needs of the population to be served. However, before designing such a system, the first step is to understand the current gaps. Does the current system meet the transit demand of the geographic area covered? If not, where are the gaps between supply and demand? To answer this question, the notion of *transit desert* has been introduced. A transit desert is an area where the supply of transit service does not meet the demand for it. While there is little ambiguity on what constitutes transit *demand*, things are more vague when it comes to transit supply. Existing efforts often define transit *supply* using *volume* metrics (e.g., number of bus stops within a pre-defined distance). However, this does not necessarily capture the quality of the transit service. In this study, we introduce a network-based transit desert index (which we call TDI) that captures not only the quantity of transit supply in an area, but also the connectivity that the transit system provides for an area within the region of interest. In particular, we define a network between areas based on the transit travel time, distance, and overall quantity of connections. We use these measures to examine two notions of transit quality: **connectivity** and **availability**. To quantify the connectivity of an area i we utilize the change observed in the second smallest eigenvalue of the Laplacian when we remove node i from the network. To quantify availability of an area i, we examine the number of routes which pass through this area as given by an underlying transit network. We further apply and showcase our approach with data from Allegheny County, Pennsylvania, USA. Finally, we discuss current limitations of TDI and how we can tackle them as part of our future research.

ACM Reference Format:
Anthony Sicilia, Alexandros Labrinidis, and Konstantinos Pelechrinis. . A Holistic Evaluation of Transit Supply and Demand using Network Analysis: The TDI Framework. In *Proceedings of Third Mining Urban Data Workshop (MUD 2018), held in conjunction with ACM KDD 2018.* ACM, New York, NY, USA, 12 pages. https://doi.org/

Third Mining Urban Data Workshop (MUD 2018), held in conjunction with ACM KDD 2018, August 20, 2018, London, UK.

1 INTRODUCTION

"A developed country is not one where the poor have cars, but one where the rich use public transit". This is what Enrique Peñalosa, the celebrated mayor of Bogota, once said. A good public transit system is needed to provide equitable access to job opportunities and upward mobility. Extensive research has shown low-income residents typically live in areas with limited transportation options, which constrains their job opportunities, upward mobility, and even access to health care services [2, 4, 8, 14, 17].

In order to be able to identify solutions to these problems, we must first have a way to quantify the extent of the transit accessibility problem. Towards this objective, the notion of *"transit desert"* has been defined after that of "food deserts". In particular, a transit desert is an area with *low* accessibility to transit and *high* dependence on transit. This definition is fairly vague, since it does not define what is low, what is high and what is considered to be close with regards to transit access. During the last years there have been efforts to measure and quantify transit deserts using information from US Census and public transit agencies.

As we will elaborate later, the majority of these efforts quantify transportation access through the number of transit stops and routes nearby. While this is intuitive and can potentially capture the reality to a large extent from a *quantity* point of view, pure access to public transit as represented through the number of transit points and routes nearby does not quantify the full image of the *quality* of the accessed service. The latter is a multifaceted notion including dimensions such as, the quality of the infrastructure (e.g., bus quality), the quality of passenger experience (e.g., overloaded buses/subway trains), the consistency of the service (e.g., operating according to publicized schedules) and of course the geographic coverage of the service (e.g., number of transit stops and routes) to name a few. However, one of the most fundamental aspects of the quality of a transportation system is the connectivity it provides over the whole coverage area. For example, a bus stop 50 feet away from your apartment may not be useful if it only takes you to specific areas (within specific time). In order to capture the multiple dimensions that compose the quality of a public transit system, including this important aspect of connectivity, we introduce in this work TDI, i.e., the *Transit Desert Index*, based on network science.

1.1 System Model and Contributions

Our method, in particular, depends on a spatially-embedded multilayer network (see Figure 1) that captures information about the transportation system under consideration. More specifically, the first layer $\mathcal{L}_T$ embeds the transportation network in the geographic area of interest, while the second layer $\mathcal{L}_G$ embeds a spatial grid over the area. By defining this network, we allow for the capture of

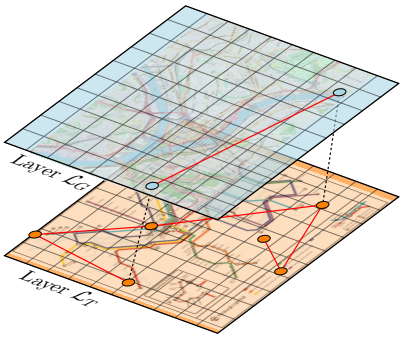

Figure 1: `TDI` is based on a spatial multi-layer network. Layer $\mathcal{L}_T$ embeds the transport network in the area under consideration, while layer $\mathcal{L}_G$ captures the connectivity between the various areas.

multiple metrics which measure quality from the perspective of both the quantity of transit stops/routes nearby *and* the connectivity of these stops, creating a more complete picture of the supply provided to a given area by a transit network. Using this network, we define κ_i the **Transit Network Availability** for every area i based on the outdegree deg^+ of the stops contained in area i. This accounts for both the number of stops and trips within a given area. We also define the **Transit Desert Eigenvalue** ϕ_i for every area i based on the eigenvalues of the Laplacian matrix for $\mathcal{L}_G$. This measures how well connected an area is with the rest of the network. By these metrics, we present a means to capture not only quantity of service, as has been previously studied, but also connectivity of service, additionally presenting a means to combine these two individual measures into a single notion of transit supply. With this, we can compare against transit demand Δ_i in order to rank areas based on their *transit gap* (between supply and demand).

Contributions: The main contributions of our work are twofold:

- We define a novel, network-based, metric `TDI` for quantifying the transit gap of an area which considers not only the "volume" of public transit in the vicinity but also the service quality in terms of connectedness. This metric acts as a generalization of previously studied versions of transit supply, allowing for the consideration of more/different notions of quality.
- We showcase the application of `TDI` using data from Allegheny County, Pennsylvania, USA (i.e., the broader Pittsburgh metro area).

The rest of the paper is organized as follows. Section 2 briefly discusses related literature and differentiates our work from the state of the art. Section 3 introduces in detail the construction of our multi-layer network, the computation of the transit desert eigenvalue, and the combination of transit supply and transit demand to form the final `TDI`. In Section 4, we apply `TDI` to the greater Pittsburgh Metropolitan Area, presenting `TDI` in view of real data. Finally Section 5 discusses future research directions and concludes our work.

2 RELATED WORK

The transit desert concept is relatively new, and hence, there are few efforts to quantify it. Jiao and Dillivan [9] calculate the difference between the level of transit dependency for an area and the amount of transit supply in the cities of Charlotte, Chicago, Cincinnati and Portland. The authors used census data to estimate the transit-dependent population. They used the number of bus/rail stops in the vicinity, the service frequency, the number of different routes, as well as the length of sidewalks and bike routes as a proxy of transit access. The authors identified transit deserts in all four cities examined. They further examined their spatial distribution, finding that they were typically concentrated close to the downtown areas. Similarly, Toms and Song [19] analyzed the gap between transit demand and supply in Jefferson County, KY, and found that most service gap areas were located in the interior of the city close to the central business district where minority and low-income populations tend to concentrate. Similar to [9], the authors developed a supply index to measure service level within each census tract, based on the bus stops and route frequency within access distance (400 m). Allen [1] provides a comprehensive review of studies in transit deserts and their relation with various demographic minorities.

One common theme of the existing literature in quantifying transit deserts is the definition of transit access through the number of bus stops, as well as the number of bus routes and their frequency, in the vicinity. While this is an intuitive proxy, it does not necessarily capture the whole picture. Our network-based approach allows us to gather further information on the quality of service in order to develop a more holistic view of the transit supply. In fact, the method of Jiao and Dillivan [9] was applied by 412 Food Rescue, a local non-profit in the City of Pittsburgh fighting food waste, in order to help quantify their impact on Allegheny County (available at [16] with corresponding paper [15]). Our results, when applied to the same county, present a different perspective via the additional consideration of connectivity; this will be discussed in Section 4. This is the main difference of our current work from existing literature. Our network-based index (described in the proceeding section) incorporates quantification of the transit-based connectivity of the various areas in the city, in addition to the previously studied notions of availability, providing a different, more holistic view of of transit supply in an area.

3 TRANSIT DESERT INDEX

In this section we introduce a network-based framework to evaluate transit supply and demand, which we quantify independently. We begin with a detailed description of the network at question and then describe its important features: connectivity and availability. We combine these two features to create a more holistic measure of transit supply. Following this, we present a definition of transit demand as seen in [9] and combine the components of supply and demand to form the `TDI` framework.

3.1 Constructing the Network

Central to all of our considerations is a spatially-embedded multi-layer network (see Figure 1) that captures information about the transportation system under consideration. More specifically, the first layer, $\mathcal{L}_T$, embeds the transportation network in the geographic

area of interest, while the second layer, $\mathcal{L}_G$, embeds a spatial grid over the area.

The underlying properties of this multi-layer network are given by the transportation network $\mathcal{L}_T = (\mathcal{V}_T, \mathcal{E}_T)$. $\mathcal{L}_T$ is simply defined as the directed network composed of all connections between transit stops: nodes consist of all transit stops within the transportation system of interest and there exists an edge $e_{ij} \in \mathcal{E}_T$ if and only if there is a transportation route passing from stop i to stop j picking up/dropping off at both stops; at this stage, parallel edges are allowed, accounting for multiple transportation routes from i to j, but this will be restricted in construction of our later embedding $\mathcal{L}_G$. Importantly, we take note of the edge attributes (τ_{ij}, d_{ij}) as they are used later in this section to compute the edge weight (velocity) between nodes at $\mathcal{L}_G$. Time, $\tau_{i,j}$, is computed from the scheduled transit stop-times within the transportation network and is given by

$$\tau_{ij} = \text{arrival time}_j - \text{departure time}_i$$

and d_{ij} is the distance between stop i and stop j given by the Haversine formula

$$d_{ij} = 2R \arcsin \sqrt{\text{hav}\left(\Delta\text{lat}\right) + \cos(\text{lat}_i)\cos(\text{lat}_j)\text{hav}(\Delta\text{lon})}$$

$$\Delta\text{lat} = \text{lat}_i - \text{lat}_j, \quad \Delta\text{lon} = \text{lon}_i - \text{lon}_j, \quad \text{hav}(\theta) = \sin^2\left(\frac{\theta}{2}\right)$$

with lat and lon referring to latitude and longitude coordinates after conversion to radians. The Haversine formula computes the great-circle, or fly-over, distance [18] under assumption of a spherical earth; in all of our calculations the radius $R = 6371.009$ km is used [11]. Additionally, we draw the reader's attention to the unweighted outdegree $deg^+(i)$ for i in $\mathcal{V}_T$. This value accounts for the number of transit scheduled trips or routes which leave a given stop i. These edge and node attributes will be important in defining our notions of quality: connectivity and availability.

From the attributes of $\mathcal{L}_T$ the second layer $\mathcal{L}_G = (\mathcal{V}_G, \mathcal{E}_G)$ can be defined. In particular, $\mathcal{L}_G$ represents an embedded grid over the area of interest with each node $v \in \mathcal{V}_G$ representing a grid cell in the area under examination. To construct this grid, derivations from the Haversine formula are used to define an equidistant partition of a bounding box for the area of interest; in the end, if a grid cell is contained within the bounding box, but its center point is not contained within the area of interest, it is excluded from consideration as a node. In all of our computation, a grid cell size of .322 km (i.e., 0.2 miles) was used.

Each $v_i \in \mathcal{V}_G$ inherits the properties of all the transit stops in $\mathcal{V}_T$ it contains, aggregating this information to allow for a targeted analysis. In particular, it inherits the paths formed by $\mathcal{E}_T$. For nodes v_i and v_j, an edge $e_{ij} \in \mathcal{E}_G$ exists if and only if there is a transportation path that connects v_i with v_j or there exists a stop within .4 km of the grid cell center of v_i or v_j from which a path can be traversed. In the latter case, we consider this extension of the transportation system to be a *walking edge*; for these edges we compute distance as before and estimate time assuming a moderate walking speed of 4.83 km (i.e., 3 miles) per hour. Note here, that given the nature of the system examined, this edge ought to be bi-directional, that is, if one can reach v_i from v_j, s/he can also reach v_j from v_i. In particular, for our later use of the Laplacian, we restrict $\mathcal{L}_G$ to be an undirected, weighted graph of shortest paths (with respect to time) in which for any $v_i, v_j \in \mathcal{V}_G$, we have that there exists a single edge. This single undirected edge e_{ij} is a representation of the shortest path with respect to time. Precisely, it is computed by taking directed, temporally shortest paths in both directions, $\mathbf{P}_{pq}, \mathbf{P}_{uw} \subset \mathcal{E}_T$ from a stop $p \in v_i$ to a stop $q \in v_j$ and $u \in v_j$ to $w \in v_i$ respectively; these can be computed using Dijkstra's algorithm, for example. Given this, the edge weight $w(e_{ij})$ is then defined to be

$$w(e_{ij}) = \max\left\{V_{ij}, V_{ji}\right\}$$

$$\text{where} \quad V_{ij} = \frac{\sum_{e_{kl} \in \mathbf{P}_{pq}} d_{kl}}{\sum_{e_{kl} \in \mathbf{P}_{pq}} \tau_{kl}} = \frac{\text{total path distance}}{\text{total path time}}$$

and V_{ji} is defined similarly. If no shortest path between v_i and v_j exists, then at least one of these grid cells contains no transit stops within the aforementioned .4 km radius; in this case we set $w(e_{ij}) = 0$, which further means the edge $e_{ij} \notin \mathcal{E}_G$. This weight represents the *velocity* of the shortest path with respect to time from v_i to v_j and will be used in analysis of the **connectivity** of our network.

Additionally, $\mathcal{L}_G$ inherits a summary of the transit stops/trips within $\mathcal{L}_T$. We define the node property κ_i for $v_i \in \mathcal{V}_G$ as the sum of the outdegree. That is, if $\mathcal{V}_{T_{v_i}}$ represents the non-empty set of stops which are contained in the grid cell, v_i, then κ_i for v_i is given by

$$\kappa_i = \sum_{j \in \mathcal{V}_{T_{v_i}}} deg^+(j)$$

This attribute plays a significant role in the proceeding section, quantifying availability.

In summary, we have the layer of interest in our analysis, $\mathcal{L}_G$, an undirected, weighted graph of shortest paths with each edge $e_{ij} \in \mathcal{E}_G$ having a connection velocity $w(e_{ij})$ and each node $v_i \in \mathcal{V}_G$ having a summary of transit stops/trips κ_i.

3.2 Defining Transit Supply

The properties of the given network $\mathcal{L}_G$ together form our notion of *transit quality*: connectivity and availability.

The availability within our network is precisely defined to be the node attribute κ_i. As discussed, this attribute takes into account all outbound trips in the underlying transit network, giving a general measure of how *available* the transit system is to a rider and acting as a network-based substitute to measure the number of nearby stops and transit routes. This value, effectively represents the number of stops contained within an area, weighted by the total number of routes which pass through this stop, providing a network-based substitution for measurement of nearby stops / routes as has been discussed by the state of the art (Section 2). In keeping with the spirit of this attribute as a measure of the *availability* of routes at a rider's *entrance* into the transit system, we say that any $v_i \in \mathcal{V}_G$ which was connected to the transit system via *walking edge*, and so by definition does not have a κ_i since it contains no transit stops, receives the κ_i of its nearest entry point into the transit system. Stops such as this will have hindered connectivity, but we recall, κ_i is a separate measure.

The notion of connectivity is more elusive than that of availability, but once $\mathcal{L}_G$ has been constructed as described, we can quantify its overall connectivity using its spectrum. In particular, we start by

calculating the Laplacian matrix of the network:

$$L_{\mathcal{L}_G} = D - A_{\mathcal{L}_G} \tag{1}$$

where $A_{\mathcal{L}_G}$ is the adjacency matrix of $\mathcal{L}_G$ and D is a diagonal matrix, with D_{ii} being equal to the weighted degree of $v_i \in \mathcal{V}_G$. The spectrum of the $L_{\mathcal{L}_G}$ includes important information with regards to the connectivity of the whole network. In particular, if the second smallest eigenvalue λ_2 of the Laplacian is 0, then the network is disconnected, while if $\lambda_2 > 0$, the network is connected [12]. If a network is connected, the magnitude of the second smallest eigenvalue of the Laplacian further informs us how easy it is to disconnect the network. In fact, λ_2, is often called the algebraic connectivity because of its relation to a spectral partitioning algorithm for graphs. Here, λ_2 is directly proportional to the minimum cut size given by a network bisection [6, 12]. Hence, in our case, with velocity as weight in our adjacency matrix, λ_2 is directly proportional to a measure of the fragility of the network in terms of this velocity. At a high level, let us consider a group of nodes with low velocity edges to the remainder of the network. These nodes are not well connected within the network since the minimum cut size required to disconnect them is lower. Hence, the algebraic connectivity of the network as a whole (which is proportional to this cut size) would be lower.

However, λ_2 gives insight on the connectivity of the network as a whole, rather than the connectivity of individual nodes. In order to get an estimate of the latter, we borrow and adopt an idea from statistical physics, namely, the cavity method [10]. In brief, the cavity method is used to obtain mean field solutions for statistical properties in lattices, low-dimensional spaces and networks. The high level idea is that, for a large enough network, removing a single node does not alter the statistical properties of the network, however, it might make the calculations easier depending on the setting. In our case, we neither have an exceedingly large network nor are we interested in a mean field solution. However, we can estimate the connectivity for every node by removing it from the network and re-calculating the eigenvalues of the Laplacian.

To illustrate, let us assume we want to estimate the connectivity of node v_i. We start by removing node v_i (and its adjacent edges) from $\mathcal{L}_G$ to obtain the *cavity network* $\mathcal{L}_{G_{-v_i}} = (\mathcal{V}_{G_{-v_i}}, \mathcal{E}_{G_{-v_i}})$. We then calculate the second smallest eigenvalue of the cavity network $\psi_2(\mathcal{L}_{G_{-v_i}})$. If $\psi_2(\mathcal{L}_{G_{-v_i}}) > \lambda_2$ the overall connectivity of the cavity network $\mathcal{L}_{G_{-v_i}}$ is better compared to that of the original network $\mathcal{L}_G$, and hence, node v_i is not well connected in $\mathcal{L}_G$. Comparably, if $\lambda_2 > \psi_2(\mathcal{L}_{G_{-v_i}})$ the overall connectivity of the cavity network has decreased compared to that of the original network, indicating v_i is well connected in $\mathcal{L}_G$. Then, by defining the transit desert eigenvalue as

$$\phi_i = \lambda_2 - \psi_2(\mathcal{L}_{G_{-v_i}})$$

we can rank the nodes in $\mathcal{L}_G$ based on their transit connectivity. Algorithm 1 summarizes the calculations of ϕ_i.

3.3 Transit Demand

Our definition for transit demand is similar to that of [9]. In particular, we say that the transit-dependent population in a given census block

Algorithm 1 Calculate Transit Desert Eigenvalue

Input: $\mathcal{L}_G = (\mathcal{V}_{\mathcal{L}_G}, \mathcal{E}_{\mathcal{L}_G}, \mathcal{W}_{\mathcal{L}_G})$
Output: $\boldsymbol{\phi}$
  $\lambda_2 = \texttt{sort}(\boldsymbol{eigen}(L_{\mathcal{L}_G}))(2)$
  for $i \in \mathcal{V}_{\mathcal{L}_G}$ **do**
    $\psi_2(\mathcal{L}_{G_{-v_i}}) = \texttt{sort}(\boldsymbol{eigen}(L_{\mathcal{L}_{G_{-v_i}}}))(2)$
    $\boldsymbol{\phi}[i] = \lambda_2 - \psi_2(\mathcal{L}_{G_{-v_i}})$
  end for
  return $\boldsymbol{\phi}$

Δ'_B is given by:

household drivers =
(population age ≥ 16) − (persons living in group quarters) (2)

Transit-household population =
(household drivers) − (vehicles available) (3)

Δ'_B = (transit-household population) + (population age 12-15)
+ (non-instituionalized population living in group quarters) (4)

In the above, as done by [9], we restrict to $\Delta'_B \geq 0$, to avoid negative transit dependent population. To map this value to our network, we say that given a census block B and a set $\mathcal{V}_{G_B}$ consisting of grid cells which center-points lie inside B, the transit dependent population for each grid cell $v_i \in \mathcal{V}_{G_B}$ is given by

$$\Delta_i = \frac{\Delta'_B}{|\mathcal{V}_{G_B}|}$$

evenly distributing the population amongst the equal area grid cells.

3.4 Synthesizing TDI

We compare supply versus demand in a similar fashion as has been done before [9], namely by standardizing and taking differences to determine a gap. We present a notable distinction in allowing for a generalizable combination function ς in order to allow for versatility in combining our different notions of transit quality.

Given that within our network we must compose multiple metrics with different units and potentially large difference in distribution, we first apply the common standard score, or z-score, to each measurement. E.g, for v_i with transit dependent population Δ_i the standard score is $z_{\Delta_i} = \frac{\Delta_i - \mu_\Delta}{\sigma_\Delta}$ where μ_Δ is the mean and σ_Δ is the standard deviation amongst the transit dependent population for all of $v_i \in \mathcal{V}_G$ which have non-empty set of edges $\mathcal{E}_{G_{v_i}}$, effectively restricting our consideration to grid cells which are actually connected to the transit network. Those not connected are handled otherwise as discussed later in this subsection. This standard score is identical in the cases of availability κ_i and connectivity ϕ_i giving z_{κ_i} and z_{ϕ_i} for $v_i \in \mathcal{V}_G$. Effectively, this standard score measures the number of standard deviations with which our measurement of interest differs from the mean across the network. We do not require assumptions of normality in our distributions when we take this standard score, but with that said, by comparing the values of different distributions by their distance from the mean μ, we place a high degree of weight on μ as a summary statistic. As we will see in our application of

TDI, it may be necessary to transform the given data in order to accommodate this.

Using these standard scores, we combine our two components of supply by allowing for an arbitrary combination function $\varsigma(z_{\kappa_i}, z_{\phi_i})$. The choice of this combination function acts as a parameter to the system, chosen by the analyst. It is an important abstraction within the definition of the index because it allows a significant degree of generalization. Notably, this allows for extension of TDI to a larger number of measures say $\{z_1, ..., z_n\}$ or a completely different choice of measures than those examined in this paper. Most importantly, it allows for versatility in the combination of different notions of quality. For example, it may often be the case, in analyzing quality of a given transit network, that the metrics used are at odds. In this case, for standardized metrics $\{z_1, ..., z_n\}$ one might be interested in best case performance which would be given by $\max\{z_1, ..., z_n\}$. In other cases, where each metric holds varied importance in tandem, a wise choice may be the convex combination

$$\varsigma(z_1, ..., z_n) = \sum_i w_i z_i \quad \text{s.t.} \quad \sum_i w = 1 \quad \text{and} \quad w_i \geq 0 \quad \forall i$$

which acts as a weighted average; e.g. taking $w_i = \frac{1}{n}$ would give $\varsigma(z_1, ..., z_n) = \bar{z}$. In [9], such an average was used to aggregate their four volumetric notions of supply. Our notion of supply requires generalization beyond averaging because we combine metrics which measure completely different *aspects* of quality; the importance of each, and required combination of each, may vary depending on the use case. In our own application, we present and compare a variety of the proposed combination functions as well as isolate each proposed metric's contribution to supply in order to better illustrate the individual aspects of quality that each measures.

With supply given by ς to combine our network-based measures of quality, we define TDI for grid cell $v_i \in \mathcal{V}_G$ by

$$\texttt{TDI}_i = \begin{cases} z_{\Delta_i} - \varsigma(z_{\kappa_i}, z_{\phi_i}), & \text{if} \quad z_{\Delta_i} > \varsigma(z_{\kappa_i}, z_{\phi_i}) \\ 0, & \text{else} \end{cases} \tag{5}$$

The implicit filtration in this case-based definition says that any v_i such that supply is greater than demand is not a transit desert; i.e. it has index 0. All other grid cells $v_i \in \mathcal{V}_G$ with non-empty set of edges $\mathcal{E}_{G_{v_i}}$ are given a positive index which increases the more severe the transit gap is. In the case of grid cells which are contained within the area of interest, but for which TDI cannot be calculated – i.e., grid cells which do not contain a transit stop within a .4 km radius and as a result have no edge in $\mathcal{E}_G$ – we give the designation of *outright desert*. These nodes have truly limited access to the transit system because the distance to an "entrance" forces an inconvenience on its use. TDI is not necessarily aimed at the analysis of these grid cells because their desert status is clear, rather TDI focuses on quantifying the desert status of grid cells which are connected to the transit system, but may still not be well served. In essence, $\texttt{TDI}_i$ takes exactly a comparison of supply and demand, as others have done before, but combines sophisticated components in a tunable fashion, in order to holistically represent the quality of the transit network in question. The interested reader is invited to the Appendix A Figures 4 and 5 where an illustrative example of the entire calculation of TDI is performed.

4 APPLICATION AND RESULTS

In this section we apply the proposed method to Allegheny County, Pennsylvania, home to the city of Pittsburgh. We first describe the data used, give details on the preprocessing necessary for analysis, and discuss some general findings on the features that compose TDI. We then begin our application of TDI to determine transit gaps. We first isolate each proposed metric as a source of transit supply to better illustrate what notions of quality these metrics capture and to empirically motivate the need for the combination of these metrics. We then explore a number of combination functions ς, illustrating how our framework can be tuned to magnify (or balance) the desired notions of quality.

4.1 Data and Preprocessing

In applying the proposed method to Allegheny County, all population data for census blocks is given by American Fact Finder [3] and all geographic information for census blocks is given by Allegheny County GIS Open Data [5]. The transit information used is provided by the Port Authority of Allegheny County [13].

In this application of TDI, slight modifications to the general procedures of Section 3 were needed. Adequate population data could not be found for age groups 12-15, only for age groups 10-14, so in following other organizations working within Allegheny County [15], we modify the equation for transit demand Δ' by this substitution. Additionally, for data used to construct $\mathcal{L}_T$ (see Section 3.1), the given arrival and departure times are specified only to the minute. To prevent cases where a trip takes zero time (e.g., arrival time is equal to the departure time), we modify by adding 30 seconds to travel time.

Lastly, we remark on some steps in preprocessing. As mentioned in Section 3.4, in using the standard score as a measure to compare difference from the mean across varying distributions, we require that the mean be an accurate representation of the data to some extent. For roughly symmetric distributions this is the case, as the median and mode should also cluster near it, so whenever our distribution is highly skewed we perform a natural log transformation to achieve better symmetry [7]. In particular, before taking the standard score, we apply the log transformation to our measure of availability κ_i as well as shifting by a constant (to account for zero values [7]) and applying the log transformation to the transit demand Δ_i. We also note that there is a single outlier in our computed values for ϕ_i; this value is included in the computation of the standard score, but is often excluded from tables and plots (e.g. for visualization, to maintain diversity in color mappings). In these cases, it is noted in the figure caption with the outlier's value as well. We invite the reader to refer to the Appendix B (Figure 6) for a visualization of the above transformations as well as this outlier.

4.2 A First Glance at Allegheny County Data

4.2.1 Outright Transit Deserts. We begin by identifying grid cells not connected within the network $\mathcal{L}_G$. These grid cells are classified as outright transit deserts within the TDI framework. In total, outright deserts consist of 47% of the total transit dependent population and 75% of the total area in Allegheny County with the remaining nodes within our network amount to roughly 4500 grid cells. Note, these numbers correspond to a .4 km max distance to a transit stop. Higher maximum distances would imply the need

Table 1: From left to right are lists of the ten highest transit gaps among census blocks given by `TDI` when: using connectivity as supply (1), using availability as supply (2), combining connectivity and availability equally as supply (3). Ranking is given by averaging the grid cells contained in each census block. A neighborhood of reference for each census block is provided and a reference map exists in Appendix B (Figure 13). In all cases, the outlier is removed (corresponding to the St. Clair neighborhood in Mount Oliver).

Block ID	Neighborhood	`TDI` (conn.)
1.1	Business/Hill Distr.	4.408
1.2	Oakland	3.948
1.3	Southside Slopes	3.870
1.4	Oakland	3.725
1.5	Oakland/Shadyside	3.630
1.6	Southside Slopes	3.553
1.7	Oakland	3.399
1.8	Mount Oliver	3.223
1.9	Oakland	3.217
1.10	Southside Slopes/Mt. Oliver	3.102

Block ID	Neighborhood	`TDI` (avail.)
2.1	Oakland	4.222
2.2	Pitcairn	3.453
2.3.	Pleasant Hills	3.254
2.4	McKeesport	3.195
2.5	McKeesport	3.028
2.6	Brookline	2.946
2.7	Natrona Heights	2.916
2.8	Corliss / Crafton	2.895
2.9	Carnegie	2.890
2.10	North Versailles	2.872

Block ID	Neighborhood	`TDI` (eq.)
3.1	Oakland	4.085
3.2	Brookline	2.942
3.3	Squirrel Hill South	2.871
3.4	Pitcairn	2.809
3.5	Corliss/Crafton	2.667
3.6	South Hills	2.648
3.7	Business/Hill Distr.	2.608
3.8	South Side Slopes	2.579
3.9	Larimer	2.557
3.10	McKeesport	2.553

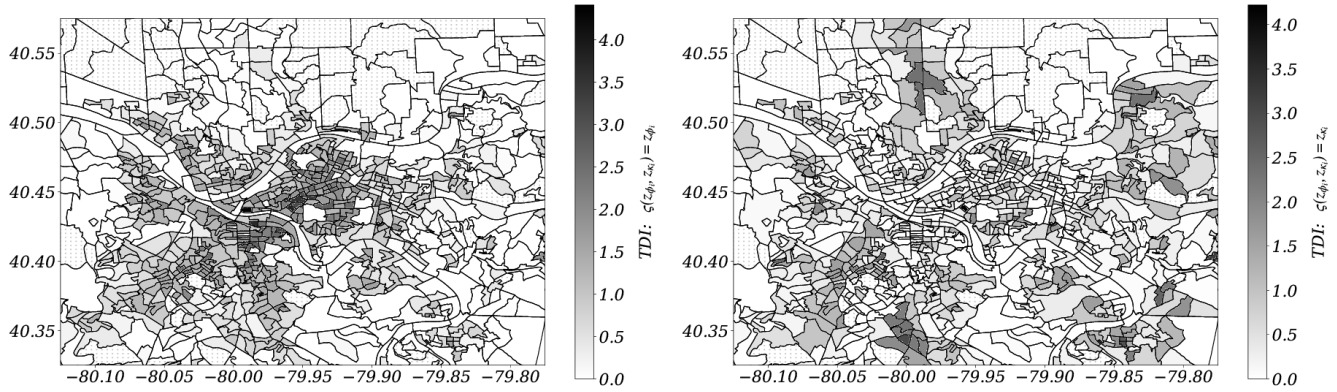

Figure 2: Application of `TDI` with isolated supply source. (Left) Connectivity z_{κ_i} as supply concentrates high `TDI` in the city, whereas (Right) availability z_{ϕ_i} as supply begins to increase `TDI` in certain outskirt areas. Values of `TDI` are averaged over non-desert grids to aggregate information at the census block level. Outright desert grids are denoted with a dot. The outlier for z_{ϕ_i} is excluded to better visualize color mappings; `TDI` for this outlier was 58.41.

to walk further to get to/from a transit stop, but will reduce the transit-dependent population/area figures. Related to these figures, we estimate that 254,924 transit-dependent persons do not have adequate access within the Allegheny County transit system; these people are located within such outright transit deserts. Appendix B contains visuals (Figure 9) and further details concerning this estimate. Also found in Appendix B is a visual comparison of outright deserts and remaining nodes (Figures 7, 8). We continue in the remainder of this section with the central focus of this work: an application of `TDI` to those nodes (grid cells) connected within the network $\mathcal{L}_G$. The primary interest of `TDI` is not to classify outright transit deserts which lack adequate access by definition, but to isolate those areas which do have access and to quantify the quality of this access.

4.2.2 Preliminary Insight into the Features which Compose `TDI`. To that end, we continue by summarizing the computed values of the features necessary to compute `TDI`. The interested reader may find accompanying visuals for this summary in Appendix B (Figures 10, 11, and 12).

Among the grid cells considered, those within the municipality of Pittsburgh have the highest demand, excluding a few cases where a census block is dominated by a wooded area or correctional facility. While it may be, as will be discussed in detail later, that these city areas are adequately serviced in terms of quantity, examination of connectivity, ϕ_i, shows relatively low values for these areas. This raises the concern that frequent local-service stops (common within the city) may contribute to congestion, making long-distance trips from these areas too inefficient. Emphasizing this point, we note that there exist many areas outside of the city limits with higher connectivity; the transit routes in these areas likely correspond to express routes into/out of the city with high velocity. For example, highly connected areas exist near the far West corner of Allegheny County. One, Moon Township, has express routes that connect the Pittsburgh Airport (28X Airport Flyer) and commuters (G3 Moon Flyer) to the Central Business District. Trips such as these make few stops after they leave the downtown area, allowing long distance coverage in a short amount of time, and hence, contributing to high connectivity.

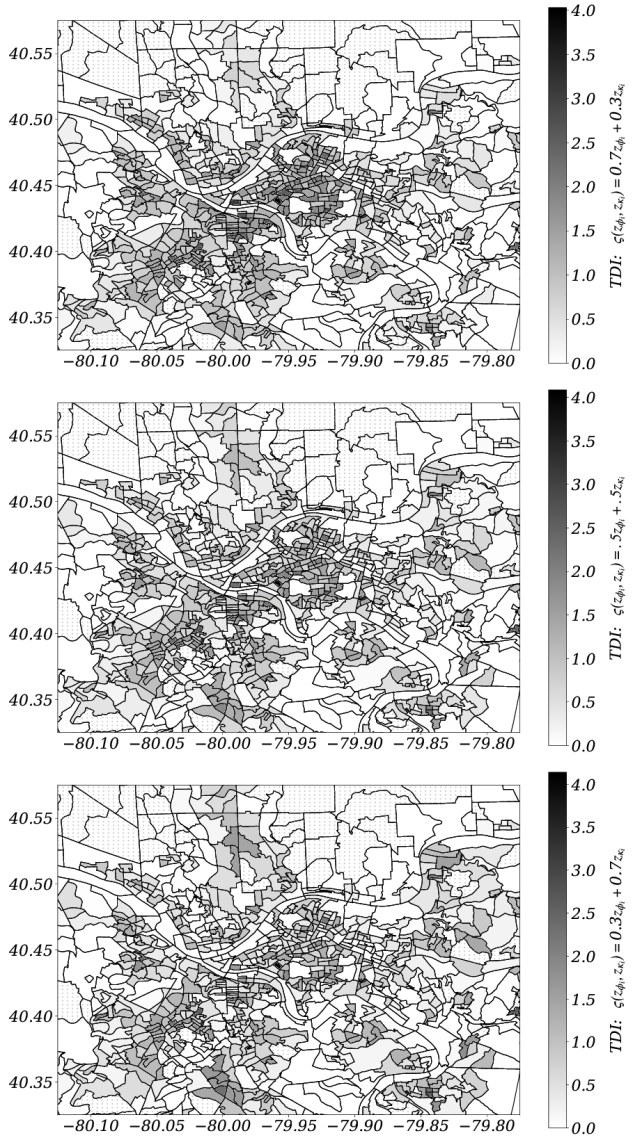

Figure 3: Rows 1-3: Application of `TDI` with weighting gradually shifted, favoring z_{ϕ_i} top-most and z_{κ_i} bottom-most. Equal weighting (middle) accompanies rankings listed in Table 1.

The notion of congestion within the municipality of Pittsburgh that connectivity addresses (in comparison to more efficient long distance trips) provides interesting information about the transit system at question, but the benefit of `TDI` is that we are capable of judging these highly congested areas from a separate point of view: *availability*. In fact, values of availability κ_i are generally higher within the municipality of Pittsburgh, contrasting those of connectivity. This will become more clear as we begin to actually apply `TDI`, but the fact that, when computed from real-world data, these two metrics do appear to capture different notions of quality is important in that it motivates the proposition that different notions of quality should be considered. Information is gained and lost when we consider one metric or the other, so one expects (as tested in the following) that these individual notions of quality used as isolated measures of transit supply will measure the transit gap differently.

4.3 Applying `TDI` with Isolated Notions of Transit Supply (Table 1, Figure 2)

In the following, we combine supply and demand in an application of `TDI`. Specifically, in this section, we pick ς such that z_{ϕ_i} and z_{κ_i} are isolated in contribution to supply (see Figure 2). We do so for two primary purposes: first, to determine differences in the notions of transit quality these individual metrics measure and, second, to conduct a more targeted comparison of these individual metrics with respect to existing methods.

4.3.1 Proposed Metrics as Isolated Sources of Supply. In isolating our proposed metrics as the source of transit supply, we observe, as expected, transit gap patterns that are inversely related to the patterns these metrics display as standalone features (see the preceding). Namely, isolating connectivity, z_{ϕ_i}, as supply, tends to give high `TDI` (i.e., a large transit gap) to areas within the municipality of Pittsburgh. This corresponds to the low-to-moderate connectivity within the city. In contrast, isolating availability, z_{κ_i}, as supply tends to give high `TDI` to certain outskirt areas. This corresponds to the strong concentration of availability within the city. We refer the reader to Table 1 giving reference neighborhoods (identified by postal code) for census blocks with the highest average `TDI`. The location of these blocks emphasize the vastly different notions of transit quality these individual metrics capture. For example, the ten areas with highest transit gap, given by the isolation of connectivity as supply, are located within the municipality of Pittsburgh, but isolating availability as supply gives only two areas within the municipality of Pittsburgh, Oakland and Brookline. The remainder lie outside this region. We note too, that Brookline is near the edge of the municipality and that Oakland is home to the University of Pittsburgh, and hence, inherits a large amount of transit-dependents in the form of non-institutionalized persons living in group quarters, i.e. students. This is likely the reason it appears frequently as a highly transit deserted area in Table 1. In any case, from Table 1, it is directly visible, at least at extrema, that our two proposed metrics capture different notions of quality. Visual examination of the notable discrepancies in Figure 2 offers additional evidence that these metrics provide a different perspective of transit deserts within Allegheny County.

4.3.2 Comparison of Proposed Metrics Against Related Work. Our metrics also provide a different perspective then is presented by the related method of [9] used by [15, 16] within Allegheny County. We proceed by providing a comparison of our metrics' transit desert rankings with the rankings of [15, 16]. Here, our metrics are individually isolated in the role of supply as in the preceding (see 2). We note, interpretation of the results in [15, 16] is predominantly visual (at the census block level) due to the lack of availability of data.

As mentioned, the Oakland Area stands as an outlier due to its high demand, so as both of our metrics (individually isolated as supply) have done, the related method [15, 16] also gives this area an

exceedingly high transit gap (between demand and supply). Besides this, when we consider the isolation of connectivity as supply, we see that we classify a number of important areas as having high transit gap which the related method does not. Specifically, one can observe visually (see Figure 2), that using connectivity as supply, we classify areas within the Central Business District and the neighboring Hill District as having moderate-to-high transit gap (Appendix B Figure 13 shows these areas grouped with those identified in Table 1). A number of these areas, for example some Hill District neighborhoods such as Bedford Dwellings and Crawford-Roberts, are not ranked as transit-deserts when using the related method. Notably, the Hill District is a community in which many residents live in poverty, serving as a reminder (see Section 1) of the importance of classifying transit deserts to improve quality of life. In other cases, where our proposed metric and the related method perceive the same area to be a transit desert, the degree of the gap differs. For example, Southside Slopes, which contains a number of areas that have very high transit gap when measured with connectivity as supply (see Table 1) have generally lower transit gap in the related work (e.g., one census block exists here with gap bigger than 3.9 and the next largest in the region is 1.28). Using connectivity as supply seems to provide a different perspective than that of our related work [15, 16], but this is to be expected as the measure of connectivity considers a different notion of transit supply then the volumetric notion, proposed in [9] and used in [15, 16].

Yet, as mentioned, we *can* still consider the notion of availability within our framework. Many areas given high transit gap when availability acts as supply are given a moderate gap by the related method as well [15, 16]. These include areas in the neighborhood of Pitcairn, McKeesport, and Carnegie (see Table 1). In Brookline, our method performs remarkably similar to the related method with only a .2 deviation difference. Visually, one also notices the two methods find additional common ground, ranking certain areas within the Central Business District as non-deserts. In contrast, we recall these same areas were ranked as deserts when connectivity acted as supply (see Figure 2). Although, we do note our notion of transit-availability does not exactly agree in all cases, for example, there is only a reduced gap in certain mentioned areas of the Hill District (while there is no gap in the related method); this is likely due to the fact that our availability metric does not consider ridership or length of walk / bike lanes in measuring volume of transit availability.

We see that our notion of availability provides a comparable alternative to that of the related work and that the notion of connectivity seems to provide a new perspective, but it is unclear which measure of quality would be more beneficial to an arbitrary rider, or even to the transit system as a whole, so in this way, isolating and comparing rankings with related work emphasizes the importance of considering multiple metrics. We see that these metrics take into account very different aspects of quality, so that their combined consideration (as will be shown in the following experiment) is necessary to more holistically evaluate poorly serviced areas.

4.4 Applying `TDI` with Combined Notions of Supply (Figure 3)

In Figure 3 we apply multiple combination functions ς given by different weighted averages. We first choose weights which favor connectivity more strongly, followed by weights which favor both connectivity and availability equally, followed by weights which favor availability more strongly. In doing so, we observe that the concentration of areas with high `TDI` gradually shifts, resembling more closely the isolation of each measure (as supply) whenever each measure is favored (see Figures 2, 3). In particular, when connectivity is favored, `TDI` gives high index predominantly to neighborhoods within the municipality of Pittsburgh. As availability is favored, areas outside of the city begin to have higher index values while locations within the city receive a lower index. Setting both weights equal provides a happy medium of comparison between these two weight schemes. Specifically, we observe areas with the highest `TDI` for this equal weight scheme (see 1). We note the Southside Slopes and Central Buisness District neighborhoods are shared with the isolation of connectivity as supply, while neighborhoods such as Pitcairn, Crafton, Mckeesport, and Brookline are shared with the isolation of availability as supply. The Oakland Area, due to its high demand, is shared with both, and some new areas including Squirrel Hill South, the South Hills, and Lariner appear (see Table 1). In displaying remnants of both isolated supply sources, the equal weight scheme provides a good reference for the transition that occurs as weights shift and illustrates the ability of our method to be tuned as desired.

5 DISCUSSION

In this study our objective has been to design a metric for quantifying access to transit in a more holistic way as compared to only considering the number of transit stops/routes in the vicinity of an area. To that end, we have utilized spectral graph theory to capture the connectivity of an area to its region through public transit, presented a network-based substitute for consideration of the number of transit stops/routes, and combined these in a generalization of transit supply into multiple notions of quality. In a demonstration of the effectiveness of our proposed method, we have applied `TDI` to the greater Pittsburgh area, effectively motivating the importance of considering multiple metrics by exhibiting their dissimilar rankings when isolated and demonstrating the capability of our proposed method to provide a versatile view of transit quality by empowering analyst choice for the combination function.

One element still missing in `TDI` is the consideration of time of the day. In particular, two areas might be well connected during most of the day, but not well connected late in the afternoon. While the transit desert eigenvalue can be extended in a fairly straightforward way to control for this through using the transit schedule, calculating the temporal profile of transit demand is more challenging. To that end one possibility is to use data from location-aware/geo-tagged social media platforms (e.g., Foursquare) to build an approximation for the temporal mobility profile between different areas. Of course, such a proxy comes with its own challenges, the most crucial being the representativeness of the platform's user population.

6 ACKNOWLEDGMENTS

This work is part of the PittSmartLiving project (https://pittsmartliving.org/) which is supported in part by National Science Foundation Award CNS-1739413. Additionally, the authors would like to thank the anonymous reviewers for their helpful suggestions.

REFERENCES

[1] Diane Jones Allen. 2017. *Lost in the Transit Desert: Race, Transit Access, and Suburban Form*. Routledge.

[2] Thomas A Arcury, John S Preisser, Wilbert M Gesler, and James M Powers. 2005. Access to transportation and health care utilization in a rural region. *The Journal of Rural Health* 21, 1 (2005), 31–38.

[3] United States Census Bureau. 2016. American Fact Finder. (2016).

[4] Karen Chapple. 2001. Time to work: Job search strategies and commute time for women on welfare in San Francisco. *Journal of Urban Affairs* 23, 2 (2001), 155–173.

[5] Allegheny County GIS Open Data. 2016. Allegheny County Census Block Groups 2016. (2016).

[6] Ulrich Elsner. 1997. Graph Partitioning – A Survey. (1997).

[7] Edward L. Fink. 2009. The FAQs on Data Transformation. *Communication Monographs* 76, 4 (2009), 379–397. https://doi.org/10.1080/03637750903310352 arXiv:https://doi.org/10.1080/03637750903310352

[8] Joe Grengs. 2010. Job accessibility and the modal mismatch in Detroit. *Journal of Transport Geography* 18, 1 (2010), 42–54.

[9] Junfeng Jiao and Maxwell Dillivan. 2013. Transit deserts: The gap between demand and supply. *Journal of Public Transportation* 16, 3 (2013), 2.

[10] Marc Mezard, Giorgio Parisi, and Miguel Angel Virasoro. 2014. *The Cavity Method*. World Scientific, 65–76.

[11] Helmut Moritz. 2000. Geodetic reference system 1980. *Journal of Geodesy* 74, 1 (2000), 128–133.

[12] Mark Newman. 2010. *Networks: an introduction*. Oxford university press.

[13] Port Authority of Allegheny County. 2018. Port Authority of Allegheny County, GTFS Data. http://www.portauthority.org/paac/CompanyInfoProjects/DeveloperResources.aspx. (2018).

[14] Paul M Ong and Douglas Miller. 2005. Spatial and transportation mismatch in Los Angeles. *Journal of Planning Education and Research* 25, 1 (2005), 43–56.

[15] 412 Food Rescue. 2018. Food Insecurity and Resource Access in Allegheny County, Pennsylvania: Using GIS to Identify High Need Communities and Assess Food Recovery and Redistribution Efficacy. (2018).

[16] 412 Food Rescue. 2018. Impact Report, April 2018. https://arcg.is/1mWbHz. (2018).

[17] Diana Silver, Jan Blustein, and Beth C Weitzman. 2012. Transportation to clinic: findings from a pilot clinic-based survey of low-income suburbanites. *Journal of immigrant and minority health* 14, 2 (2012), 350–355.

[18] Roger W Sinnott. 1984. Virtues of the Haversine. *Sky Telesc.* 68 (1984), 159.

[19] Kaitlin Toms and Wei Song. 2016. Spatial analysis of the relationship between levels of service provided by public transit and areas of high demand in Jefferson county, Kentucky. *Papers in Applied Geography* 2, 2 (2016), 147–159.

APPENDICES

A An Example of `TDI` Applied to a Small Network

We provide here an application of `TDI` on a small toy network (see Figure 4). For simplicity, and to exemplify the proposed metric which is perhaps the least intuitive, we consider only the connectivity of the network when computing supply. The edge weights in Figure 4 represent the transit velocities to our grid cells (i.e., nodes A, B, C, D, and E; onward, we will index them in this order as well). We first assume a given demand, listed below and visible (as standard scores) in Figure 5

$$A:2 \quad B:6 \quad C:5 \quad D:4 \quad E:3$$

We next must compute the weighted Laplacian of the graph in 4. We do so as given in Equation 1 which results in

$$\begin{pmatrix} 10 & -4 & -1 & -2 & -3 \\ -4 & 15 & -3 & -5 & -3 \\ -1 & -3 & 7 & -1 & -2 \\ -2 & -5 & -1 & 10 & -2 \\ -3 & -3 & -2 & -2 & 10 \end{pmatrix}$$

giving the algebraic connectivity of the whole network to be $\lambda_2 = 8.368$. When we proceed as in Algorithm 1. We remove nodes to produce the cavity network, recompute eigenvalues of the new Laplacian, and subtract these from the original algebraic connectivity λ_2 to arrive at our measure of connectivity for each individual node, given below and visible (as standard score) in Figure 5

$$A:0.821 \quad B:3.271 \quad C:-1.631 \quad D:0.862 \quad E:2.083$$

We see that the connectivity of the network improved when C was removed (evidenced by the negative value) and the connectivity of the network decreased most when B was removed. According to our heuristic, this indicates that B is well connected within the network, while C is poorly connected. In examining Figure 4 we can

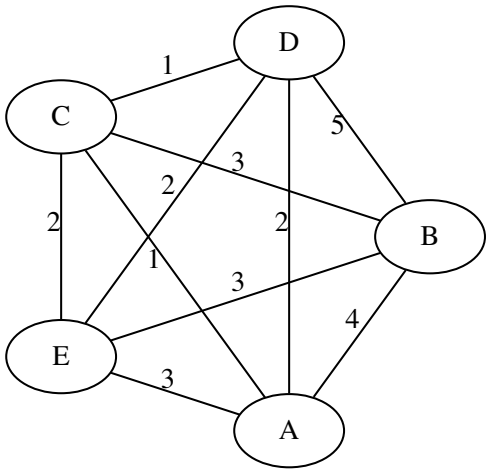

Figure 4: A sample network. Here edge labels correspond to the edge weight velocity. The demand for each node is given in Figure 5.

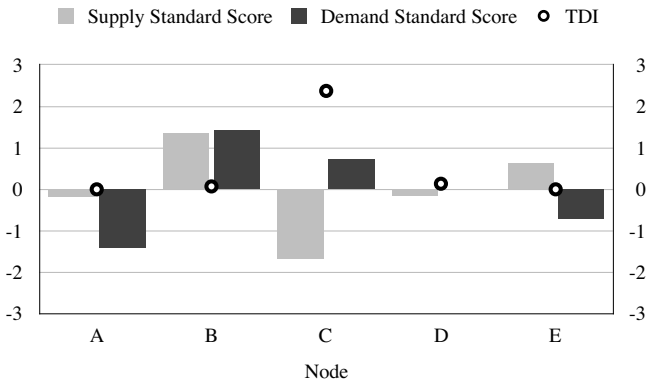

Figure 5: Displayed is supply and demand (after taking the standard score) for nodes in our sample network. Dots display resulting value of `TDI`.

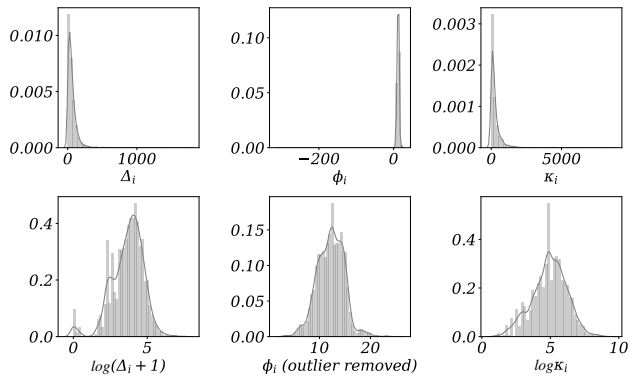

Figure 6: Δ_i and κ_i before and after log **transformation. ϕ_i with and without out lier. The constant 1 is added in the** log **transformation of Δ_i to handle 0 values [7] . The out lier for ϕ_i is included in calculation of the standard score and other analyses, but often left out for visualization purposes.**

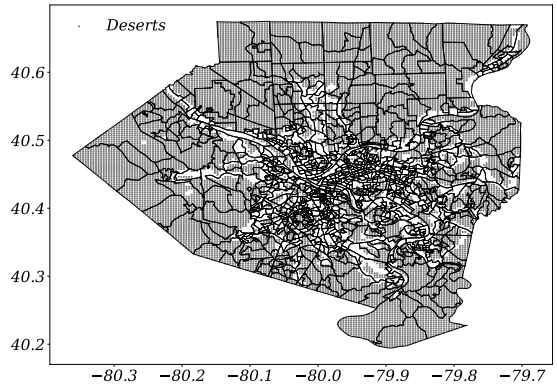

Figure 7: Grids marked as outright deserts. These grids have no transit stop within a .4 **km radius.**

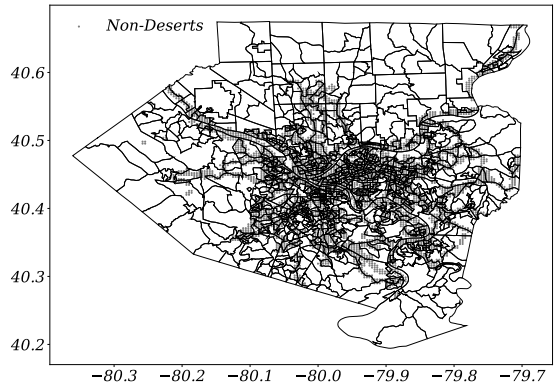

Figure 8: Grids which have an accessible transit stop. These grids have edges within $\mathcal{L}_G$ and are given an index by TDI **which is** ≥ 0

understand this by noticing C has relatively low velocity edges to the other nodes (e.g., 2, 1, 3, 1) compared to B (e.g., 5, 3, 3, 4). We recall that the algebraic connectivity is proportional to a minimum cut size in a network bisection and B's high velocity edges, in this case with a complete graph, would contribute to any cut we make.

If we take the standard scores (see Figure 5), we can now subtract supply from demand, remembering our cases as in Equation 5 to arrive at the final TDI for each node in our network. These are visible in Figure 5. We have that the low connectivity of C, compared to its above average demand, cause it to be considered a transit desert. On the other hand, B which has even higher demand is not considered a transit desert; this is because of its high connectivity. With supply as connectivity, the increased transit supply of B is able to counteract its high demand. We remark too on A, D, and E (see 5). Node A has below average demand, so because its supply is average, it is not ranked as a transit desert. Node D has average demand and average supply, so the two counteract as with B. Node E, like node A, has below average demand; its supply is above average so it is not ranked as a transit desert. This exemplifies that the predominant cases of interest here, where areas become classified as transit deserts, are those where demand far exceeds supply.

B Additional Figures From the Application of TDI to Allegheny County

Figure 6 displays all transformation done to the data before taking the standard score; this includes log transform of both the transit dependent population Δ_i and the availability κ_i. Also shown, is the distribution of connectivity ϕ_i with and without a single outlier.

Figure 7 displays all of the nodes within our network $\mathcal{L}_G$ which are classified as outright deserts and Figure 8 displays remaining nodes for which TDI can be calculated. Recall that outright transit deserts are classified as those grids which are not connected to the transit network; for these nodes TDI is not computed.

Figure 9 visualizes the total number of people who are effectively *left behind* within the Allegheny County transit system. Here, we define the number of people left behind to be

$$\#\text{ Left behind} = \Delta' \frac{|\mathcal{V}_{G_B}^{\text{OD}}|}{|\mathcal{V}_{G_B}|}$$

where $\mathcal{V}_{G_B}^{\text{OD}}$ are the grids within the block which are classified as outright deserts. Effectively, the multiplicative factor applied to demand Δ' is the percent of grids with reasonable access to transit stops, so if we assume the population to be evenly distributed amongst grids within a block, this multiplication yields the number of transit-dependent people without reasonable access to public transit. Our estimated total of transit-dependent population without access is 254,924: a telling statistic in consideration of the impact access to transit can have on quality of life (see Section 1). As mentioned earlier, this number is computed assuming a .4 km maximum distance to a transit stop. Higher max distances lead to lower transit-dependent population figures.

Figures 10, 11, 12 display the values of our standalone features (before combination into TDI) mapped across Allegheny County. The maps display much of the detail described in Section 4.2.2. Namely, low-to-average connectivity is present within the municipality of Pittsburgh, while high availability is present here. Also visible

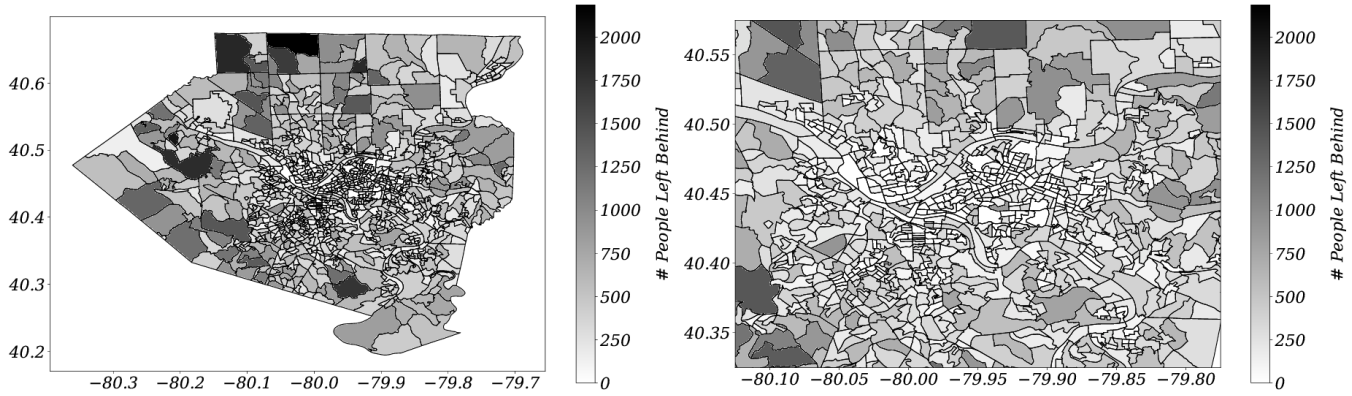

Figure 9: Estimated number of people who do not have access to a transit stop within .4 km. The entire area of Allegheny County is shown (Left) with zoomed area for detail (Right). Outskirts of Allegheny county away from the city are particularly affected.

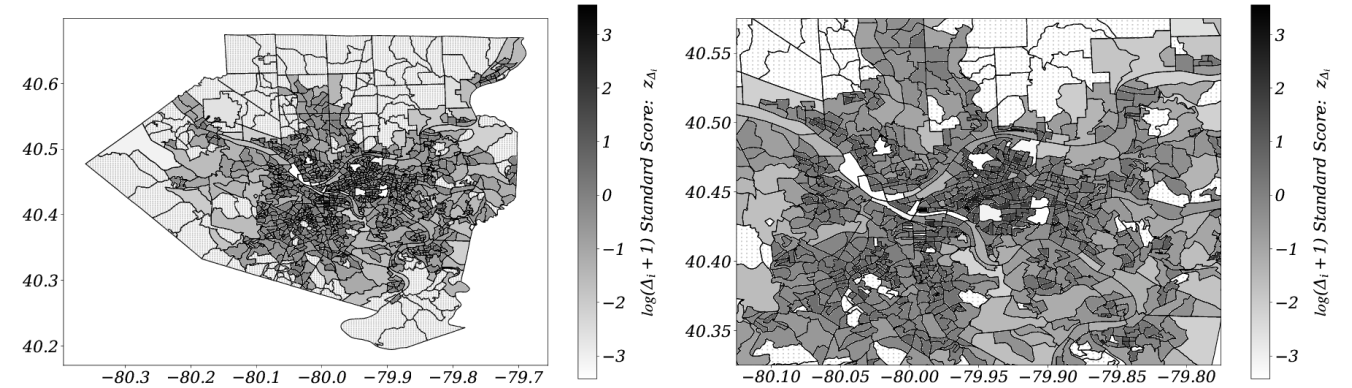

Figure 10: Visualization of Δ_i (transit demand) standard score. Left displays entire area, while Right is zoomed for more detail. For each census block z_{Δ_i} is averaged over non-desert grids to produce an aggregate value. Census blocks with no transit connected nodes, i.e. only outright deserts, are instead dotted by their grid centers. High demand is concentrated within the municipality of Pittsburgh.

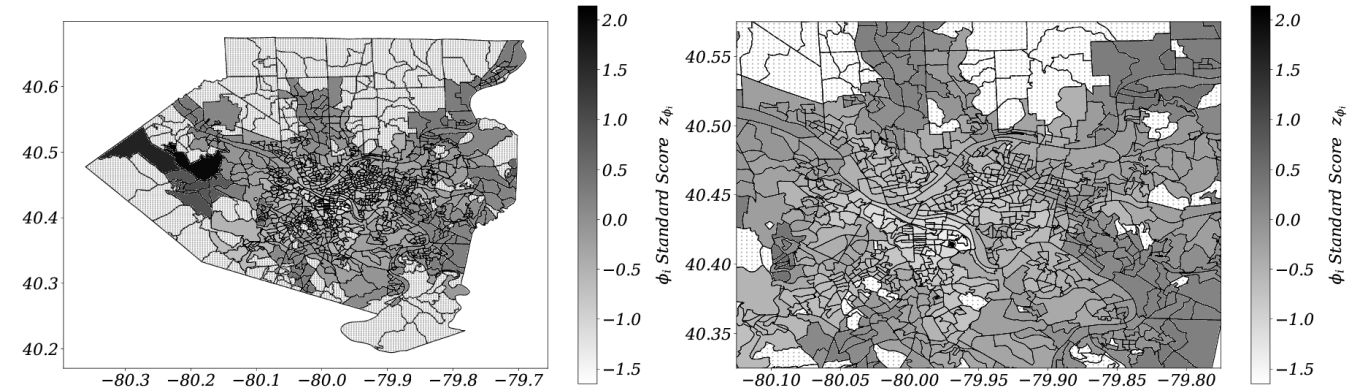

Figure 11: Visualization of ϕ_i (transit connectivity) with standard score averaged for each census block; those with no transit connected nodes are labeled as in Figure 10. Higher connectivity is present outside of metropolitan Pittsburgh. An outlier grid is removed to present better visualization via colormap; it is marked with a large solid point. The value of this grid was -59.35.

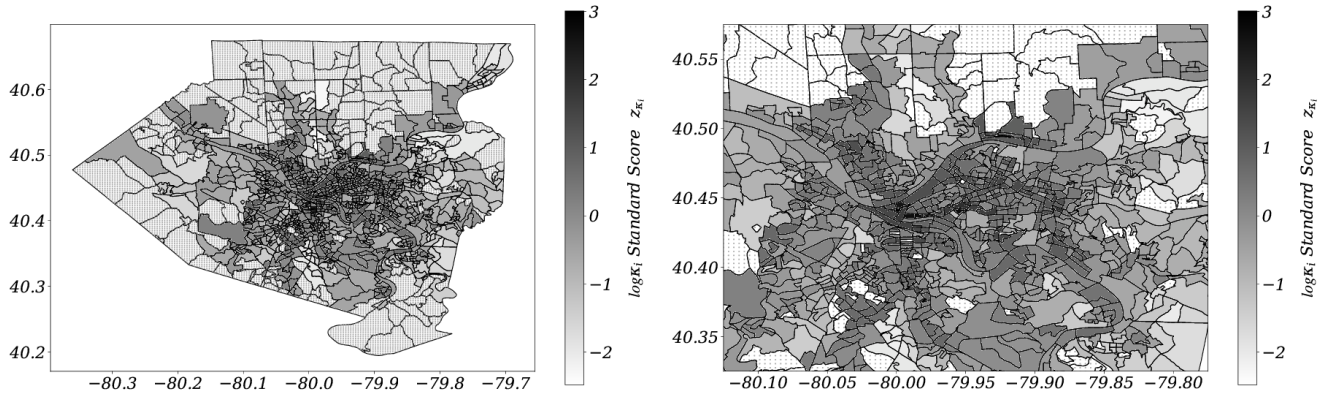

Figure 12: Visualization of κ_i (transit availability) standard score averaged for each census block; those with no transit connected nodes are labeled as in Figure 10. Availability is sporadic with concentration within the municipality of Pittsburgh. Outside of metropolitan Pittsburgh, some highly available areas do exist.

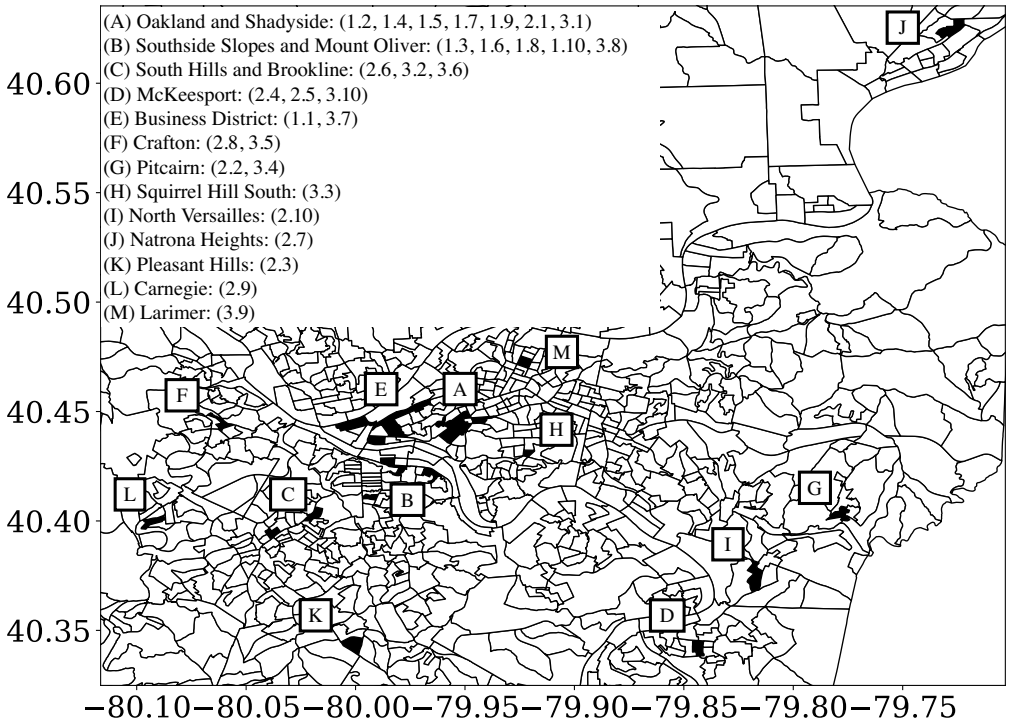

Figure 13: A reference for neighborhoods shown in Table 1 and discussed in Section 4. Key displays neighborhoods and the Block IDs (as in Table 1) which they contain. Relevant census blocks are shaded black. The Hill District is grouped under E.

are the aforementioned far West census blocks with exceedingly high connectivity.

Figure 13 displays a reference guide for neighborhoods that encompass specific census blocks discussed in Section 4. The neighborhoods were identified by the postal codes. This map contains all of the neighborhoods referenced in Table 1 as well as certain other neighborhoods mentioned such as those in the Central Business District and Hill District.