Computers, Environment and Urban Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Assessing the delineated commuter sheds of various clustering methods

D. Schleith[a,*], M.J. Widener[b], C. Kim[c], L. Liu[d]

[a] Brown University, 306 Maxcy Hall, Providence, RI 02912, USA
[b] University of Toronto, Sidney Smith Hall, 100 St. George St, Room 5037, Toronto, ON M5S 3G3, Canada
[c] University of Cincinnati, 400B Braunstein Hall, Cincinnati, OH 45221-0131, USA
[d] University of Cincinnati, 400E Braunstein Hall, Cincinnati, OH 45221-0131, USA

ARTICLE INFO

Keywords:
Commuting
Commuter Shed
Spatial Clusters
MAUP

ABSTRACT

In commuting research the geographic area under investigation is of central importance. When examining commutes occurring in a region of interest, the selection and use of different city, county, or metropolitan region boundaries will have a large impact on analyses of travel times and distances, whether a transit network provides adequate access to jobs, levels of congestion, and so on. This research examines two cluster detection methods to delineate the commuter shed of Miami, Florida and Cincinnati, Ohio. These clusters are then compared to a series of government delineated boundaries often used in commuting research for a jobs-housing balance analysis of each. The results of the clustering methods are different for the two regions due to the differing urban forms of each. The results demonstrate that the decision of what boundary to use for research on commuting can lead to drastically different results. While there is not necessarily any correct boundary, one may be more appropriate and the rationale for such should be more thoroughly discussed.

1. Introduction

In commuting research, the geographic area under investigation is of central importance. When examining commuting, the selection and use of different city, county, or metropolitan regional boundaries will have a large impact on analyses of travel times or distances, whether a transit network provides adequate access to jobs, levels of congestion, and so on. This is closely linked to the spatial form of cities (especially in the North American context) where a relatively dense city is surrounded by suburbs with progressively lower densities (Garreau, 1991). Determining what actually constitutes a commuting region (or "commuter shed") is typically a matter of using administrative boundaries prescribed by the U.S. Census Bureau. In general, the metropolitan region is often used because it represents a big enough area to capture most of the economic activity occurring in and around a smaller, specific urban place like a city. The issue with metropolitan boundaries, however, is summarized in Morrill, Cromartie, and Hart (1999), "… metropolitan areas are widely recognized as far from consistent in meaning or adequate in definition." The problem is largely attributed to the use of counties as building blocks. Counties that are selected to comprise a metropolitan region are those neighboring the county or counties containing the largest principal city. The neighboring counties are included if they are socially and economically connected to the principal county, as measured by the number of commuters into and out of the central county. Counties have a large spatial extent, and oftentimes include vast rural spaces with little relationship to the urbanized area of interest to many researchers. A method for providing a more precise measure is warranted.

Researchers have implicitly used a handful of different names for a commuter shed in the literature but often without an accompanying definition of what such a thing is (Axisa, Newbold, & Scott, 2012). Here, we are interested in the attraction of a central city and all of the commuting that occurs within some boundary given the pull of the central city. Henry, Barkley, and Bao (1997) referred to this as the functional economic area. In their work they determined that if the pace of growth at the fringe is faster than that of the center city the growth spreads out and results in more suburbanization. These suburbs are both intricately tied to and relatively distinct from the center city they surround. People live, shop, and have their children attend schools in these more distant areas but still often travel into the center city for work. The growth at the fringe where people live, along with the commuting ties to the center city still need to be better defined to exclude the very rural fringes but capture the activity from suburb to central city and the increasingly common suburb to suburb.

In many ways a commuter shed is like a cluster of commuting activity, where there are significant links between residents moving in and out of relevant, contiguous zones (like census blocks or tracts). In this research commuter sheds are created using two spatial clustering

* Corresponding author.
E-mail addresses: daniel_schleith@brown.edu (D. Schleith), michael.widener@utoronto.ca (M.J. Widener), changjoo.kim@uc.edu (C. Kim), lin.liu@uc.edu (L. Liu).

https://doi.org/10.1016/j.compenvurbsys.2018.04.004
Received 10 May 2017; Received in revised form 27 February 2018; Accepted 15 April 2018

methods and those boundaries are compared to boundaries defined by local or the federal government. In order to more thoroughly compare the boundaries, a standard commuting study analyzing the jobs-housing balance (JHB) of each of the 12 boundaries (six per city, Miami, Florida and Cincinnati, Ohio) is done. This facilitates a real world comparison of how the different clustering methods perform relative to the other "prescribed" boundaries, besides simply number of tracts in the commuter shed.

The following section highlights several studies that have largely ignored justifying the chosen boundary. Also included are examples of research that have examined some of the ways that the boundary of a study area can affect the results of the work. After that section we describe the clustering methods and the excess commuting (EC) framework, which we use to analyze the JHB and thereby assess the quality of our created sheds. This is followed by a discussion of the Longitudinal Employer-Households Dynamics (LEHD) data used in our example study. The clusters created are examined relative to the prescribed boundaries with a comparison of the EC results. Next, the implications of the clusters and how commuter sheds may be better defined and applied across research is discussed. Finally, we conclude with some general thoughts, specific limitations, and possible directions for future work.

2. Discrepancies across commuting research

As previously mentioned, commuter sheds are the de facto analysis areas of most commuting research, the results of which are sensitive to the definition of the study area. Researchers are therefore interested in these definitions, as they attempt to accurately describe settlement patterns across the country and provide a reasonable assessment of how people move within an urban region. The "default" commuter shed used by the U.S. Census Bureau is the Metropolitan Statistical Area (MSA). The U.S. Census Bureau defines these by calculating the percent of commutes occurring to and from central and outlying counties. If the combined interaction is greater than some threshold, the outlying county can be included in the MSA as long as it does not have a higher interaction percent with a different central county (where the central city is located). Generally, this effort is to determine the location of suburbs of various urban areas and which rural places are not exurban suburbs but places unto themselves.

To illustrate how the boundary question is often overlooked, several papers are discussed here with their various study region rationale. First, Hu and Wang investigated the jobs-housing balance and test a Monte-Carlo simulation approach to EC in two papers that both use East Baton Rouge Parish in Louisiana as the study area (Hu & Wang, 2015, 2016). In their work they state, "[East Baton Rouge] is the core of the Baton Rouge metropolitan area (other surrounding parishes are mostly rural)." The On The Map website (onthemap.ces.census.gov), which provides a simple tool for examining commuting patterns, shows that in 2010 East Baton Rouge parish had 124,722 primary commutes originate and terminate within the parish. However, another 121,258 commutes originated outside of the parish but terminated within it. Almost 50% of the commutes into East Baton Rouge are excluded from this analysis. We do not intend to make the case here that excluding the surrounding counties is or is not the appropriate scale of analysis, just that an explanation for the exclusion of 50% of the commutes is likely warranted.

Horner and Schleith (2012) investigated the jobs-housing balance of Leon County, Florida. No explanation is given for why the metropolitan boundary was not used. However, in this case, it is also true that the surrounding counties are also rural. Leon County does completely encapsulate the city of Tallahassee and so the county may be the appropriate study region. In this case though, 85,091 commutes originated and terminated within Leon County in 2010. This excludes 67,838 commutes that originated outside of Leon County but terminated within it. Here, 44% of the commutes into the study area are excluded from the analysis.

Kim, Sang, Chun, and Lee (2012) investigated commuting in Hamilton County, Ohio disaggregated by occupation and gender. They chose Hamilton County as it completely contains the city of Cincinnati. In 2010 however, 229,578 commutes originated and terminated within Hamilton County. Another 231,069 commutes originated outside of the county and terminated within it. Fifty percent of the commutes into the county are excluded from the analysis. In the case of the Cincinnati metro region, 736,000 commutes originate and terminate within the MSA boundary. So in this case we might reasonably assume that most of those ~460,000 commutes would have been included in an analysis using the MSA boundary.

Axisa et al. used essentially the equivalent of the Toronto Metropolitan Region (although some rural zones are excluded) as the area for two studies examining migration and commute distance in what they refer to as the Toronto commuter shed (Axisa, Newbold, & Scott, 2012; Axisa, Scott, & Newbold, 2012). These papers are the first that we could find that specifically refer to the area under investigation as a commuter shed. In their study, however, some census subdivisions are excluded for not having enough commuting interaction with the census subdivisions with the highest job density. Again this is at the discretion of the authors and what effect including other subdivisions would have on the final results of an analysis is unclear. In these two papers though, a justification for the inclusion of different subdivisions is at least based on the percentage of commutes into core subdivisions.

Wang (2001) used the urbanized area (UA) boundary of Columbus, Ohio, to examine intraurban variations in commuting. The urbanized boundary is described to represent the commuter shed for the city of Columbus. Unlike other authors, Wang does include statistics on the proportions of residents and workers who are included and excluded by the urbanized boundary. This information may be important in order for researchers and planners to better use or apply the findings of the work.

Another important study is that of Frost, Linneker, and Spence (1998). In this work, the authors included commutes that occur within cities in England, as well as trips that originate outside of the city boundaries but terminate within them. They found that the commuters whose trips originated outside of the city accounted for the largest increases in commute times. This study is unique in that it does not include commutes from one suburb to another, only trips into the city While this paper stands out as an excellent way to investigate commuting regarding some boundary, we might currently (especially in a U.S. context) wonder about commutes occurring between suburbs of various cities. In choosing a particular commuter shed, all commutes that originate and terminate with the shed are examined, including those between smaller suburban cities.

Additionally, Niedzielski (2006) investigated commuting in two cities in Poland. This work was important because it examined commutes at the zonal level within a city or region. To do this, commutes have to be examined by commutes in to and commutes out of all zones. This work showed that commutes in and out vary drastically by kind of commuter (industry).

Finally, Horner and Murray's (2002) and Niedzielski, Horner, and Xiao's (2013) papers investigated the Modifiable Areal Unit Problem (MAUP) and the ways in which the EC measures scale. These papers demonstrated the various ways that the EC measures can change based on the underlying zonal geography. While the present work does not specifically address changing the underlying units, each of the commuter sheds examined are comprised of a different number of census tracts and so are a series of modified units. Both previously mentioned papers found that some of the EC measures scale with size of the study region while others are scale independent. This is important to consider for the present analysis as the size of the study region changes with each boundary investigated.