I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13642

Delineating Functional Labour Market Areas with Estimable Classification Stabilities

Benjamin Davies
David C. Maré

AUGUST 2020

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13642

Delineating Functional Labour Market Areas with Estimable Classification Stabilities

Benjamin Davies
Motu Economic and Public Policy Research

David C. Maré
Motu Economic and Public Policy Research and IZA

AUGUST 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Delineating Functional Labour Market Areas with Estimable Classification Stabilities*

We describe an unsupervised method for delineating functional labour market areas (LMAs) in national commuting networks. Our method uses the Louvain algorithm, which we extend to support top-down hierarchical LMA classification and estimable classification stabilities. We demonstrate our method using historical Census commuting data from New Zealand.

JEL Classification: J61, R12, R23

Keywords: community detection, commuting, functional boundaries, labour market areas, networks

Corresponding author:
David C. Maré
Motu Economic and Public Policy Research
Level 1, 97 Cuba Street
PO Box 24390
Wellington 6142
New Zealand

E-mail: dave.mare@motu.org.nz

* The research discussed in this paper was supported by Te Pūnaha Matatini, and by the 'Building Better Homes Towns and Cities' National Science Challenge. We thank Nathan Chappell for his contributions to the early stages of this research. Access to the data used in this study was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. The results presented in the study are the work of the authors, not Statistics New Zealand.

Introduction

Identifying functional, rather than administrative, labour market areas (LMAs) is important for analysing spatial patterns of economic activity. Functional boundaries capture the geography of interactions among employers and employees (Goodman, 1970; Brown & Holmes, 1971; Casado-Diaz, 2000), whereas administrative boundaries typically ignore such interactions. By construction, functionally-defined LMAs are largely self-contained, and thus represent an obvious unit of analysis for studies of local labour markets. The self-containment also reduces the impact of the modifiable areal unit problem (Openshaw, 1984).

This paper describes a method for delineating functional LMAs in national commuting networks. Our method has four desirable properties that are generally not discussed in the extant literature. First, it can be applied without supervision. Early attempts at functional LMA delineation (Goddard, 1970; Brown & Holmes, 1971; Masser & Brown, 1975; Coombes, Green, & Openshaw, 1986; Tolbert & Sizer, 1996) use semi-supervised methods, such as factor analysis and hierarchical clustering of origin-destination commuter matrices. These methods require subjective judgments about where LMA boundaries lie, which may lead to ad hoc modifications and biased inferences. In contrast, our method requires no subjective judgments beyond choosing to use the method.

Second, our LMA delineation method is scale invariant. The LMAs identified using our method do not depend on the units in which commuting flows are measured. This scale-invariance validates temporal and subgroup comparisons.

Third, our method supports top-down, hierarchical classifications of LMAs and sub-LMAs. This allows us to “zoom in” on the commuting patterns within each LMA, facilitating deeper spatial analyses of economic activity.

Finally, our method yields a set of stability measures that summarise the reliability of specific allocations of areas to LMAs. These measures allow us to evaluate the robustness of our LMA classifications to noise in the underlying commute data.

We demonstrate our LMA delineation method using historical Census commuting data from New Zealand. This demonstration has the additional benefit of updating the current LMA boundaries (Newell & Perry, 2002; Papps & Newell, 2002; Ralphs & Goodyear, 2008) used to conduct spatial economic analysis in New Zealand. Our method can be applied readily to other commuting data, such as those derived from future Censuses and from administrative data (Fabling & Maré, 2020).

This paper contributes to the literature on functional LMA classification. We describe four desirable features of a classification method—lack of supervision required, scale-invariance, hierarchicality, measurable stability—and present a method that possesses these features. We use tools from network science to study the spatial extent of labour market interactions. Previous studies apply similar tools to analyse how such interactions contribute to growth (Davies & Maré, 2020; Rigby, Roesler, Kogler, Boschma, & Balland, 2019), innovation (Balland, Boschma, Crespo, & Rigby, 2019), and knowledge and skill concentration (Balland & Rigby, 2017; Balland, et al., 2020). Our method for delineating functional LMAs supports future such studies by providing a conceptually robust and measurably stable means of identifying distinct areas in which workers and firms interact.

Delineating functional LMAs

We delineate functional LMAs by partitioning the nodes in a commuting network. This network captures the commuting patterns between residences and workplaces. We extract from such patterns sets of nodes with relatively self-contained commuting flows. These sets correspond to functional LMAs; they comprise locational entities with more internal than external interaction or connections (Brown & Holmes, 1971). This section describes how we construct the commuting network and partition its nodes into LMAs.

Consider an economy comprising a set $\mathcal{A}$ of areas that partition the set of residential and workplace addresses. Let F be the origin-destination commuting matrix with ij^{th} entry equal to the commuting flow f_{ij} from area i to area j. Following De Montis et al. (2013), Pálóczi (2016), and Adam et al. (2018), we measure the strength of the commuting flows between areas via the symmetric matrix

$$M = F + F^T - \operatorname{diag}(F),$$

where F denotes the transpose of F, and where $\operatorname{diag}(F)$ is the matrix with ii^{th} entry equal to f_{ii} and off-diagonal entries equal to zero. The ij^{th} entry of M counts the round-trips between residences and workplaces, aggregated to the area level. Subtracting $\operatorname{diag}(F)$ avoids double-counting intra-area flows. We interpret M as the adjacency matrix for a weighted, undirected network N in which nodes correspond to areas and edges have weight equal to the pairwise strength of the commuting flows between incident areas.

The Louvain algorithm

We use the Louvain algorithm (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) to partition the network N into disjoint subsets that correspond to LMAs. This algorithm is a heuristic for maximising "modularity" (Newman & Girvan, 2004; Newman, 2004), which captures the extent to which groups of nodes are intra-connected densely but inter-connected sparsely.[1] The modularity of a partition $\mathcal{P}$ is defined as

$$Q(\mathcal{P}) = \frac{1}{m} \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{A}} \left[M_{ij} - \frac{k_i k_j}{m} \right] \delta(P_i, P_j),$$

where $k_i = \sum_{j \in \mathcal{A}} M_{ij}$ is the weighted degree of area i, $m = \sum_{i \in \mathcal{A}} k_i$ is the sum of weighted degrees, P_i is the part containing area i, and

$$\delta(P_i, P_j) = \begin{cases} 1 & \text{if } P_i = P_j \\ 0 & \text{otherwise} \end{cases}$$

is the Kronecker delta function. The modularity $Q(\mathcal{P})$ equals zero if the partition $\mathcal{P}$ produces no stronger within-part commuting flows than would be expected if commuting flows were random, and is larger for partitions with more self-contained parts. Thus, maximising modularity is akin to maximising self-containment relative to a "null model" in which edge weights are random.

The Louvain algorithm is one of many algorithms that partition the nodes in a network into "communities" that "share common properties and/or play similar roles" (Fortunato, 2010). It is applied in several studies of local labour markets (De Montis, Caschili, & Chessa, 2013; Pálóczi, 2016; Thomas, Adam, & Verhetsel, 2017; Adam, Delvenne, & Thomas, 2018; Dannemann, Sotomayor-Gómez, & Samaniego, 2018). Other algorithms have different objectives, such as maximising random walk containment (Pons & Latapy, 2006) or minimising the random walk description lengths (Rosvall & Bergstrom, 2008). We use the Louvain algorithm because it performs well, in terms of speed and accuracy, compared to alternative algorithms (de Sousa & Zhao, 2014; Yang, Algesheimer, &

[1] A heuristic is necessary because modularity maximisation is NP-complete (Brandes, et al., 2008) and thus infeasible for large commuting networks.

Tessone, 2016). However, our method supports any algorithm that partitions nodes into communities.

The Louvain algorithm comprises four (possibly repeated) steps:

1. Assign each node to a unique community.
2. Sequentially reassign each node to the community of one of its neighbours so as to deliver the greatest increase in modularity. Do so until no such increase is possible.
3. Construct a new network with nodes equal to the communities identified in step 2 and with edge weights equal to the sum of edge weights among nodes in the two incident communities.[2]
4. If the networks considered in steps 1 and 3 are identical then stop. Otherwise, go to step 1.

These four steps deliver an assignment of nodes to communities, corresponding to an assignment of areas to LMAs. Thus, the Louvain algorithm provides an unsupervised method for delineating LMAs by endogenously determining the number of, and boundaries between, LMAs based on the commuting strength matrix M.

The modularity $Q(\mathcal{P})$ of a partition $\mathcal{P}$ is invariant to multiplying M by a real scalar. Therefore, the LMAs identified by the Louvain algorithm depend only on the proportion of commuters travelling between areas rather than the absolute number. Such scale-invariance ensures that LMA partitions delivered by the algorithm are comparable across time because population growth that increases all commuting flows by a constant factor will leave the set of identified LMAs unchanged. Scale-invariance also validates subgroup comparisons because LMA boundaries do not depend on the units in which flows are measured (e.g., number of people vs. number of males or adults of a specific age).

Estimating LMA stabilities

The communities identified by the Louvain algorithm vary with the order in which nodes are considered for reallocation during step 2. We overcome this ambiguity using the method proposed by Adam et al. (2018): we run the algorithm many times, each time randomising the order in which nodes are considered for reallocation, and extract the modal community allocations across runs. This method allows us to estimate the stability of each area's LMA classification via the proportion of runs in which nodes are assigned to their modal communities. These estimates allow us to evaluate the robustness of a given LMA classification to measurement errors, which perturb the commuting matrix randomly.

Randomising the order in which nodes are considered for reallocation by the Louvain algorithm introduces randomness into our LMA delineation method. Consequently, we cannot guarantee the equality of independently generated LMA boundaries derived from the same commuting data. However, estimating LMA classification stabilities helps to identify where independently generated boundaries are most likely to disagree.

Resolution limit

The Louvain algorithm—along with other modularity maximisation algorithms—has a "resolution limit" (Fortunato & Barthélemy, 2007; Good, de Montjoye, & Clauset, 2010): it may fail to detect small communities, even if those communities are unambiguously defined. For example, if the underlying network is sparse then any edge with positive weight may represent a stronger

[2] This method preserves modularity (Arenas, Duch, Fernandez, & Gomez, 2007) but double-counts within-community flows between distinct nodes, which contribute to self-loops in the new network.

connection between disparate communities than would be expected if edges were distributed randomly. Consequently, merging two or more "true" communities may increase modularity.

We combat the resolution limit by reapplying the Louvain algorithm to the commuting flows contained within each LMA to identify sub-LMA boundaries.[3] This allows us to "zoom in" on the structure of commuting patterns within each LMA and to potentially recover "true" LMAs that get amalgamated by the Louvain algorithm.

Our top-down approach has at least three alternatives. First, we could exploit the hierarchical nature of the Louvain algorithm: each iteration over steps 1–3 delivers an increasingly coarse assignment of nodes to communities, so the "true" communities could be recovered by terminating the algorithm prematurely. However, this bottom-up approach requires prior knowledge of which hierarchical level represents the "true" communities. Our top-down approach requires no such knowledge.

Second, we could apply a convex transformation to each commuting flow f_{ij} before defining the commuting strength matrix M. Such a transformation increases the variation in relative flow sizes and, therefore, results in more granular LMA partitions.[4] However, convex transformations amplify measurement errors in the relative size of flows in and out of small areas. Moreover, non-linearly transforming commuting flows would violate the scale-invariance property that we desire our method to possess. This violation would preclude clean temporal and subgroup comparisons by making the relative size of M's entries sensitive to the units in which flows are measured.

Third, we could modify the Louvain algorithm to accept a "resolution parameter" that influences the granularity of the detected communities (Lambiotte, Delvenne, & Barahona, 2014; Reichardt & Bornholdt, 2006). However, introducing such a parameter would make LMA delineation require supervision in the form of choosing the parameter's value. Moreover, this choice would require prior knowledge of the "true" LMA boundaries. Because we desire an unsupervised delineation method, and because the resolution parameter must be chosen arbitrarily without prior knowledge, we do not introduce the parameter into our method.

New Zealand as a case study

Data

We demonstrate our LMA delineation method by applying it to historical Census commuting data from New Zealand.[5] These data provide employee counts by usual residence and workplace, both aligned to 2013 area unit codes, at the dates of 2001, 2006, and 2013 Censuses. Area units are statistical areas with an average population of around 2,000 and an average size of 140 km^2.

Figure 1 plots the employed resident population density of each area unit, averaged across the three Census years in our data. These mean densities capture the spatial distribution of New Zealand's labour force, which is concentrated in small, often coastal pockets separated by large forests and agricultural areas.

We exclude from our analysis all employees whose residence or workplace address belonged to an area unit that was unknown, outside a territorial authority, or contained no land. We also exclude employees with commute distances beyond 150 kilometres, which we estimate using the Euclidean

[3] Beckers (2019) uses this approach to study hierarchical layers of the Belgian logistics network.

[4] For example, transforming intra-area convexly flows makes each area appear more self-contained and, consequently, less likely to be merged into adjacent communities because the modularity gain from doing so decreases.

[5] The data and code used in our analysis are available at https://doi.org/10.5281/zenodo.4003346.

distance between origin and destination area unit centroids. Excluding long commutes dampens the influence of idiosyncratic travel patterns arising through genuine but unusual circumstances, or through miscoded residence or workplace addresses (Newell & Perry, 2002).

Table 1 and Table 2 report the proportion of employees excluded from our data for each Census year, broken down by exclusion criterion and origin region. Together, our criteria exclude about a fifth of commuting flows in Census years 2001 and 2006, and about 15% of such flows in Census year 2013. Gisborne and Tasman tend to be the regions most affected by our exclusions, while Wellington tends to be the least affected. The "long commute" exclusion rate grows across Census years, consistent with the nationwide increase in commute distances portrayed in Table 3.

LMA and sub-LMA boundaries

Table 4 reports the number of area units included in our analysis, and the number of LMAs and sub-LMAs identified, for each Census year. The number of LMAs and sub-LMAs fell between 2006 and 2013, consistent with the rise in commute distances and consequent expansion of labour market boundaries. The decrease in LMAs is also consistent with the decrease in area units' supply- and demand-side self-containment shown in Table 5, which also shows that both LMAs and sub-LMAs fell in supply- and demand-side self-containment.[6] These patterns suggest that, on average, New Zealand's labour force became less segregated spatially between the 2001 and 2013 Censuses.

Our method delineates a smaller number of larger LMAs than previous studies of commuting patterns in New Zealand (Newell & Perry, 2002; Papps & Newell, 2002; Ralphs & Goodyear, 2008). These studies apply Coombes et al.'s (1986) semi-supervised method, which the authors modify to better identify smaller rural settlements.

Table 6–Table 8 describe attributes of the LMAs we identify for Census years 2001, 2006, and 2013. Across all three years, the LMAs we identify contain between one and 379 area units, with land areas and employee populations varying over several orders of magnitude. The variation in LMA employee populations reflects the variation in population density across area units: LMAs in urban centres (i.e., Auckland, Wellington, and Christchurch) tend to have large employee populations but relatively small land areas. The LMAs we identify tend to have supply- and demand-side self-containments above 90%. The least supply-side contained LMAs are near the interface of the Auckland and Waikato regions, reflecting inter-LMA commuting between two relatively dense areas with strong transport links.

Figures 2–4 present the boundaries of the LMAs described in Table 6–Table 8. In each figure, we shade in grey the area units with LMA classification stabilities less than 100%. These unstable area units tend to appear in large, contiguous blocks, suggesting that the instability comes from whether LMAs are merged rather than from whether individual area units are on a particular side of an LMA boundary. Excluding LMAs with islands, the only non-contiguous LMA is Wellington in 2001 and 2006 (respectively, LMA 16 and 14), which contains the area unit of Mara on the North Island's lower east coast. This non-contiguity appears to arise from a handful of Wellington residents commuting to Mara on the date of the 2001 and 2006 Censuses.[7]

Figure 5 presents an alluvial diagram that visualises the reallocation of area units between LMA boundaries across Census years. We weight area units by employed resident population so that the

---

[6] Demand-side self-containment is the proportion of local jobs that are filled by local residents. Supply-side self-containment is the proportion of local residents who work locally.

[7] This may reflect a mis-recorded reference to the *Mana* area unit, which is within the main part of the Wellington LMA.

presented flows represent the movement of labour force shares between LMAs in successive years. The diagram shows that there is very little change, in terms of reallocations of employees among LMAs, in LMA boundaries between the 2001, 2006 and 2013 Censuses.

Figure 6 presents sub-LMA boundaries in Wellington for each Census year.[8] The 2001 LMA boundaries roughly delineate the Kapiti Coast, Upper Hutt, Lower Hutt, Porirua, and Wellington Central districts. The sub-LMAs approximating the Upper and Lower Hutt districts are merged in 2013, and there are some area unit reallocations between the sub-LMAs approximating the Porirua and Wellington Central districts between Census years. However, Figure 7 shows that these large land area reallocations correspond to negligible employee reallocations.

Classification stabilities

Table 9 reports the distribution of the number of communities identified across runs for each Census year. Our final LMA counts match the integer-rounded means and medians for each year, with the most variation in 2001 and the least in 2013.

Table 10 reports the employee-weighted, resident-weighted and unweighted mean LMA allocation stabilities by Census year. Comparing weighted and unweighted means suggests that, on average, LMA allocations are more stable for area units with greater employee and resident populations.

Table 10 also reports weighted and unweighted means of the variable "Stability = 100%," which equals one if an area unit is allocated to an LMA unambiguously and equals zero otherwise. The weighted means capture the proportion of commuters with unambiguous LMA memberships, while the unweighted means capture the proportion of area units with such memberships. On average, our LMA delineation strategy allocates more than 95% of employees to unambiguous LMAs in Census years 2001 and 2013, and more than 84% of employees to unambiguous LMAs in 2006.

Conclusion

This paper describes a method for delineating functional LMAs in national commuting networks. Our method has four desirable properties: it is unsupervised, scale-invariant, hierarchical, and measurably stable. We demonstrate our method by applying it to New Zealand Census commuting data. However, our method can be applied to any commuting data.

Our objective with this paper is to present an LMA delineation method requiring minimal subjective judgments. However, some such judgments are unavoidable. For example, raw commuting data typically require cleaning (e.g., removing unknown addresses and implausibly long commutes), which invites subjectivity via the choice of cleaning procedure. Likewise, our method relies on a subjective choice of community detection algorithm. There are many such algorithms, all with histories of successful use, and the choice between them is also subjective. The Louvain algorithm, on which our delineation method relies, delivers reasonable results and is commended by review papers. We encourage further research exploring alternative algorithms, and their relative merits and drawbacks.

[8] The figure excludes the Mara area unit, which comprises Wellington sub-LMA 1 in Census years 2001 and 2006 but is not contiguous with the rest of the Wellington LMA.

References

Adam, A., Delvenne, J.-C., & Thomas, I. (2018). Detecting communities with the multi-scale Louvain method: robustness test on the metropolitan area of Brussels. *Journal of Geographical Systems, 20*(4), 363-386. doi:10.1007/s10109-018-0279-0

Arenas, A., Duch, J., Fernandez, A., & Gomez, S. (2007). Size reduction of complex networks preserving modularity. *New Journal of Physics, 9*(6), 176.

Balland, P.-A., & Rigby, D. L. (2017). The Geography of Complex Knowledge. *Economic Geography, 93*(1), 1-23.

Balland, P.-A., Boschma, R., Crespo, J., & Rigby, D. L. (2019). Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional Studies, 53*(9), 1252-1268.

Balland, P.-A., Jara-Figueroa, C., Petralia, S. G., Steijn, M. P., Rigby, D. L., & Hidalgo, C. A. (2020). Complex economic activities concentrate in large cities. *Nature Human Behaviour, 4*, 248-254.

Beckers, J. R. (2019). *The logistics sector in a consumer driven society, essays on location and network structure.* PhD thesis, University of Antwerp.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. doi:10.1088/1742-5468/2008/10/P10008

Brandes, U., Delling, D., Gaertler, M., Gorke, R., Hoefer, M., Nikoloski, Z., & Wagner, D. (2008). On modularity clustering. *IEEE Transactions on Knowledge and Data Engineering, 20*(2), 172-188. doi:10.1109/TKDE.2007.190689

Brown, L. A., & Holmes, J. (1971). The delimitation of functional regions, nodal regions, and hierarchies by functional distance. *Journal of Regional Science, 11*(1), 57-72. doi:10.1111/j.1467-9787.1971.tb00240.x

Casado-Diaz, J. M. (2000). Labour market areas in Spain: A case study. *Regional Studies, 34*(9), 843-856. doi:10.1080/00343400020002976

Clauset, A., Newman, M. E., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E, 70*(6).

Coombes, M. G., Green, A. G., & Openshaw, S. (1986). An efficient algorithm to generate official statistical reporting areas: the case of the 1984 travel-to-work areas revision in Britain. *Journal of Operational Research Society, 37*(10), 943-953. doi:10.2307/2582282

Dannemann, T., Sotomayor-Gómez, B., & Samaniego, H. (2018). The time geography of segregation during working hours. *Royal Society Open Science, 5*. doi:10.1098/rsos.180749

Davies, B., & Maré, D. C. (2020). Relatedness, Complexity and Local Growth. *Regional Studies*. doi:10.1080/00343404.2020.1802418

De Montis, A., Caschili, S., & Chessa, A. (2013). Commuter networks and community detection: a method for planning sub national areas. *European Physical Journal Special Topics, 215*(1), 75-91. doi:10.1140/epjst/e2013-01716-4 de Sousa, F. B., & Zhao, L. (2014). Evaluating and comparing the igraph community detection algorithms. *Proceedings of the 2014 Brazilian Conference on Intelligent Systems.* doi:10.1109/BRACIS.2014.79

Fabling, R., & Maré, D. C. (2020). *Commute patterns over time: Using administrative data to identify where employees live and work.* Motu Working Paper 20-05, Motu Economic and Public Policy Research.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports, 486*(3-5), 75-174. doi:10.1016/j.physrep.2009.11.002

Fortunato, S., & Barthélemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences, 104*(1), 36-41. doi:10.1073/pnas.0605965104

Goddard, J. B. (1970). Functional regions within the city centre: a study by factor analysis of taxi flows in Central London. *Transactions of the Institute of British Geographers, 49*, 161-182.

Good, B. H., de Montjoye, Y.-A., & Clauset, A. (2010). Performance of modularity maximisation in practical contexts. *Physical Review E, 81*. doi:10.1103/PhysRevE.81.046106

Goodman, J. F. (1970). The definition and analysis of local labour markets: some empirical problems. *British Journal of Industrial Relations, 8*(2), 179-196. doi:10.1111/j.1467-8543.1970.tb00968.x

Lambiotte, R., Delvenne, J.-C., & Barahona, M. (2014). Random walks, Markov processes, and the multiscale modular organisation of complex networks. *IEEE Transactions on Network Science and Engineering, 1*(2), 76-90.

Liu, X., Gong, L., Gong, Y., & Liu, Y. (2015). Revealing travel patterns and city structure with taxi trip data. *Journal of Transport Geography, 43*, 78-90. doi:10.1016/j.jtrangeo.2015.01.016

Masser, I., & Brown, P. J. (1975). Hierarchical aggregation procedures for interaction data. *Environment and Planning A, 7*, 509-523. doi:10.1068/a070509

Newell, J., & Perry, M. (2002). Functional labour markets revealed by travel to work data 1991 and 2001. *Labour, Employment and Work in New Zealand*, 17-24. doi:10.26686/lew.v0i0.1208.

Newman, M. E. (2004). Analysis of weighted networks. *Physical Review E, 70*(5).

Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*, 026113. doi:10.1103/PhysRevE.69.026113

Openshaw, S. (1984). *The modifiable areal unit problem.* Norwich: Geo Books.

Pálóczi, G. (2016). Researching commuting to work using the methods of complex network analysis. *Regional Statistics, 6*(1), 3-22. doi:10.15196/RS06101

Papps, K. L., & Newell, J. O. (2002). Identifying functional labour market areas in New Zealand: a reconnaissance study using travel-to-work data. *IZA Discussion Paper 443*.

Peel, L., Larremore, D. B., & Clauset, A. (2017). The ground truth about metadata and community detection in networks. *Science Advances, 3*(5). doi:10.1126/sciadv.1602548

Pons, P., & Latapy, M. (2006). Computing Communities in Large Networks Using Random Walks. *Journal of Graph Algorithms and Applications, 10*(2), 191-218.

Ralphs, M., & Goodyear, R. (2008). The daily commute: An analysis of the geography of the labour market using 2006 Census data. *Labour, Employment and Work in New Zealand*.

Reichardt, J., & Bornholdt, S. (2006). Statistical mechanics of community detection. *Physical Review E, 74*(1), 016110.

Rigby, D. L., Roesler, C., Kogler, D., Boschma, R., & Balland, P.-A. (2019). *Do EU regions benefit from smart specialisation?* Papers in Evolutionary Economic Geography 19.31, Utrecht University.

Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences, 105*(4), 1118-1123. doi:10.1073/pnas.0706851105

Statistics New Zealand. (2013). *2013 Census definitions and forms.*

Thomas, I., Adam, A., & Verhetsel, A. (2017). Migration and commuting interactions fields: a new geography with community detection algorithm? *Belgio, 4*. doi:10.4000/belgeo.20507

Tolbert, C. M., & Sizer, M. (1996). *U.S. commuting zones and labour market areas: a 1990 update.* ERS Staff Paper 9614, Economic Research Service, U.S. Department of Agriculture.

Yang, X., Fang, Z., Yin, L., Li, J., Zhou, Y., & Lu, S. (2018). Understanding the spatial structure of urban commuting using mobile phone location data: a case study of Shenzhen, China. *Sustainability, 10*(5). doi:10.3390/su10051435

Yang, Z., Algesheimer, R., & Tessone, C. J. (2016). A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Scientific Reports, 6*. doi:10.1038/srep30750

Figures and tables

Table 1: Employee exclusion rates (%) by criterion and Census year

Exclusion criterion	2001 Census	2006 Census	2013 Census
Residence or workplace address belongs to excluded area unit	18.93	19.99	13.53
Estimated commute distance beyond 150km	0.87	1.03	1.46
Total	19.79	21.02	14.99

Table 2: Employee exclusion rates (%) by origin region and Census year

Region of origin	2001 Census	2006 Census	2013 Census
Auckland	19.67	19.53	14.76
Bay of Plenty	21.80	22.27	16.38
Canterbury	18.81	18.83	14.86
Gisborne	24.76	25.53	16.41
Hawke's Bay	22.58	23.82	16.58
Manawatu-Wanganui	18.73	23.61	13.43
Marlborough	21.53	22.23	17.31
Nelson	22.28	20.62	16.49
Northland	22.10	24.86	16.94
Otago	19.47	22.58	17.46
Southland	19.04	23.82	16.96
Taranaki	17.63	23.97	13.88
Tasman	23.65	22.57	17.48
Waikato	21.62	22.44	15.72
Wellington	17.43	19.85	12.25
West Coast	19.19	23.85	15.52

Table 3: Cumulative proportion of commuting flows by estimated distance and Census year, after removing flows with unknown or oceanic origin or destination

Distance d (km) band	2001 Census	2006 Census	2013 Census
$d < 50$	97.53	97.23	96.30
$d < 100$	98.61	98.37	97.79
$d < 150$	98.93	98.71	98.31
$d < 200$	99.13	98.91	98.59
$d < 250$	99.25	99.05	98.78
$d < 300$	99.35	99.16	98.91
$d < 350$	99.45	99.26	99.05
$d < 400$	99.53	99.37	99.17

Table 4: Number of area units, LMAs and sub-LMAs by Census year

Variable	2001 Census	2006 Census	2013 Census
Number of area units	1,881	1,882	1,887
Number of LMAs	31	30	29
Number of sub-LMAs	251	213	208

Table 5: Area unit, LMA, and sub-LMA attribute means (standard deviations) by Census year

Attribute	2001 Census	2006 Census	2013 Census
		Area units	
Employed residents (000s, before exclusions)	0.92 (0.67)	1.05 (0.77)	1.06 (0.79)
Employed residents (000s)	0.74 (0.55)	0.83 (0.63)	0.90 (0.69)
Employees (000s)	0.74 (1.66)	0.83 (1.90)	0.90 (1.99)
Supply-side self-containment (%)	30.02 (22.38)	27.98 (21.24)	27.82 (20.34)
Demand-side self-containment (%)	44.95 (25.30)	42.67 (24.40)	41.84 (23.64)
		LMAs	
Employed residents (000s, before exclusions)	55.70 (96.50)	66.18 (114.36)	68.99 (120.22)
Employed residents (000s)	44.69 (77.97)	52.28 (92.38)	58.65 (102.88)
Employees (000s)	44.69 (78.64)	52.28 (93.43)	58.65 (104.27)
Supply-side self-containment (%)	96.09 (5.32)	95.29 (7.59)	93.62 (9.65)
Demand-side self-containment (%)	96.83 (3.35)	96.47 (4.27)	94.96 (6.64)
		Sub-LMAs	
Employed residents (000s, before exclusions)	6.88 (18.59)	9.32 (23.65)	9.62 (24.48)
Employed residents (000s)	5.52 (14.98)	7.36 (19.03)	8.18 (20.94)
Employees (000s)	5.52 (16.41)	7.36 (20.76)	8.18 (22.67)
Supply-side self-containment (%)	69.59 (20.14)	68.94 (20.29)	67.43 (20.04)
Demand-side self-containment (%)	77.67 (14.72)	76.77 (14.45)	73.57 (16.93)

Note: Demand-side self-containment is the proportion of local jobs that are filled by local residents. Supply-side self-containment is the proportion of local residents who work locally.

Table 6: Attributes of LMAs identified using 2001 Census data

LMA	Number of area units	Land area (km²)	Employed residents (before exclusions)	Employed residents	Employees	Supply-side self-containment (%)	Demand-side self-containment (%)
1	33	6,678	19,437	14,640	14,556	97.54	98.10
2	66	7,148	43,599	34,782	34,254	95.27	96.74
3	379	2,900	512,778	412,389	416,019	99.18	98.31
4	1	277	399	324	318	93.52	95.28
5	25	2,067	19,992	15,369	13,137	71.15	83.24
6	108	9,923	98,670	77,649	77,232	96.62	97.14
7	44	13,062	19,509	14,799	14,619	95.13	96.31
8	21	3,439	18,603	14,388	14,025	93.16	95.57
9	29	3,574	23,514	18,951	19,395	92.72	90.60
10	53	2,042	54,450	43,104	42,621	96.76	97.85
11	41	4,454	29,517	22,758	22,869	96.01	95.54
12	23	5,518	17,298	13,416	13,524	97.29	96.52
13	32	12,463	21,429	16,170	16,206	99.17	98.94
14	69	8,664	60,378	46,764	46,899	99.19	98.9
15	79	11,892	73,614	60,480	60,174	96.72	97.21
16	185	2,182	191,289	158,307	159,588	99.29	98.50
17	62	7,175	45,264	37,284	37,320	99.51	99.41
18	34	3,142	18,540	14,526	14,292	94.22	95.76
19	23	6,785	18,309	14,832	13,947	90.25	95.98
20	23	9,125	19,503	15,321	15,267	97.94	98.29
21	53	11,372	40,047	30,894	30,909	99.24	99.19
22	9	7,777	3,885	3,114	3,123	97.21	96.93
23	55	23,244	13,944	11,262	11,250	98.91	99.01
24	182	12,486	190,740	155,136	155,157	99.67	99.65
25	3	438	846	675	675	91.11	91.11
26	1	794	381	237	237	100	100
27	16	6,183	13,314	10,851	10,908	96.54	96.04
28	47	19,435	33,015	26,553	26,451	98.71	99.09
29	84	8,415	59,946	48,612	48,588	99.14	99.18
30	28	18,675	17,496	13,803	13,854	98.48	98.12
31	73	33,502	46,896	37,977	37,944	99.00	99.09

Table 7: Attributes of LMAs identified using 2006 Census data

LMA	Number of area units	Land area (km^2)	Employed residents (before exclusions)	Employed residents	Employees	Supply-side self-containment (%)	Demand-side self-containment (%)
1	33	6,678	22,599	16,557	16,440	97.45	98.14
2	65	7,086	51,435	39,468	38,910	94.37	95.72
3	379	2,839	598,866	482,562	488,289	98.96	97.80
4	4	439	4,200	3,261	2,664	74.15	90.77
5	25	2,067	24,099	18,804	15,456	65.63	79.85
6	121	11,678	129,159	101,412	101,298	97.27	97.38
7	97	16,582	61,338	46,602	46,590	96.21	96.23
8	21	3,439	21,300	16,107	15,612	92.20	95.12
9	53	2,042	67,926	53,238	52,500	96.20	97.55
10	77	8,998	80,568	62,007	61,557	96.56	97.27
11	23	5,518	19,089	14,643	14,997	96.60	94.32
12	32	12,463	23,001	17,190	17,208	99.18	99.08
13	69	8,664	69,387	52,884	52,938	99.34	99.24
14	184	2,179	214,461	172,527	173,622	99.16	98.53
15	62	7,175	50,532	38,442	38,472	99.58	99.50
16	43	9,710	25,542	19,059	18,870	94.60	95.55
17	21	5,867	19,095	14,706	14,085	90.23	94.21
18	23	9,125	22,710	17,457	17,469	98.54	98.47
19	53	12,448	45,435	35,619	35,562	99.20	99.36
20	1	535	-	-	-	-	-
21	8	7,242	4,350	3,420	3,462	96.75	95.58
22	54	22,168	16,122	12,279	12,297	99.14	99.00
23	182	12,486	221,583	180,900	180,996	99.65	99.59
24	3	438	936	765	798	90.98	87.22
25	1	794	363	225	225	100	100
26	16	6,183	15,060	12,138	12,081	95.53	95.98
27	47	19,435	36,054	28,005	27,864	98.55	99.05
28	28	18,675	23,874	17,940	17,979	98.98	98.77
29	70	31,194	48,204	36,729	36,750	99.19	99.13
30	33	6,678	22,599	16,557	16,440	97.45	98.14

Notes: LMA 20 comprises the Lake Tennyson area unit, which has zero land area, and which had at least one employee or employed resident but fewer such people than the suppression threshold imposed by Statistics New Zealand for public data.

Table 8: Attributes of LMAs identified using 2013 Census data

LMA	Number of area units	Land area (km^2)	Employed residents (before exclusions)	Employed residents	Employees	Supply-side self-containment (%)	Demand-side self-containment (%)
1	84	12,499	61,323	50,949	51,285	98.10	97.46
2	13	1,244	10,236	8,613	8,100	73.18	77.81
3	378	2,808	618,825	527,790	535,041	98.49	97.16
4	1	22	33	27	30	88.89	80.00
5	5	442	4,308	3,738	3,162	72.71	85.96
6	24	1,986	25,215	21,225	16,698	60.04	76.32
7	22	3,521	20,367	16,752	15,900	88.14	92.87
8	121	11,678	133,137	113,613	112,923	95.83	96.41
9	88	12,607	53,829	44,058	44,247	95.43	95.02
10	10	4,022	3,456	2,865	2,820	92.67	94.15
11	53	2,042	69,801	58,701	57,822	95.39	96.84
12	23	5,518	17,604	14,583	14,907	95.78	93.70
13	32	12,463	21,339	17,781	17,808	99.17	99.02
14	69	8,664	65,376	54,585	54,651	99.40	99.28
15	119	18,727	97,749	84,723	83,961	97.54	98.42
16	62	7,175	51,936	44,742	44,838	99.68	99.47
17	183	2,110	215,829	190,038	192,222	99.35	98.22
18	1	2	-	-	-	-	-
19	21	5,867	19,683	16,716	15,270	88.17	96.52
20	23	9,125	21,921	18,138	18,000	97.25	98.00
21	55	12,449	46,200	38,376	38,244	98.72	99.06
22	9	7,777	4,434	3,762	3,777	94.98	94.60
23	54	22,168	16,032	13,545	13,620	98.98	98.44
24	186	12,925	228,018	194,193	194,514	99.55	99.39
25	1	794	348	267	267	100	100
26	63	25,617	53,421	45,201	44,961	97.21	97.73
27	87	10,723	65,706	54,231	54,162	98.88	99.01
28	28	18,675	26,433	21,792	21,726	98.50	98.80
29	72	31,195	48,078	39,933	40,011	99.29	99.10

Notes: LMA 18 comprises the Mana Island area unit, which had at least one employee or employed resident but fewer such people than the suppression threshold imposed by Statistics New Zealand for public data.

Table 9: Distribution of community counts across 250 runs, by Census year

	Communities					
Census year	Mean	Std. dev.	Min	Median	Max	LMAs
2001	31.35	1.03	28	31	33	31
2006	29.87	0.37	28	30	31	30
2013	28.78	0.88	26	29	30	29

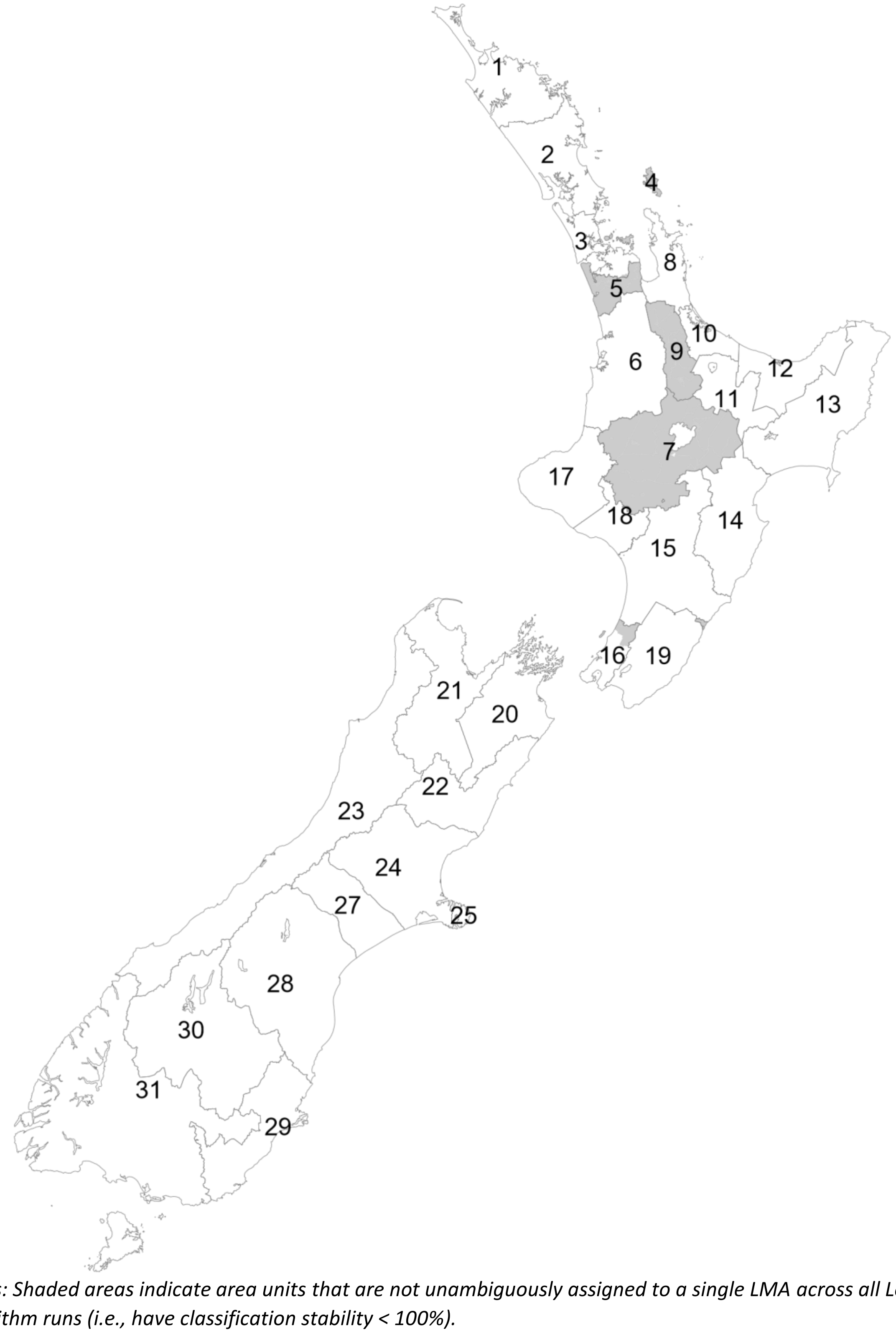
Table 10: Mean LMA allocation stabilities by Census year

Variable	2001 Census	2006 Census	2013 Census
		Stability	
Employee-weighted	99.30	97.57	98.68
Resident-weighted	99.27	97.29	98.57
Unweighted	98.88	96.97	97.97
		Stability = 100%	
Employee-weighted	96.28	86.37	96.07
Resident-weighted	96.06	84.58	95.67
Unweighted	94.26	83.26	93.80

Notes: "Stability = 100%" denotes the indicator variable for the event in each the LMA-specific stability measure equals unity.

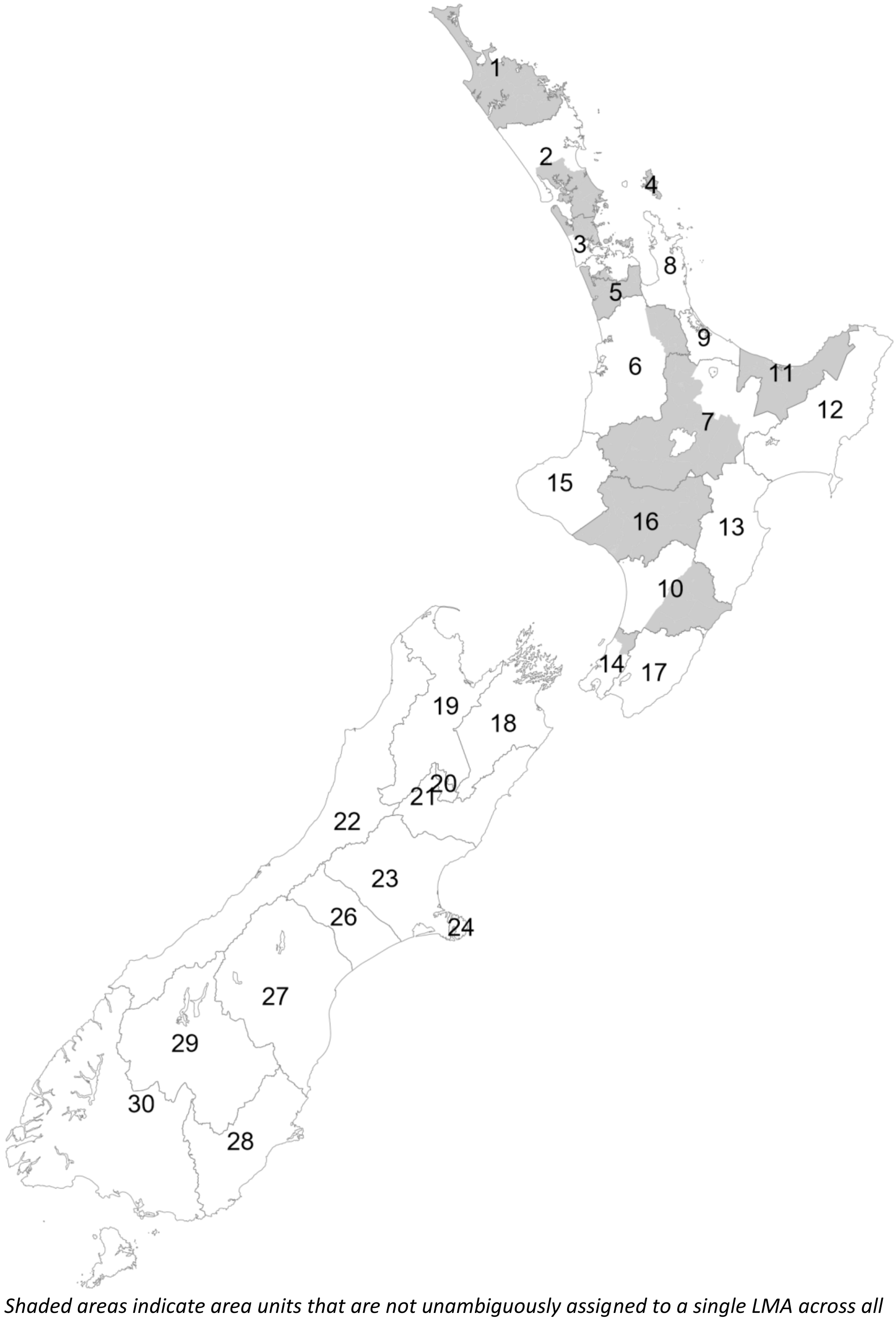
Notes: Boundaries delineate Regional Council areas. Darker area units have more employed residents per square kilometre.

Figure 1: Area unit employed resident densities, averaged over census years 2001-2013



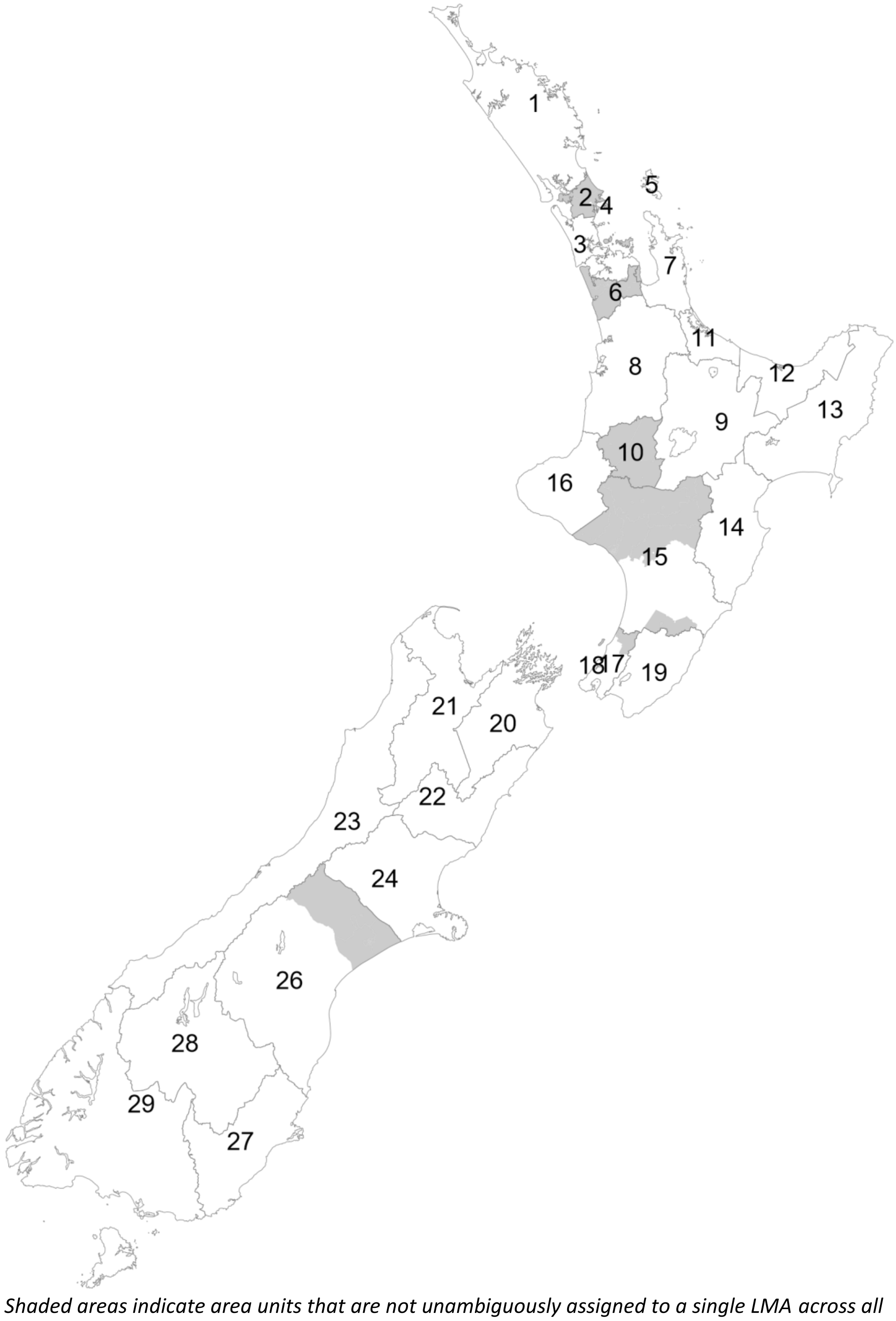
Notes: Shaded areas indicate area units that are not unambiguously assigned to a single LMA across all Louvain algorithm runs (i.e., have classification stability < 100%).

Figure 2: LMA boundaries using 2001 Census data



Notes: Shaded areas indicate area units that are not unambiguously assigned to a single LMA across all Louvain algorithm runs (i.e., have classification stability < 100%).

Figure 3: LMA boundaries using 2006 Census data



Notes: Shaded areas indicate area units that are not unambiguously assigned to a single LMA across all Louvain algorithm runs (i.e., have classification stability < 100%).

Figure 4: LMA boundaries using 2013 Census data

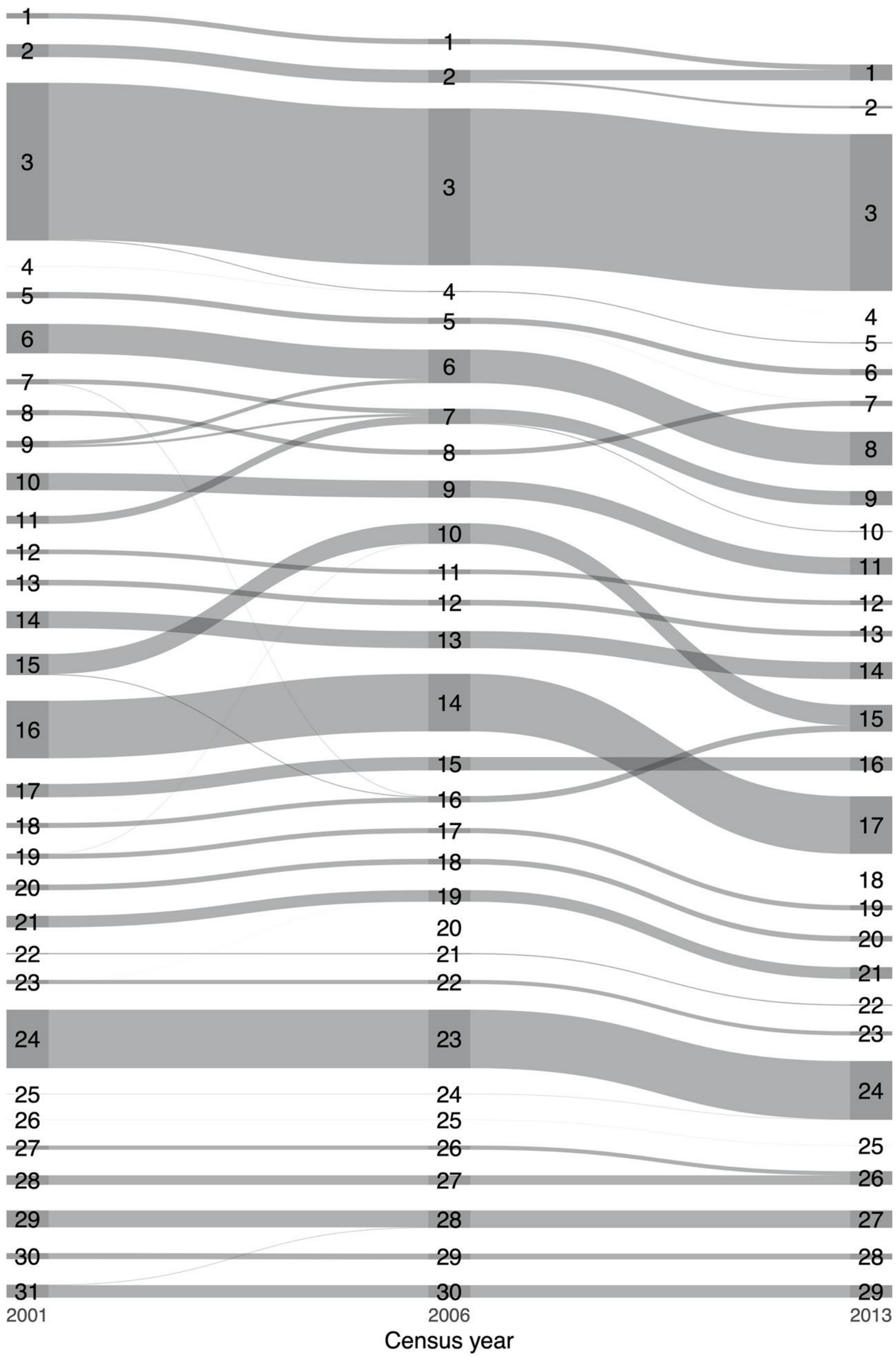


Figure 5: Employee-weighted LMA reconfigurations across census years

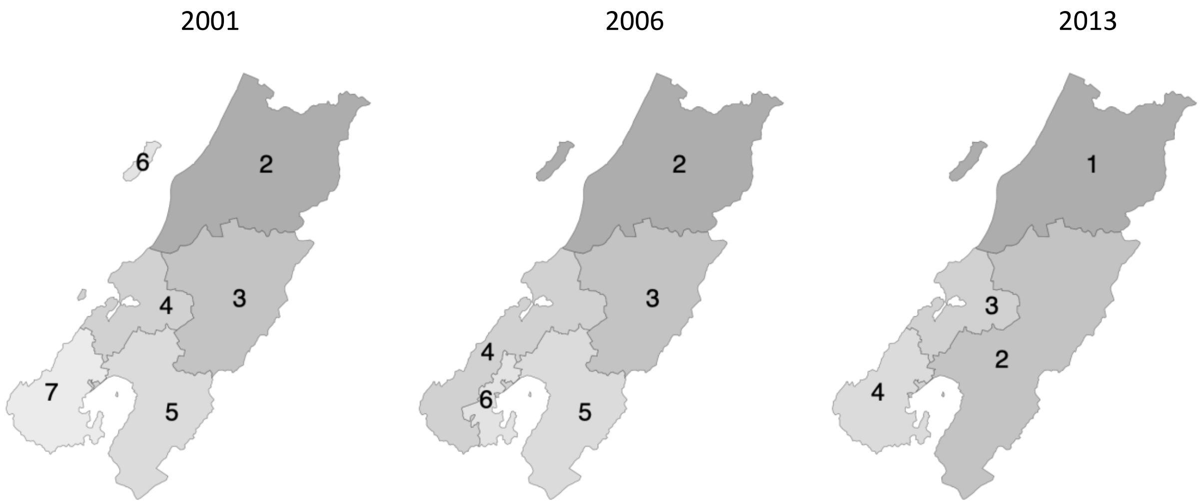


Figure 6: Sub-LMA boundaries in Wellington using 2001, 2006, and 2013 Census data

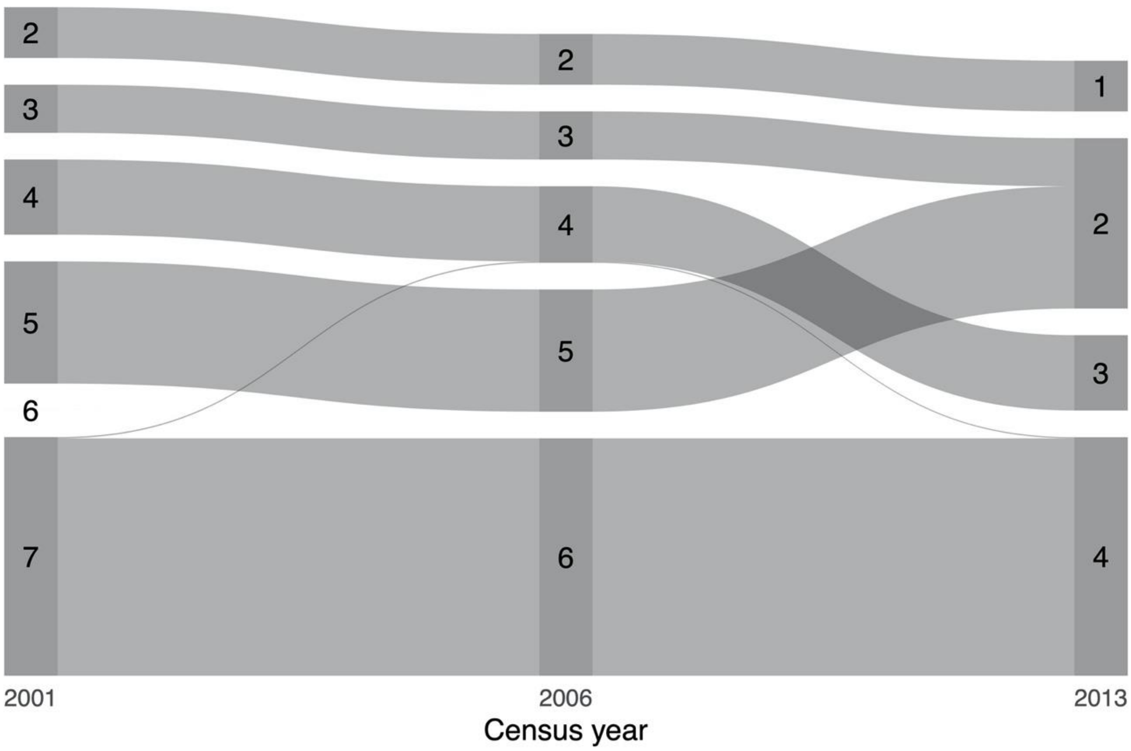


Figure 7: Employed resident-weighted area unit transitions among Wellington sub-LMAs