What's in a region? An algorithmic approach to regional delineation

Alasdair Rae[*1] and Ruth Hamilton[†1]

[1]Department of Urban Studies and Planning, University of Sheffield

January 12, 2018

Summary

This paper considers the question of what is 'in' a region from an economic perspective, based on commuting data. It follows in a long line of studies on the subject and builds on previous approaches to labour market delineation, including the well-known 'travel-to-work area' (TTWA). Using Combo, a network partitioning algorithm, we analyse commuter flow data from the 2011 Census in order to define a discrete set of regions for Scotland, as a test case. Our aim is to contribute to methodological advances in regional delineation and to put forward results that can have real-world utility for policy makers.

KEYWORDS: regions, delineation, algorithm, commuting, interaction

1. Background

In the field of regional studies, and cognate disciplines, the fundamental question of the scale and size of regions has been approached from a number of different perspectives. The topic also has a long lineage, with seminal contributions from Ohlin (1933), Lösch (1954) and Meyer (1963) amongst the most well known. This is testament both to the enduring importance of the subject matter and to periodic advances in methodological thinking on the subject. Regional delineation also retains a high degree of policy relevance, with recent regional reorganisations in France, Denmark and Poland suggesting there is a place for methodologically robust approaches to understanding what is 'in' a region and where the boundaries ought to be drawn if they are to retain some functional basis.

2. Our approach to regional delineation

In this paper, we seek to build on previous approaches to understanding regions from an economic perspective (e.g. Parr, 2008) through the use of an algorithmic approach to regional delineation. As we do so, however, we remain aware of the inherently political nature of the topic and the need to understand that however we define regions, issues of identity, culture, history and policy will inevitably come to the fore. We mention this as a reminder that any set of regions we may define - be they economic, policy-related or otherwise - can never perfectly capture the full complexity of the underlying nexus of social and economic interactions they seek to represent. Nonetheless, regions remain a critically important part of national policy infrastructures throughout the world and a robust and defensible approach to their definition is essential.

We aim to make a contribution along two lines. First, we hope to make a *methodological* contribution in relation to how economic regions can be defined, from 'the ground up'. Rather than starting with a series of nodes or using a set of spatial parameters on geography or population, our approach is based

[*] a.j.rae@sheffield.ac.uk

[†] r.hamilton@sheffield.ac.uk

on network partitioning (Sobolevsky, et al., 2014), using commuting data. It is based on the strength of connections between places rather than on their geographical proximity, though inevitably there will be a close relationship between the two.

Secondly, we aim to make a contribution with *a set of results which can have practical* use by informing ongoing debates about regional re-organisation in a policy context. We have selected Scotland as a case study, since there have been a number of reports and consultations on the possibility of reducing the current set of 32 local authority areas. Crucially, however, the Scottish Government have not yet taken any decisions on the form of a revised regional geography for local authorities. The recent experience of France, which in 2016 moved from a geography of 22 metropolitan regions to a new set of 13 highlights the fact that the geography of regions is not simply an arcane academic enterprise: it can have very important political and social implications.

3. Methods

Over the past 15 years, community detection has received a lot of attention in the field of network analysis (e.g. Newman, 2006) and several different algorithms have been proposed to address. While there is, as yet, no consensus on method or terminology, the key features of a network are clear: it contains a set of nodes connected by edges. If an edge represents a directional link, such as the flow of money between banks, it is said to be directed as opposed to being undirected. It may also be weighted if the link represents a value, for example volume of trade or value of transaction. A cluster or community is then defined as a group of nodes with a high density of connections between them and relatively few connections to nodes outside of the cluster. This can also serve as a definition of a functional region.

Community detection algorithms have been applied to a wide range of datasets including social networks, bank transactions, and even the Marvel Universe. Here we apply a general optimization method (Sobolevsky et al., 2014) to commuting data in Scotland, building on similar work done on the United States (Nelson and Rae, 2016). We use data from the 2011 Census which recorded both the place of residence and location of workplace for all respondents over the age of 16 and in employment at the time of the census.

The full technicalities of the process are more than can be described in detail here, so we have set out below the steps involved and refer to Sobolevsky et al.'s precise exposition of the method in their 2014 paper.

An overview of the Combo algorithm (Sobolevsky et al., 2014)

- **INPUT:** a network of origin-destination nodes is required. In our case, these were the origin and destination area centroids from the 2011 commuting dataset for Scotland. By default, at the outset, all nodes are assigned to a single community.
- *Step 1:* for each origin, destination pair Combo searches for the optimal community partitioning based on the best gain from moving a node from an origin to destination community (i.e. Combo iteratively tests a variety of network partitions for each node pair).
- *Step 2:* assign nodes from origin to destination community according to partition optimisation.
- *Step 3*: select optimal partitions based on best gains achieved by moving nodes to different partitions.
- *Step 4*: test optimality of final partitioning solution and if it is maximised, end. This final step ensures that the modularity score for the final partitioning is as high as possible.

- **OUTPUT:** the final output of Combo is the partition of a network of nodes *n* into a set of 'communities', which in this case is a set of labour market regions for Scotland, within which individual nodes are connected along individual 'edges' or vertices, *v*.

As the authors' note, 'the fulcrum of the algorithm is the choice of the best recombination of vertices between two communities' (Sobolevsky et al., 2014, p. 5). This is how we define the new regions to be presented in the full paper.

4. Results

We will present the results of our algorithmic partitioning of Scotland, which has produced a discrete set of 17 new 'regions', based on 2011 commuter data. We will discuss how these compare to previous regional delineations (e.g. TTWAs, Council Areas) and recent proposals relating to local government organisation. We will also raise the question of how far we should trust algorithms when it comes to regional reorganisations and the extent to which, as Smelcer and Carmel (1997) state, algorithms always produce the 'correct solution'. We do not agree, but we think there is much to be gained by taking an algorithmic approach.

5. Biography

Alasdair Rae is a Professorial Fellow in Urban Studies and Planning at the University of Sheffield. His work focuses on spatial data analysis and applied GIS as it relates to housing, neighbourhoods, deprivation, regions and commuting.

Ruth Hamilton is a Teaching Associate in Urban Studies and Planning at the University of Sheffield. Ruth has a background in biological computation and population dynamics. She leads the MSc in Applied GIS and was lead researcher on a recent JRF study on neighbourhood deprivation and disconnection in the UK.

References

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. Proceedings of the National Academy of Sciences, 99(12), 7821–7826.

Lösch, A. (1954). *The Economics of Location*. New Haven: Yale University Press.

Meyer, J.R. 1963. Regional Economics: A Survey. American Economic Review 53: 19-54.

Newman, M. E. J. (2006). Modularity and community structure in networks. Proceedings of the National Academy of Sciences, 103(23), 8577–8582.

Nelson, G. G. D., & Rae, A. (2016). An Economic Geography of the United States: From Commutes to Megaregions. PLOS ONE, 11(11), e0166083.

Ohlin, B. (1933). *Interregional and International Trade*. Cambridge: Harvard University Press.

Parr J. B. (2008) Cities and regions: problems and potentials, Environment and Planning A 40, 3009–3026.

Smelcer, J. B., & Carmel, E. (1997). The Effectiveness of Different Representations for Managerial Problem Solving: Comparing Tables and Maps. Decision Sciences, 28(2), 391–420.

Sobolevsky, S., Campari, R., Belyi, A., & Ratti, C. (2014). General optimization technique for high-quality community detection in complex networks. Physical Review E, 90(1), 012811.