Data driven functional regions

Carson J. Q. Farmer[1],

[1]National Centre for Geocomputation
National University of Ireland Maynooth,
Co. Kildare, Ireland
Email: Carson.Farmer@nuim.ie

1 Introduction

National scale changes in social, economic, and demographic variables are the result of a diverse range of interactions at the local, regional, national, and international level. Even within a single local area, there are a myriad of factors driving change and promoting stability and/or instability (Green & Owen 1990). The past several decades have seen increasing recognition that the standard administrative areas used by governments for policy making, resource allocation, and research do not provide meaningful information on actual conditions of a particular place or region (Ball 1980, Casado-Díaz 2000). As such, there has been a move towards the identification and delineation of functional regions, commonly based on the conditions of local labour markets (Smart 1974, Casado-Díaz 2000, Coombes 2002, Watts 2004). Ball (1980) suggests that the use of a spatial definition of local labour market conditions, such as functional regions, is of particular relevance for several reasons, including as a means of presenting information and statistics on employment and socio-economic structures, as well as assessing the effectiveness of regional policy decisions, and local government reorganisation.

Journey to work behaviour is widely cited as one of the most appropriate indicators of local labour market dimensions (e.g., Ball 1980, Gerard 1958, Vance 1960, Hunter 1969), and as a result, is often used to develop functional regions. As such, a functional region is often defined as a geographical region in which a large majority of the local population seeks employment, and the majority of local employers recruit their labour (Ball 1980, Coombes & Openshaw 1982, Casado-Díaz 2000). Functional regions are generally produced via a delimitation procedure which considers the direct and indirect relationships between regions by analysing the behaviour of individual commuters (Cörvers et al. 2008).

A number of regionalisation procedures have been suggested in the litterature (e.g., Masser & Brown 1975, Slater 1981, Coombes et al. 1986, Flórez-Revuelta et al. 2008), the most successful likely being that proposed by Coombes et al. (1986). The aim of this procedure is to define as many functional regions as possible, subject to certain statistical constraints which ensure that the regions remain statistically and operationally valid (Coombes & Casado-Díaz 2005). Similar definitions of functional regions has been employed in the Netherlands (Laan 1991) and other parts of Europe (Eurostat 1992), and more recently in Spain (Casado-Díaz 2000), and New Zealand (Papps & Newell 2002).

2 Issues

One thing that becomes immediately apparent when considering the regionalisation procedure of Coombes et al. (1986) and related algorithms is the seemingly arbitrary choice of threshold values. These threshold values will largely determine the size and number of functional regions defined, and can greatly influence the results of the regionalisation exercise. Others have argued against the use of situation-dependant absolute threshold values (e.g., Cörvers et al. 2008), or have suggested alternatives, such as using relative instead of absolute values (e.g., Laan & Schalke 2001). Despite these attempts, there remains few alternative methods in the geographical literature which do not rely to some degree on arbitrary criteria, and even fewer still being used in practise.

An additional problem with many functional regionalisation procedures is that they cannot be used directly for selecting the *number* of functional regions, k. Indeed many procedures require the value of k to be specified *a priori* Brown & Holmes (e.g., 1971), Masser & Scheurwater (e.g., 1980), Baumann et al. (e.g., 1983), Cörvers et al. (e.g., 2008), or determine k through the use of *ad hoc* assessments of the data. For example, many researchers have relied on subjective assessments of the configuration of functional regions, often based the authors' perceptions of local environments and specific application contexts to determine the optimal number of functional regions (Noronha & Goodchild 1992). As Noronha & Goodchild (1992) point out, it is difficult to maintain confidence in assessments of validity such as "These results agree *fairly well* with the regions used by the government for planning and statistics..." (Hollingsworth 1971,emphasis added by Noronha & Goodchild (1992)), or "The functional regions produced by the approach outlined above ... conformed *extremely closely* to an *intuitive knowledge* of the study area." (Brown & Holmes 1971,emphasis added by Noronha & Goodchild (1992)).

3 Alternative

Alternatives to standard clustering methods for interaction data include a range of network-based methods which have been developed in fields as diverse as computer science, sociology, biology, and physics (White & Smyth 2005, Newman 2006a,b). Early examples include the Kernigan-Lin algorithm (Kernighan & Lin 1970), spectral clustering (e.g., Ng et al. 2002, Verma & Meila 2003, Fischer & Poland 2004), as well as several hierarchical clustering methods (e.g., Murtagh 1983). One thing that many of these methods have in common is that they are designed to find the 'community structure' of a network, which is generally defined as the underlying clustering of nodes within the network. In this sense, community structure refers to the tendency for nodes in a network to form groups of high within-group edge connections, and low between-group edge connections (Newman 2004b). There are a range of applications for this type of cluster analysis, including finding groupings in social networks, describing the structure of communication and distribution networks, as well as finding functional regions within interaction data.

While many of the methods described above are fast, effective tools for partitioning a network dataset into relevant clusters, there are several shortcomings inherent in these methods which make them difficult to use for many applications. For example, the majority of these approaches attempt to balance the size of the detected clusters while minimising the interaction between groupings, which leads to a partitioning of the dataset into clusters of relatively equal size (Ng et al. 2002). While this may be optimal for parallel processing in computer science, many real world phenomenon do not exhibit this type of clustering, and as such, the optimal grouping in this type of data may not be found using spectral or hierarchical clustering methods. Furthermore, most spectral and hierarchical clustering algorithms are not able to provide an estimate of the number of clusters or grouping in a dataset, since they are designed to find groupings based on a predefined value of k.

Recently, a range of new clustering algorithms based on the 'modularity function' of Newman & Girvan (2004), have been developed (e.g., Newman 2004a,b, Clauset et al. 2004, Leicht & Newman 2008). These methods have in turn been further refined to take advantage of the speed of spectral clustering methods, while maintaining the benefits of the 'modularity function' Q of Newman & Girvan (e.g., White & Smyth 2005, Newman 2006a,b, Leicht & Newman 2008). This Q function directly measures the quality of a particular cluster arrangement, providing a means to automatically select the optimal number of clusters (or functional regions) k, in a network by choosing the cluster arrangement where Q is maximised. This has the direct benefit of eliminating the need for arbitrary cut-off values or parameters, as well as providing a concrete justification for the number of detected functional regions or clusters. Furthermore, spectral clustering algorithms that attempt to optimise the Q function will preferentially choose cluster arrangements where "there are many edges within [clusters] and only a few between them." (Clauset et al. 2004,p.066111-1). This definition of an optimal cluster arrangement agrees closely with the aim of most functional regionalisation procedures, and as such is particularly suited to our functional regionalisation problem.

4 Origins & destinations

For the present study, travel-to-work data was obtained for the entire Republic of Ireland from the Place of Work Census of Anonymised Records (POWCAR) from the Census of Population of Ireland 2006 (CSO 2006). This dataset contains anonymised geo-coded journey to work details (origin and destination) of all employed individuals in Ireland who regularly commute, as well as a range of demographic and socio-economic characteristics, including sex, age group, marital status, education, socio-economic group, employment type, and means of travel (CSO 2006). For our analysis, the origins and destinations are geo-coded to their corresponding electoral district (ED), and from this we are able to generate a generalised network of flows between each ED. In this way, we are able to examine the effects of different scocio-economic variables by changing the weights of the network edges to reflect the movements of particular employment- and person-types, such as 'highly educated female executives'. The allows us to develop *socio-economic functional regions*, which will help to provide a perspective on socio-economic influences on the local labour market. By subjecting the above commuting network to a spectral variant of the Newman modularity cluster algorithm (e.g., White & Smyth 2005, Newman 2006b, Leicht & Newman 2008), we will be able to highlight the underlying community structure of the travel to work network, effectively generating a range of functional regions based on the interconnections of the origins and destinations in the POWCAR dataset.

5 Acknowledgements

Research presented in this paper was funded by the Irish Social Sciences Platform (ISSP) under PRTLI4. Partial funding by a Strategic Research Cluster grant (07/SRC/I1168) by Science Foundation Ireland under the National Development Plan is also gratefully acknowledged.

References

Ball, R. M., 1980. The use and definition of Travel-to-Work areas in great britain: Some problems. *Regional Studies: The Journal of the Regional Studies Association*, *14*(2), 125–139.

Baumann, J. H., Fischer, M. M., & Schubert, U., 1983. A multiregional labour supply model for austria: The effects of different regionalisations in multiregional labour market modelling. *Papers in Regional Science*, *52*(1), 53–83.

Brown, L. A., & Holmes, J., 1971. The delimitation of functional regions, nodal regions, and hierarchies by functional distance approaches. *Journal of Regional Science*, *11*(1), 57–72.

Casado-Díaz, J. M., 2000. Local labour market areas in spain: A case study. *Regional Studies*, *34*(9), 843–856.

Clauset, A., Newman, M. E. J., & Moore, C., 2004. Finding community structure in very large networks. *Physical Review E*, *70*(6), 066111–1–066111–6.

Coombes, M., 2002. Travel to work areas and the 2001 census. *Report to the Office of National Statistics. Centre for Urban and Regional Development Studies*.

Coombes, M., & Casado-Díaz, J. M., 2005. The evolution of local labour market areas in contrasting regions. *ERSA conference papers, European Regional Science Association.*

Coombes, M. G., Green, A. E., & Openshaw, S., 1986. An efficient algorithm to generate official statistical reporting areas: The case of the 1984 Travel-to-Work areas revision in britain. *The Journal of the Operational Research Society*, *37*(10), 943–953.

Coombes, M. G., & Openshaw, S., 1982. The use and definition of travel-to-work areas in great britain: Some comments. *Regional Studies: The Journal of the Regional Studies Association*, *16*(2), 141–149.

CSO, 2006. Census 2006 Place of Work - Census of Anonymised Records (POWCAR). http://www.cso.ie/census/POWCAR_2006.htm.

Cörvers, F., Hensen, M., & Bongaerts, D., 2008. The delimitation and coherence of functional and administrative regions. *Regional Studies*, *Forthcoming*.

Eurostat, 1992. Study on employment zones. Tech. rep., Eurostat (E/LOC/20), Luxembourg.

Fischer, I., & Poland, J., 2004. New methods for spectral clustering. Technical Report IDSIA-12-04, Dalle Molle Institute for Artificial Intelligence, Manno, Switzerland.

Flórez-Revuelta, F., Casado-Díaz, J., & Martínez-Bernabeu, L., 2008. An evolutionary approach to the delineation of functional areas based on travel-to-work flows. *International Journal of Automation and Computing*, *5*(1), 10–21.

Gerard, R., 1958. Commuting and the labor market area. *Journal of Regional Science*, *1*(1), 124–130.

Green, A. E., & Owen, D. W., 1990. The development of a classification of travel-to-work areas. *Progress in Planning*, *34*(1), 1–92.

Hollingsworth, T. H., 1971. Gross migration flows as a basis for regional definition: an experiment with scottish data. In *Proceedings, IUSSP Conference, London, 1969*, vol. 4, (pp. 2755–2765).

Hunter, L. C., 1969. Planning and the labour market. In S. C. Orr, & J. Cullingworth (Eds.) *Regional and Urban Studies*. George Allen & Unwin, Glasgow.

Kernighan, B. W., & Lin, S., 1970. An efficient heuristic procedure for partitioning graphs. *Bell System Technical Journal*, *49*(2), 291–307.

Laan, L. V. D., 1991. *Spatial labour markets in the Netherlands*. Eburon.

Laan, L. V. D., & Schalke, R., 2001. Reality versus policy: The delineation and testing of local labour market and spatial policy areas. *European Planning Studies*, *9*(2), 201–221.

Leicht, E. A., & Newman, M. E. J., 2008. Community structure in directed networks. *Physical Review Letters*, *100*(11), 118703–1–118703–4.

Masser, I., & Brown, P. J. B., 1975. Hierarchical aggregation procedures for interaction data. *Environment and Planning A*, *7*(5), 509–523.

Masser, I., & Scheurwater, J., 1980. Functional regionalisation of spatial interaction data: an evalution of some suggested strategies. *Environment and Planning A*, *12*(12), 1357–1382.

Murtagh, F., 1983. A survey of recent advances in hierarchical clustering algorithms. *The Computer Journal*, *26*(4), 354–359.

Newman, M., 2004a. Detecting community structure in networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, *38*(2), 321–330.

Newman, M. E. J., 2004b. Fast algorithm for detecting community structure in networks. *Physical Review E*, *69*, 066133. Phys. Rev. E 69, 066133 (2004).

Newman, M. E. J., 2006a. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, *74*(3), 036104–19.

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

Newman, M. E. J., & Girvan, M., 2004. Finding and evaluating community structure in networks. *Physical Review E*, *69*(2), 026113–1–026113–15.

Ng, A. Y., Jordan, M. I., & Weiss, Y., 2002. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, *14*, 849–856.

Noronha, V. T., & Goodchild, M. F., 1992. Modeling interregional interaction: Implications for defining functional regions. *Annals of the Association of American Geographers*, *82*(1), 86–102.

Papps, K. L., & Newell, J. O., 2002. Identifying functional labour market areas in new zealand: A reconnaissance study using Travel-to-Work data. Discussion Paper 443, Institute for the Study of Labor (IZA), Bonn.

Slater, P. B., 1981. Comparisons of aggregation procedures for interaction data: An illustration using a college student international flow table. *Socio-Economic Planning Sciences*, *15*(1), 1–8.

Smart, M. W., 1974. Labour market areas: Uses and definition. *Progress in Planning*, *2*, 239–353.

Vance, J. E., 1960. Labour shed, employment field and dynamic analysis in urban geography. *Economic Geography*, *36*(3), 189–220.

Verma, D., & Meila, M., 2003. A comparison of spectral clustering algorithms. Technical Report UW-CSE-03-05-01, Department of CSE, University of Washington, Washington, U.S.A.

Watts, M., 2004. Local labour markets in New South Wales: Fact or fiction? Working Paper No. 04-12, Centre of Full Employment and Equity, The University of Newcastle, Callaghan NSW 2308, Australia.

White, S., & Smyth, P., 2005. A spectral clustering approach to finding communities in graph. In *Proceedings of the 2005 SIAM International Conference on Data Mining*, (pp. 274–286). Newport Beach, CA, USA.