COMMUNITIES IN COMMUTING NETWORKS

Carson J. Q. Farmer and A. Stewart Fotheringham

National Centre for Geocomputation,
National University of Ireland Maynooth,
Maynooth, Co. Kildare, Ireland

KEY WORDS: Networks, commuting, modularity, community structure, sub-populations, Local labour market

1 Introduction

One of the most promising approaches to mitigating land-use and transportation problems is research on urban commuting (Horner, 2004). One way of understanding the underlying structure of commuting, and how this structure may influence the observed patterns of commuting in large travel-to-work networks is by exploring the community structure of commuting networks. Community structure in a network exists when a network may be naturally divided in groups of nodes, where within-group edge connections are dense, and between-group edge connections are sparse. In the context of commuting, if the spatial nature of the network is considered, community structure may be used to delineate contiguous regions of relatively cohesive commuting patterns, or, what some might term travel-to-work areas (TTWA). These TTWAs provide planners and geographers with a socio-economic view of urban life, highlighting regions where people both live and work.

While analysis of aggregate commuting networks may provide useful insights into the structure and regularities of travel-to-work, disaggregate spatial networks, where attention is paid to the non-topological, non-spatial attributes of commuting, may provide further insights. When seeking to identify the community structure inherent in travel-to work-networks, the commuting patterns of population sub-groups often become averaged. This has the potential to hide important sub-communities which, through their interactions, lead to the observed aggregate commuting network. Therefore, in order to understand the structure of commuting as a whole, it is important to explicitly examine the spatial expression of network sub-groups.

2 Networks

One of the most widely used methods for quantifying the 'quality' of a particular partition of a network into communities is the modularity quality function, Q, of Newman and Girvan (2004). In geographical terms, modularity is based on the notion that a good division of a network into communities is one in which there are fewer than expected connections between communities (Newman, 2006a). The concept of modularity can therefore be used to find statistically surprising spatial interactions by measuring the difference between the total fraction of connections that fall within communities, and the expected fraction of such connections based on some null model (Newman, 2006a). This definition of an optimal cluster arrangement agrees closely with the aim and objectives of many geographical regionalisation procedures, and is particularly suited to locating communities in commuting networks.

The literature on spatial interaction suggests that there are limits on the distances individuals are willing to travel to bridge home and work (Singell and Lillydahl, 1986). It is therefore beneficial to factor into our equations the fact that longer commutes are less likely, providing a more nuanced approach to community detection. This can be done via a geographically weighted variant of the modularity function. This modification is designed to facilitate smaller, more compact functional regions by emphasising spatially local connections.

3 Assessing Stability

The validity of a regionalisation is undoubtedly related to the stability of its communities. A community is stable if it remains relatively invariant to random- or sampling-error and noise. Therefore, it is important to distinguish between communities which reflect the commuting patterns observed, and those generated as a result of data uncertainties or measurement error (Nemec and Brinkhurst, 1988). An effective method for assessing the stability of communities is via resampling methods (Hennig, 2007). There are a number of studies which utilise 'bootstrap resampling' (Efron and Tibshirani, 1993) to assess the stability of detected clusters, as well as to determine the 'true' number of clusters in a dataset (Felsenstein, 1985, Hennig, 2007). This type of analysis is particularly useful for spatial networks, as it explicitly considers the variability within the observed network and does not rely on any predefined distribution to assess significance. Our paper describes this procedure applied to network-based communities.

4 Application

For the present study, travel-to-work data were 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 commuting details of all employed individuals in Ireland who regularly commute, plus a range of demographic and socio-economic characteristics (CSO, 2006). The origins and destinations are geo-coded to their corresponding electoral district (ED), and from this we generate a network of commuting flows between all EDs. The network contains 3409 nodes and 222,484 non-zero edges, which form a sparse, weakly connected network. By assigning socio-economic attributes to the weights of the network edges, we can then characterise the movements of particular employment- and person-types, producing a range of different networks based on specific socio-economic factors. Following a more general analysis of aggregate commuting patterns, we examine the structure of disaggregate communities for males and females, as well as select socio-economic groups, two age-group aggregations, and private modes of transport, by subjecting the the set of sub-group commuting networks to a spectral variant of the Newman modularity cluster algorithm (Newman, 2006b).

Furthermore, we will examine the stability of said communities, to objectively assess the validity of the detected community structure.

5 Results & discussion

5.1 Aggregate commuting network

Figure 1 shows the community structure of the network using the proposed algorithm. There are a total of 65 contiguous regions with an average size of 1070.07 km^2, and a range of 57.44 to 5401.44 km^2. Modularity for the final regionalisation is 0.53. The radial hierarchy in Figure 2a represents the hierarchical structure of the communities given in Figure 1. Each subsequent stage divides the network into increasingly smaller communities, and provides some clues as to the effectiveness of the overall procedure. For instance, the initial split in the figure separates Dublin and surrounding regions from the rest of Ireland. This is not surprising, as the regions surrounding Dublin have been shown to be highly mono-centric in terms of employment (Vega and Reynolds-Feighan, 2008). The next split in the hierarchy separates the remaining EDs into two classes of communities: hub and periphery regions.

The stability of the above community structure was measured using a resampling scheme based on replacing a proportion k, of the network edges with random noise drawn from a Poisson distribution with λ equal to the initial observed edge weight. For the present analysis, k is fixed at 0.15 to simulate possible sampling errors in the POWCAR dataset. The distribution of stability values for the above communities is given in Figure 2b. Under the resampling scheme, only seven communities have a stability measure ($\bar{\gamma}$) less than 0.6 (with only one less than 0.5), suggesting that despite the added error in the dataset, the procedure is able to consistently extract the initial set of communities. The least stable communities are primarily associated with the periphery regions, where strong centralised commuting is less common. Indeed, 72.72% of the communities with $\bar{\gamma} < 0.80$ are contained within a single sub-group of the periphery regions, and are labelled in Figure 2a as Unstable.

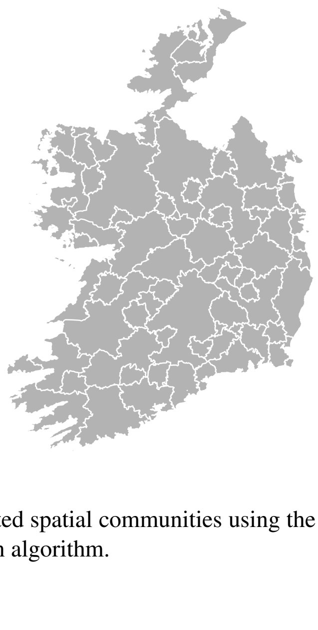
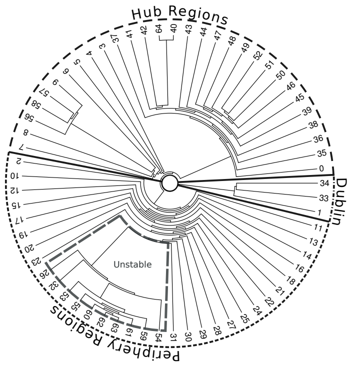
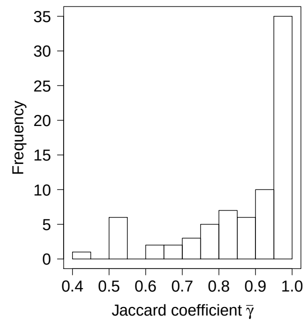


Figure 1: Detected spatial communities using the Newman community detection algorithm.



(a)



(b)

Figure 2: Radial hierarchy of the commutnity structure (a), and the distribution of stability values (b), for the community structure given in Figure 1. Note that stability values above 0.5 are considered 'stable'.

5.2 Population sub-groups

Examining the differences between various disaggregations of the POWCAR data into population sub-groups reveals additional commuting (sub)-structures. The sub-groups in this case are not mutually exclusive, and represent components of the population that are relevant to policy and resource allocation issues. Sub-groups considered include: 1) males, 2) females, 3) employers, managers, and higher and lower professionals (white-collar), 4) manual skilled, semi-skilled, and unskilled labourers (labour), 5) individuals with at least a third level degree or diploma (third-level), 6) individuals with no formal education, and/or no formal third level education (secondary), 7) individuals aged 15 to 39 years (young), and 8) individuals aged 40 years and greater (experienced). We also consider private means of transport separately from all other commuting types. Note that an individual's socio-economic group is determined by their occupation and employment status (see (CSO, 2006) for further information). Maps of community structure for the various sub-groups are given in Figure 3.

The analysis shows clear differences in the community structure of population sub-group commuting patterns. While several of these differences are predictable in nature (i.e., are strongly associated with periphery or hub regions in all sub-groups), others are less obvious at the aggregate level. It is therefore important to consider how the structure of aggregate commuting reflects the intricacies of sub-group commuting behaviour.

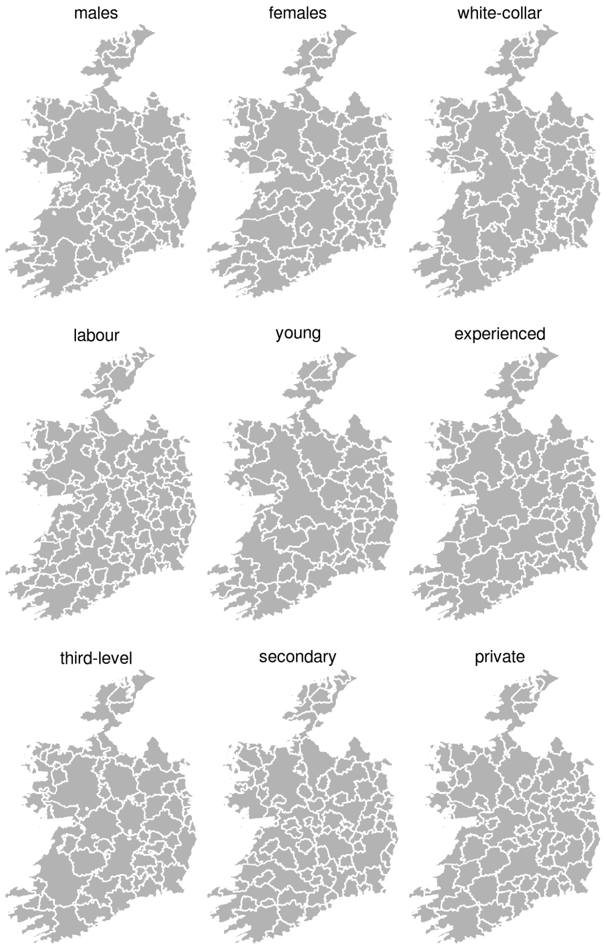


Figure 3: Detected functional regions for sub-group populations derived from the Irish commuting data.

5.3 Conclusions

The community detection procedure utilised in this paper treats commuting patterns as a network of flows in order to delineate spatial communities useful for representing local commuting patterns. Because the procedure uses the modularity quality function to determine the 'best' regionalisation for a given commuting network, it requires no *a priori* specification of community size or count, and is not reliant on the underlying population values or other external data sources. Furthermore, by weighting the commuting flows by geographical distance, the procedure presented here has been shown to find stable, spatially-constrained communities in a real-world geographical network. In addition, stability of the regionalisation produced via the proposed algorithm was tested using bootstrap resampling techniques designed to measure the effects of noise or random error on the algorithm's performance. Finally, the sub-group community structure maps presented here reveal important variations in sub-group commuting patterns not evident when simply using aggregate commuting data.

A main finding of this research is that attention should be paid to examining the 'stability' of these and previous regionalisation methods. The resampling strategy presented here is not limited to any particular community detection procedure, and can therefore be used to assess the stability of any procedure designed to find groups in network data. How stable are these communities, and how much confidence can one have in their boundaries? Would the same or similar communities be observed under different amounts of error or noise? These are important questions to consider when choosing a community detection algorithm, and statistically exploring these questions will ultimately provide the justification necessary for moving forward with a particular procedure.

We have shown that network-based partitioning methods lend themselves well to the delineation and analysis of spatial communities, and in many respects may offer benefits over legacy methods in the geographical literature. Furthermore, the theory underlying modularity provides an intuitive link to the concept of TTWAs and local labour markets, and as a result, methods which explicitly consider the modular nature of commuting data should be further explored. While no single set of communities, whether they be based on aggregate or disaggregate data, can capture the true structure of complex commuting patterns (Green et al., 1986), it is clear that structure does indeed exist, and that modularity provides an intuitive means of describing, and evaluating said structure.

REFERENCES

CSO, 2006. Census 2006 - Place of Work Census of Anonymised Records (POWCAR).

Efron, B. and Tibshirani, R. J., 1993. An introduction to the bootstrap. Chapman & Hall.

Felsenstein, J., 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39(4), pp. 783–791.

Green, A., Coombes, M. and Owen, D., 1986. Gender-specific local labour market areas in England and Wales. Geoforum 17(3-4), pp. 339–351.

Hennig, C., 2007. Cluster-wise assessment of cluster stability. Computational Statistics & Data Analysis 52(1), pp. 258–271.

Horner, M. W., 2004. Spatial dimensions of urban commuting: a review of major issues and their implications for future geographic research. The Professional Geographer 56(2), pp. 160–173.

Nemec, A. F. L. and Brinkhurst, R. O., 1988. Using the bootstrap to assess statistical significance in the cluster analysis of species abundance data. Journal Canadien des Sciences Halieutiques et Aquatiques 45(6), pp. 965–970.

Newman, M. E. J., 2006a. Finding community structure in networks using the eigenvectors of matrices. Physical Review E.

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

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

Singell, L. D. and Lillydahl, J. H., 1986. An empirical analysis of the commute to work patterns of males and females in two-earner households. Urban Stud 23(2), pp. 119–129.

Vega, A. and Reynolds-Feighan, A., 2008. Employment subcentres and travel-to-work mode choice in the Dublin region. Urban Studies pp. 1747 –1768.