Identifying fuzzy land use from network and graph theoretical approaches

Alexis Comber
Department of Geography
University of Leicester
Leicester, LE1 7RH, UK
ajc36@le.ac.uk

Masahiro Umezaki
Department of Human Ecology,
Tokyo University,
Tokyo, Japan
umezaki@humeco.m.u-tokyo.ac.jp

Chris Brunsdon
Department of Geography
University of Leicester
Leicester, LE1 7RH, UK
cb179@le.ac.uk

Abstract— **This paper describes the application of network science community finding techniques to networks of land cover objects in order to identify areas of land use. The results show that areas of homogenous land use can be identified and fuzzy land use determined using training data generated from a linear discriminant analysis. Further work will explore the application of different parameters and attributes to community identification algorithms.**

Keywords: network, graph. communities, land use, land cover

I. INTRODUCTION

In many remote sensing surveys, the concepts of 'Land Cover' and 'Land Use' are confused and treated as if they are the same thing in many surveys. Despite being illogical (land use and land cover are fundamentally different things), the confusion is convenient (Comber, 2008a) as surveys frequently have to satisfy a number of divergent stakeholder objectives. Fisher et al (2005) document the origins and nature of these confusions. Land use is difficult measure directly from remotely sensed data as land use describes *activity* rather than features on the earth's surface. Barnsley and Barr (2001) note that spectral radiance values recorded in remotely sensed data are only indirectly related to the attributes and dimensions of land use. As a result land use information is commonly inferred from land cover data. However land use and land cover rarely have one to one relationships: different land covers can constitute the same use (eg 'residential') and any given land cover type can be related to a number of uses (eg 'grass'). For many policy related application, land use is of greater relevance than land cover. For analyses that seek to identify and report on the spatial distribution of land use from remotely sensed data, the problem is how to translate land cover information that is recorded directly by the sensing instrument to land use information, whilst taking account of a number of considerations:

- the different land covers that are associated with any given land use (thematic);
- the impacts of different 'kernel' or 'window' sizes on aggregations of land cover and the land uses they infer (spatial, granular); and
- knowledge of the landscape and anthropogenic processes that result in specific cover / use combinations (temporal, knowledge-based).

The work reported in this paper explores and applies some network and graph theory techniques for community identification in order to generate land use information from land cover data.

II. BACKGROUND

A number of workers have sought to tackle the problem of identifying land use from remotely sensed data using secondary techniques based on the spatial configuration of land cover elements. Much of the relevant previous work and background to the problem of identifying land use from land cover can be traced to the work of Stuart Barr and Mike Barnsley in the 1990s. In a series of excellent papers Barr and Barnsley explored a number of techniques for determining areas of land use. Their early work analysed pixel based land cover data using a moving kernel to group clusters of pixels into discrete land use categories (Barr and Barnsley, 1996). Later work developed graph theory approaches to infer land use from the structural properties of and relations between land cover objects (vector regions) derived from Ordnance Survey LandLine base mapping (Barr and Barnsley 1997; Barnsley and Barr 1997). They developed an extended relational attribute graph model that was able to infer land use from the spatial pattern of land cover objects and their structural differences. This work was extended to consider the morphological properties of land cover derived from high-resolution satellite data (Barr and Barnsley 2000). Their corpus of work was summarised in 2004 and they concluded that different types of urban land use may be identified through analysis of the spatial disposition of their constituent land cover parcels and suggested that a quantifiable mapping exists between urban form (land cover) and urban function (land use). However this work was neither extended operationally nor into the more theoretical aspects of algorithm development.

Other work grounded in the remote sensing literature has applied alternative methods for shifting from land cover to land use. Herold et al (2002) used landscape metrics to describe urban land use structures and land cover changes. Their results showed that different urban land use types could be identified and land use changes quantified. Jansen and Di

Gregorio (2003) identified agricultural production systems by analysing field patterns and man-made structures. They found that the regular shape of land cover polygons indicated commercial production systems and irregular forms may indicate protective and conservation uses. Wastfelt (2009) developed a hybrid approach that combined spatial approaches with land use knowledge in an object oriented application. Comber (2008b) provided a method for quantifying the distributions of uncertainties associated with re-casting land cover as land use. Many of these approaches are reviewed in Lu and Weng (2007).

The work of Barr and Barnsley is especially relevant to this study as their work presaged two developments in spatial information science. First, the increased number of object-oriented techniques in remote sensing, led principally, but not uniquely, by the uptake of Definiens eCognition / Developer. Second, the expansion in computer science research of the development of network and graph theory approaches for identifying communities.

Object-oriented classification allows the user to manipulate groups of pixels (or objects) that have been segmented from image data. A hierarchical rule-based approach incorporating ancillary data is used to classify the segmented objects. Rules are encoded in a 'knowledge base', which specifies at what stage in the classification process they are applied. The advantages of the object oriented approaches are i) that the objects produced by segmentation may reflect more intuitively the features of interest on the ground compared to traditional pixel based classifications, and ii) that the knowledge base allows application of different rules to be manipulated in order to reflect heuristic manual classification procedures. In this way many authors have noted that object oriented approaches are able to better represent 'reality' as perceived by ecologists, field surveyors and air-photo interpreters than others remote sensing approaches (eg Lucas et al., 2007; Comber et al, 2010). For example Lackner and Conway (2008) apply knowledge incorporated into decision rules in object oriented classification software to identify land use. Blaschke (2010) provides a review of the recent literature on object based image analysis in remote sensing.

The outputs of object-based classifications are conceptually closer to community finding approaches in network and graph theory than pixel based classifications. The resulting objects have explicit and variable spatial structures (areas, perimeter length, etc), topological characteristics (ie spatial relations to other objects) as well as the thematic attributes also associated with pixel classifications. The land cover objects form nodes (vertices) in a network. They are connected to other nodes by edges (links) defined on their relationships with other nodes (commonly spatial but thematic relationships are definable). There has been a recent expansion of interest in network and graph theoretical models in computing and mathematical research which was partly due to the recognition of network ubiquity and interest in analyzing different types of networks for communities. Examples include academics who publish together, business organisational structures, cell phone networks, social networks (Facebook) etc. Additionally Girvan and Newman (2002) exposed graph-partitioning problems to the attention of the statistical physics and mathematics scientific communities. Porter et al (2009) comment that "Suddenly, community detection in networks became hip among physicists and applied mathematicians all over the world, and numerous new methods were developed to try to attack this problem" (p1083) and that "the study of what has become known as community structure is now one of the most prominent areas of network science" (p1084). The identification of communities from networks is the fundamental goal in such research and a number of techniques have been developed. Interestingly, Newman (2006) divides community finding research into two groups: "graph partitioning" in computer science whose main applications have been in parallel computing, and "community structure detection" sociology and applied sciences with many applications especially in social and biological networks. Newman notes that the main differences between the two approaches relates to the number of groups being known *a priori* in the former (the number of processors is usually known) and the objective of finding the best division. Whilst in community structure detection, the objective is data mining to identify communities *de novo* and to reveal community sub-structures under the assumption that the network divides naturally into subgroups. In this case the network itself determines the number and size of the groups. Porter et al (2009), Newman (2006) and Leicht and Newman (2008) provide excellent reviews of work in this area.

In summary, an extensive body of earlier research in spatial information science and remote sensing has indicated that graph or network theory may be appropriate for identifying regions of land use from groups land cover features. There are many techniques for identifying communities from networks, many of which are written into freely available libraries. The outputs of object based image analysis have many characteristics can be analysed using these network science techniques.

III. METHODS

A. Land Cover Data

A small area of Infoterra's LandBase © was provided for an area north of the city of Leicester in the UK (Figure 1).

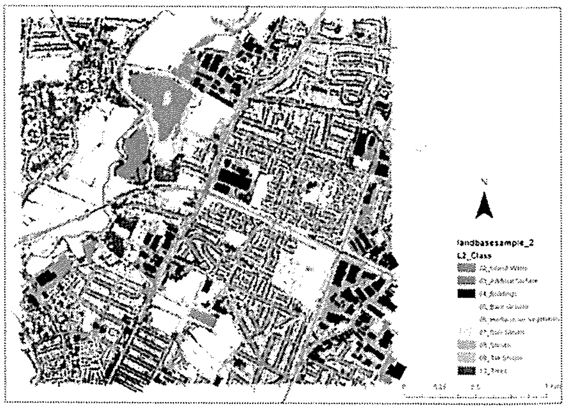


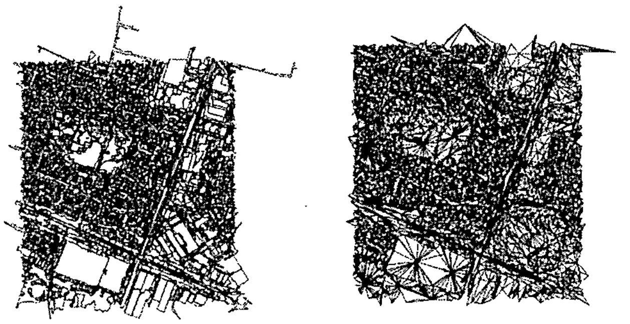
Figure 1. A sample of Landbase for the study area.

Landbase is constructed from an object oriented classification of a multiple layered image mosaic of Colour

Infra-red Imagery (CIR), Natural Colour Imagery (RGB), a digital surface model and a DTM. The segments are allocated to one of 10 land cover classes and each segment carries extensive contextual attributes describing the adjacent proportions of each class

B. Adjacency

The adjacency relations between objects for a sample of the data are shown in Figure 2.



Land cover objects   Adjacency

Figure 2. The derivation of adjacency in the network.

C. Community Identification in Networks

One of the most common approaches for identify network communities is to optimize modularity. Modularity is quality function that seeks to measure how well any given partition of a network groups its communities (Newman, 2004).

Modularity measures the difference between the total fraction of edges that fall within groups versus the fraction one would expect if edges were placed at random. High values indicate network partitions in which more of the edges fall within groups than one would expect by chance, that is a when a particular division of the network has more edge weight within groups than expected. Modularity functions provide measures of the total strength of connections within communities versus those between communities. Communities were identified by optimizing modularity, Q, for the undirected network of land cover objects:

$$Q=\frac{1}{2m}\sum_{ij}\left[A_{ij}-\frac{k_i k_j}{2m}\right]\delta(c_i c_j) \qquad (1)$$

Where m is the number of edges, Aij is the element of the A adjacency matrix in row i and column j, ki is the degree of i, kj is the degree of j, ci is the type (or component) of i and cj that of j.

D. Analysis

Each land cover object has attributes of local cover statistics that describe the proportions of the ten different level 2 classes that are within 50m of the object. These 'neighbourhood' attributes characterize what surrounds each object and are used as inputs to the community finding analysis.

There are a number of variations to the method that relate to a) the thresholds for identifying discontinuities and b) the distance measures used in community identification. Figure 3 shows the impact of 3 different discontinuity thresholds and Figure 4 shows the number of communities identified by different distance measures at different community partitioning thresholds. It is noteworthy that the distance measures identify approximately the same number of communities at any given threshold.

T = 0.004   T = 0.4

Figure 3. The impact of different thresholds for determining community discontinuities.

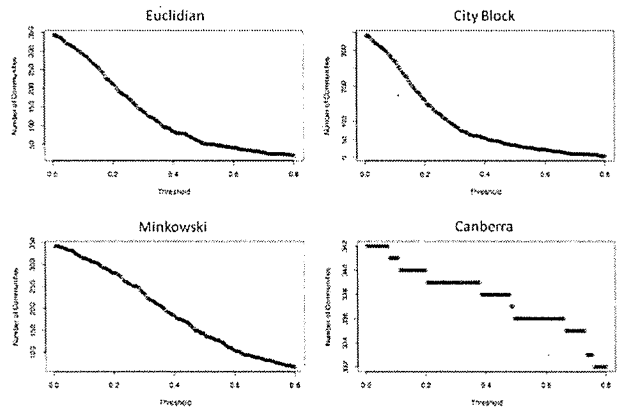


Figure 4. The number of communities identified at different thresholds using different distance measures.

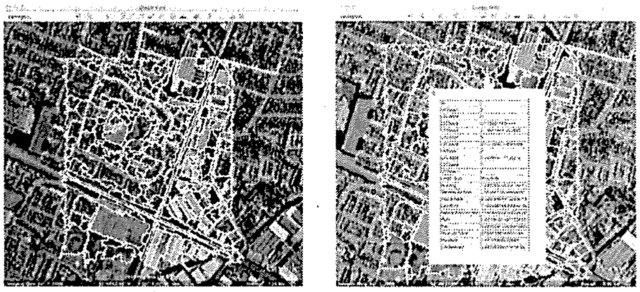
The analysis proceeds in the following way:

- For each land cover object a list of neighbours is constructed based on shared boundaries and converted into an adjacency matrix.
- The attributes for each object are analysed to identify discontinuities and thereby to identify communities within the network. This is done using the distance attribute feature space between pairs of objects (ie whose coordinates are *neighbourhood* proportions of classes). This example used a threshold of 0.004.
- Objects with 'close' attribute profiles (ie similar land cover class attribute proportions) are merged together.
- The proportions of different land cover classes in each community are used to characterise the land use and compared with validation data generated for another area by overlaying LandBase with Ordnance Survey's MasterMap®.

The communities identified in this way are sets of nodes (land cover objects) whose interconnection weights are significantly higher than those to other parts of the network.

IV. RESULTS

The communities of land use that are identified are homogenous in terms of their 'neighbourhood' attributes and adjacency. That is, their *within* community relations and links are stronger than their *between* community links (Figure 5). They are attributed with the area-weighted proportions of the different LandBase L2 land cover classes that they contain. The attributes are compared with training data that predicts the relationship between LandBase L2 attribute proportions and OS MaterMap 'Descriptive' land use classes from a linear discriminant analysis. The posterior probabilities for each MasterMap class give an indication of the fuzziness for each land use class.



a) b)

Figure 5. a) the land use communities identified with b) an example of the attribution, predicted class and posterior probabilities.

V. CONCLUDING REMARKS

The results show that 'communities' of land use can be identified from compositions of land cover objects and that network science approaches to community detection have much to offer in this area. Equally, the domain of geographical information provides a rich arena to test and develop community finding algorithms, as the results can be compared to reality on the ground. Porter et al (2009, p1096) note that "There are almost no theorems, and few methods have been developed to use or even validate the communities that we find".

The reliability of the 'results' will be strongly influenced by selection of the training data and whether comparing aggregated Landbase objects with MasterMap is reasonable or like comparing apples and oranges. Future work will develop this analysis to accommodate a number of factors that have an influence on the validity of the results: the specification of thresholds, different distance measures in attribute feature space and alternative training data.

ACKNOWLEDGEMENTS

Carson Farmer (National University of Ireland, Maynooth), Andrew Tewkesbury and Infoterra, OS MasterMap Topography Layer [GML geospatial data], Coverage: Leicester, Ordnance Survey, GB. Using: EDINA Digimap Ordnance Survey Service, http://edina.ac.uk/digimap, Downloaded: March 2010.

REFERENCES

Barnsley, M.J. and Barr S.L, 1996. Inferring urban land use from satellite sensor images using kernel-based spatial reclassification. *Photogrammetric Engineering & Remote Sensing*. 62(8), 949-958.

Barnsley, M.J. and Barr S.L, 1997. Distinguishing urban land-use categories in fine spatial resolution land-cover data using a graph-based, structural pattern recognition system. *Computers Environment and Urban Systems*. 21(3/4), 209-225.

Barnsley. MJ and Barr SL (2001). Monitoring urban land use by earth observation. *Surveys in Geophysics*. 21, 269–289, 2000.

Barr, S.L. and Barnsely, M.J., (1997). A region-based, graph- theoretic data model for the inference of second-order thematic information from remotely-sensed images. *International Journal of Geographical Information Science*. 11(6), 555- 576

Barr, SL and Barnsley, MJ, (2000). Reducing structural clutter in land cover classifications of high spatial resolution remotely-sensed images for urban land use mapping. *Computers & Geoscience*. 26, 433-449

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*. 65(1), 2-16

Comber A.J., (2008a). Land cover or land use? *Journal of Land Use Science*. 3(4), 199–201.

Comber, A, Medcalf, K, Lucas, R, Bunting, Brown, A, P Clewley, D, Breyer, J and Keyworth, S, (2010). Managing uncertainty when aggregating from pixels to objects: context sensitive mapping and possibility theory. *International Journal of Remote Sensing*. 31(4), 1061-1068.

Comber, A.J., (2008b). The separation of land cover from land use with data primitives. *Journal of Land use Science*. 3(4), 215–229

Fisher, P.F., Comber, A.J., and Wadsworth, R.A., (2005). Land use and land cover: contradiction or complement. in Peter Fisher, David Unwin (eds). *Re-Presenting GIS*, (pp. 85-98). Chichester: John Wiley.

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

Herold M., Scepan J. and Clarke, K.C., (2002). The use of remote sensing and landscape metrics to describe structures and changes in urban land uses, *Environment and Planning A*, 34(8): 1443-1458.

Jansen, LJM and Di Gregorio, A, (2003). Land-use data collection using the "land cover classification system": results from a case study in Kenya. *Land Use Policy*. 20, 131–148

Lackner, M. and Conway, T.M., (2008). Determining land-use information from land cover through an object-oriented classification of IKONOS imagery. *Canadian Journal of Remote Sensing*, 34(2), 77-92

Leicht, EA and Newman, MEJ (2008), Community structure in directed networks, *Physical Review Letters*, 100: 118703

Lu, D. and Weng, Q., (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5), 823-870

Lucas, R., Rowlands, A., Brown, A., Keyworth, S. and Bunting, P. 2007. Rule-based classification of multi-temporal satellite imagery for habitat and agricultural land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*. 62(3), 165-185.

Newman, MEJ (2004). Finding and evaluating community structure in networks. *Physical Review E*. 69, 026113

Newman, MEJ (2006). Modularity and community structure in networks, *Proceedings of the National Academy of Sciences*. 103, 8577–8582.

Porter, MA, Onnela, J-P and Mucha, PJ, (2009). Communities in networks. *Notices of the AMS*. 56(9), 1082-1166.

Wastfelt, A., (2009). Land use qualities identified in remotely-sensed images, *International Journal of Remote Sensing*. 30(9), 2411-2427.