From spatial interaction data to spatial interaction information? Geovisualisation and spatial structures of migration from the 2001 UK census

Alasdair Rae *

Department of Town and Regional Planning, University of Sheffield, Western Bank, Sheffield, S10 2TN, United Kingdom

1. Introduction

‘Knowledge is always gained by the orderly loss of information.’

Kenneth Boulding (1970, p. 2)

The task of mapping migration, or any other kind of spatial interaction data, has preoccupied a select number of researchers over the years. Whilst Waldo Tobler has been the most prolific member of this group, the origins of such an approach do of course extend back in time to Ernst Ravenstein and his 1885 paper on the subject of ‘The Laws of Migration’. The topic is revisited in the literature at points which coincide with advances in computer techniques but we have not yet arrived at a point where standardised flow mapping techniques are in the public domain. In fact, point-to-point geovisualisations of spatial interaction data are few and far between and therefore do not properly illustrate the inter-linkages between places that would allow us to better understand such themes as urban spatial structure, rural–urban interaction or large volume, long-distance flows. Many contemporary illustrative approaches at mapping spatial interaction data are nonetheless important in their own right (e.g. Nielsen & Hovgesen, 2008) but it remains true that the vast riches of national census data tables for migration and commuting (not to mention other datasets relating to business relocation or civil aviation), are largely locked away from view. At present, they remain the domain of specialist analysts in university research centres, geography departments or urban planning departments. This situation stands in marked contrast to developments in dynamic geovisualisation more generally, where there have been groundbreaking advances in the presentation and investigation of multivariate spatio-temporal data (e.g. Guo, Jin, MacEachren, & Liao, 2006; MacEachren, Wachowicz, Edsall, & Haug, 1999). The contention here is that through new geovisualisation techniques very large spatial interaction datasets can be mapped, analyzed, queried and supplemented in order to increase practitioner understanding of the ways in which local areas are connected to all other areas. Put simply, it is now possible to produce at-a-glance geovisualisations that begin to unlock the potential of datasets traditionally closed to public view. Before we can begin to derive new knowledge from spatial interaction data, then, we must first find ways to convert data into some kind of intelligible information. Such an approach is described herein.

This paper is comprised of four sections, in addition to the final concluding section. First, there follows a discussion of advances in relation to flow mapping; that is, mapping of data with both an origin and destination point in geographic space. In this section, the...
historical lineage from Ravenstein to the present day is covered and the theoretical underpinnings outlined in more detail, including more recent advances in dynamic geovisualisation techniques. The second section provides an overview of the approach to mapping a large migration dataset with proprietary GIS software; in this case the 2001 ward-level migration dataset from the United Kingdom census. The third section presents the results from the foregoing technical procedure and illustrates them in the form of a series of exemplary geovisualisations that depict the results of different queries entered into the GIS. A fourth section considers how we can move from data to information and considers online dissemination as the most appropriate vehicle (a project website).

Kingdom census. The third section presents the results from this case the 2001 ward-level migration dataset from the United

2. Advances in mapping spatial interaction data

Contemporary techniques in geovisualisation (e.g. Guo, Gahegan, MacEachren, & Zhou, 2005; Yan & Thill, 2007) have been brought to the fore in recent years owing to advances in technology, but the essential aim remains the same: ‘to turn large heterogeneous data into information (interpreted data) and subsequently, into knowledge (understanding derived from information)’ (Hernandez, 2007, p. 249; see also Carr, White, & MacEachren, 2005; Guo et al., 2005). Fundamentally, then, geovisualisation is concerned with what might be termed ‘geographical ontology’ in that it seeks to discover ‘what exists’ from a spatial perspective, or, more specifically, ‘what exists where’ (see Goodchild, Yuan, & Cova, 2007; Longley, Goodchild, Maguire, & Rhind, 2005, pp. 82; Galton, 2003 MacEachren, 1994; MacEachren et al., 2003). This type of question is embodied in Ravenstein’s, 1885 study where he lays down ‘The Laws of Migration’ for the United Kingdom and produces a series of 12 maps illustrating the various characteristics of residential mobility in the counties of the UK. The fifth such map, entitled ‘Currents of Migration’ (Ravenstein, 1885, p. 183) is of most interest since it displays the movement of people around Great Britain and Ireland by means of a series of single-headed arrows, crossing county boundaries and typically flowing towards major urban centres. Despite there being no mention of it in the paper itself, it is the first example of the kind of geovisualisation proposed here and the seminal reference upon which subsequent approaches have sought to build.

One has to move forward 74 years to arrive at the first examples of computer-based flow mapping, produced by the Chicago Area Transportation Study (1959) and using a specially constructed cathode ray tube system. In keeping with the pragmatic and rational planning ethos of such approaches (see Black, 1990), the results of this study were used to support decision making on the location of new interstate highways in the Chicago area (Tobler, 1987, p. 156). Kern and Ruston (1969) then took the next significant step in their approach of plotting single lines on a map with the aid of a dedicated computer program. It was in the fields of transport engineering and transport planning more specifically, however, that approaches to mapping dynamic flows developed next. As Tobler (1987) notes, Wittick (1976), Noguchi and Schneider (1977), McLaughlin (1977) and Beddow (1978) all made valuable contributions to the subject, both from a broadly-conceived geographic ontological standpoint and in terms of exploring the epistemological terra incognita of flow mapping. Tobler then advanced the subject matter further, with his ‘model of geographic movement’ (Tobler, 1981), his ‘experiments in migration mapping’ (Tobler, 1987), finally culminating in a user-friendly, dedicated Windows-based flow mapping program in Tobler (2003) (see also Tobler, 1995). In addition to Tobler’s Flow Mapper application, collaboration between Utrecht University and Gadjah Mada University in Indonesia resulted in the production of a dedicated Flowmap application in 1990 (see Geertman, de Jong, & Wessels, 2003). This was initially a DOS-based package, although it now operates on a Windows platform and has been successfully used in a number of studies (e.g. de Jong & van Eck Ritsema, 1996; Ritsema van Eck & de Jong, 1999). Despite the merits of the Utrecht approach, and its compatibility with proprietary GIS file formats, it is not yet widely used. Other applications that allow for some form of flow mapping include MapInfo (via the MapBasic window using the ‘creatable’ command) Mapititude, MapViewer, GRASS, BusinessMap 4.5, the ET GeoWizards extension for ArcGIS 9.x, transportation planning software such as TransCAD, and a bespoke Flowmap tool developed by researchers at Stanford University (Phan, Xiao, Yeh, Hanrahan, & Winograd, 2005).

Other significant contributions have also resulted in advances in the field, though once again it seems the GIS community at large have been slow to embrace these developments. This situation may be changing, so that the work of Gou (1993), Liu (1994, 1995) and Marble, Gou, and Liu (1995) from the early to mid-1990s, and more recent work by Yan and Thill (2007, 2008, in press) can now be situated within a broader geovisualisation paradigm which includes truly dynamic approaches of the kind developed by Carr et al. (2005), Gahegan (2000, 2001) and Gahegan, Takatsuka, Wheeler, and Hardisty (2002). Carr et al.’s ‘conditioned choropleth map’ approach and other similar work by MacEachren and allied researchers (e.g. Guo et al., 2005, 2006; MacEachren, 2001; see also Wachowicz, 2001) represents the kind of advance in geovisualisation that can truly bridge the gap between data and knowledge and provide new research questions which hitherto might have remained unasked. At this stage, however, flow map functionality and development remains some way behind these advances.

Within the leading proprietary GIS software packages flow map functionality has remained unusually underdeveloped, even during the last ten years of rapid growth in the tools we have at our disposal and despite the size of the GIS user community. Furthermore, awareness and uptake of the tools available is also less than optimal. For example, in an ArcView-Listserv comment in October 1998, a highly respected and renowned GIS contributor commented that: ‘we write our own software. Trying to do flow mapping in [proprietary GIS] products is just not viable’ (Marble, 1998; see also Marble, Zaiyong, Lin, & Saunders, 1997). However, it was during this period that flow map functionality actually became more widely available in ArcView with O’Malley’s (1998) Desire Line Maker; an extension freely available from ESRI’s ArcScripts website which can produce flow maps from either a combination of a shapefile and data table or a data table with origin and destination coordinate information. Despite the availability of this tool, its evident usefulness to our understanding of spatial dynamics and spatial structures, and the subsequent development of others like it, by 2008 relatively little progress had been made (cf. Holland & Plane, 2001, p. 90). This was exemplified in another online posting, where the contributor states that ‘unfortunately, flows and geodynamics are not very well represented in modern GIS’ (Gibin, 2008). In essence, then, developments in flow mapping seem somewhat stunted when compared to developments in geospatial technology more generally.

Notwithstanding Tobler’s massive individual contribution to the mapping of spatial interaction data, and the potential utility of point-to-point ‘spider diagrams’ (Davis, 2001, p. 275), Dorling’s work on the visualisation of spatial structure (Dorling, 1991) stands alone as the other major contribution to the field. In this re-

spect he was clearly well ahead of his time in the implementation of such techniques, although they remain more a testament to individual brilliance than they do to the development of methods to support planning or public policy decision making, remaining very much in the domain of the technical specialist until flow map functionality was made available in proprietary GIS software. In addition, and as a direct extension of Tobler’s earlier work, flow map functionality was made available in ArcGIS with the development of Glennon’s Flow Data Model Tools in a series of ArcGIS 9 Visual Basic for Applications (VBA) macros (Glenon, 2005). At this stage, then, the potential user community for such approaches reached an all time high, given the integration of the techniques into a leading proprietary GIS package (rather than a stand-alone application), and the growing dominance of ArcGIS in many nations. However, the approach remains under-utilised and its potential still needs to be developed if the path from spatial interaction data to spatial information, knowledge, and even wisdom is to be made clear (Longley et al., 2005, p. 13).

3. An approach to mapping a large migration matrix

The intention now is to provide a general overview of the procedure used to produce geovisualisations of a large spatial interaction dataset, in addition to some further technical detail. In keeping with the desire to promote the approach more widely, the Flow Data Model Tools (FDMT) for ArcGIS were used as the mapping engine within ArcMap 9.0, whilst the specific dataset used was the ward-level ‘age by sex’ table (MG201) from the United Kingdom’s 2001 census (see Stillwell & Duke-Williams, 2007). It is important to emphasise that the methods adopted are not predicated upon the use of any pre-formatted or country-specific dataset but rather are flexible enough to allow for the integration of any data structure or format. In common with most GIS procedures, a degree of pre-processing was required and in this regard the MG201 data provided a robust test case which others could logically follow. The method outlined below is therefore the first step towards unlocking the knowledge contained within a vast matrix of data that represents the spatial interactions and household mobility patterns of millions of individual people (the MG201 table is a 10,608 × 10,608 interaction matrix).

After some initial pre-processing of the dataset, the procedure of flow mapping could then commence using Glennon’s FDMT. When the flow line tool is activated, it is possible to calculate two-way, net, or gross flow lines and in the case of the MG201 data two-way was chosen since there are typically flows in both directions between destinations and origins and this method provides a geovisualisation which most accurately reflects the original data-set, whereas the other options involve some loss of information relating to directionality either through addition of flows (gross) or subtraction (net). In order to facilitate replication, the individual steps followed in the procedure are detailed below.

1. Enter name of ‘output file’ for new flow shapefile (excluding file suffix).
2. Select ‘flow type’ (two-way, net or gross).
3. Specify the ‘points layer name’ containing the origins and destinations.
4. Specify the ‘points ID field name’ (so that the FDMT can match the points layer in step 3 to the spatial interaction data matrix in step 6).
5. Specify the ‘shape field name’ (typically, this will simply be ‘Shape’).
6. Specify the ‘flow table name’ (i.e. the dbf flow data table).
7. Specify the ‘from point ID field name’ (i.e. the origin).
8. Specify the ‘to point ID field name’ (i.e. the destination).
9. Specify the ‘magnitude field name’ (e.g. the number of migrants).
10. Finally, initialise the FDMT by clicking ‘OK’.

Owing to the large size of the MG201 dataset (in total, 851,318 rows), the processing power and time required to run the FDMT in ArcGIS 9.0, and the in-built restriction on the number of rows it can comfortably handle on a standard desktop computer (~30,000 on a 3.20 GHz machine with 2 GB RAM), the original MG201 dbf was divided into 43 smaller parts with a maximum size of 20,000 rows (about 5 MB each). Using a desktop PC with these specifications, the mapping of 20,000 flows took approximately 80 minutes with no other applications running. Since the flow mapping project here is deliberately ambitious in scale and uses the MG201 dataset, the procedure of mapping all 43 flow tables would take more than 57 h, thereby making the process rather onerous, and working against the original goal of replicability and practicality. As an alternative, and in order to significantly speed up the process, the FDMT was re-initialised using a RAM disk, a software abstraction that treats part of the random access memory (RAM) as if it were a hard drive and appears as a new physical disk to the operating system (in this case Windows XP) and the application. In simple terms, a RAM disk is a small application that can be downloaded and run on any standard Windows-based desktop computer. When it is running, a new drive letter appears within the operating system environment and the FDMT tool can simply be directed to process on that drive, rather than the default C: \. Since disk access time is significantly improved using this approach, the mapping of a 20,000 row flow table took only 9 minutes; a highly significant time saving. Thus, even though the approach outlined here is based on one of the largest tables available from the UK national census, the mapping of a spatial interaction table with more than 850,000 flows is possible in only one working day with standard GIS software.

After completing 43 iterations of the flow mapping procedure, all 43 resultant flow line shapefiles were merged together using the Append tool in ArcToolbox, culminating in a single shapefile containing 851,318 individual lines and an attribute table with a column containing a unique identifier for all pairs of origins and destinations and one column containing the total number of migrants. In order to make subsequent analysis more powerful and insightful, a series of joins were performed between the attribute table of the new shapefile and separate, imported dbf tables containing further information relating to each origin and destination. Additionally, the straight line distance for each flow line was calculated using a simple VBA script. The fields contained in the final attribute table are summarised in Table 1. With all the information now contained in the shapefile’s attribute table, it was then possible to perform a wide range of queries using ArcMap’s in-built SQL query builder. The results of these queries illustrate the visual and analytical power of the approach and demonstrate the immediacy of the technique in terms of bridging the gap between data and knowledge. The next section presents these results more fully.

4. Seeing the forest and the trees: geovisualisation of spatial structures

4.1. An overview of migration patterns

The geovisualisation equivalent of the complete MG201 dataset is, owing to its sheer size, incomprehensible; an issue also noted by Holland and Plane (2001) in their analysis of interregional migration in the US. The vast number of flow lines connecting origins and destinations across the UK obfuscates what is potentially a very rich and valuable resource. The analytical skill required
now, therefore, is to turn the resultant spatial data into spatial information that can have some real knowledge-generating power for analysts of population dynamics and residential mobility. It is axiomatic, then, that in order to provide the kind of knowledge which has historically been locked away within these kinds of dataset, an orderly and logical loss of information must follow (Boulding, 1970). Such efforts echo those of Nielsen and Hovgesen (2008) who used ward-level commuting data for England and Wales (albeit not using a flow mapping approach), Yan and Thill (2007, 2008, in press) with their visual data mining of spatial interaction data, and more comprehensive dynamic geovisualisation approaches such as those adopted by Carr et al. (2005) and Guo et al. (2005). Guo et al. (2006). This section of the paper presents the results of the foregoing analysis in a form that both exceeds the capabilities of a non-geovisualisation approach and draws attention to patterns of movement associated with issues such as distance decay, spatial scale and the asymmetry of migration fields.

Before proceeding with more selective queries which can produce very detailed and specialised geovisualisations of migration, it is useful in the first instance to provide an overview of migration in the UK more generally. In Fig. 1, patterns of migration involving inter-ward moves of 50 or more persons are identified; with moves over 50 km displayed separately since this represents a significant migration threshold for many cities in the United Kingdom (see Table 3; Champion & Coombes, 2007; Long, Tucker, & Urton, 1988, p. 638). The longer distance flow lines here are displayed with arrows in order to identify directionality. It should be noted that the FDMT process is also able to handle intra-ward moves but these simply generate individual points which are not visible in visualisations involving inter-ward moves. At this more general level of analysis, three factors are evident. First, the geographic patterns of movement are a close reflection of the existing urban footprint in the UK. Second, the spatial structures are affected by the centroids of wards, particularly where longer distance movements are more obvious (i.e. in more rural areas such as northern Scotland). Finally, and most obviously, the impact of armed forces redeployments and as such do not operate in the same way as typical household mobility patterns. Three such examples are identified in Fig. 1. Nonetheless, this initial geovisualisation provides the kind of information which helps us gain an understanding of patterns of migration between 2000 and 2001 and generates questions which would not be evident from the tabular equivalent (e.g. relating to the nature of the larger volume, long-distance moves).

A useful second stage geovisualisation of the current dataset can be produced by initiating a simple query based on flow magnitude and distance. Besides longer distance armed forces movements, it is also evident that inter-ward moves of 12 or more (a multiple of 3, the lowest reported inter-ward flow owing to small cell adjustment methodology (SCAM); the UK census' data disclosure mechanism) accounted for a majority of the migration occurring in the UK between 2000 and 2001. In total, then, inter-ward moves of 50 km or less accounted for just 10.9% of all point-to-point flow lines but 55.8% of total migration. At an aggregate level, and in order to capture the essence of population mobility, Fig. 2 provides an overview of the geography of migration in the UK. The urban structure is once again evident, but this is also supplemented by a more widespread pattern of migration across the entire UK which essentially reflects population density. Nonetheless, we do begin to understand more about the patterns of migration and the centrality of clearly identifiable locations than would previously have been the case with a non-geographic approach. Also important here is the extent to which patterns of functional polycentricity dominate the geography of migration processes (particularly around London; see Fig. 3).

In addition to the vector-based approach described above, a more experimental approach was taken in representing the entire MG201 dataset in raster format (see also Nielsen & Hovgesen, 2008). Fig. 4 displays the pattern of residential mobility for the UK as a line density raster, with a cell size of 2.5 km. The total number of migrants moving along flow lines through each cell is summed and used to derive a smooth surface so the resultant image is not only indicative of the urban spatial structure of the UK (in terms of ward population densities and geographic sizes) but also of the overall pattern of interactions between places. Therefore, cell densities are not only indicative of ward geography, but also of the ‘flow geography’ of migration in the whole UK. Most obvious here is the pull of London in the UK migration system, and the general concentration of movement among England’s major cities; a pattern which differs from the more familiar ‘urban footprint’ pattern associated with flows in Fig. 2. Once again, such a representation is a compromise but it is illustrative of the geography of

### Table 1

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCODE</td>
<td>The unique identifier from the original centroids layer used to identify the flow origin and used as the basis for subsequent joins</td>
</tr>
<tr>
<td>DCODE</td>
<td>The unique identifier from the original centroids layer used to identify the flow destination and used as the basis for subsequent joins</td>
</tr>
<tr>
<td>MIGRANTS</td>
<td>Total number of migrants moving between the origin and destination from 2000 to 2001</td>
</tr>
<tr>
<td>OWARD_CODE</td>
<td>The standardised, six character ward code for each origin</td>
</tr>
<tr>
<td>OWARD_NAME</td>
<td>The ward name for each origin</td>
</tr>
<tr>
<td>DLA_CODE</td>
<td>The standardised, four character code for each origin's local authority</td>
</tr>
<tr>
<td>DLA_NAME</td>
<td>The name of each origin's local authority</td>
</tr>
<tr>
<td>OCOUNTRY</td>
<td>The name of the UK country each origin is located in</td>
</tr>
<tr>
<td>OPOP2001</td>
<td>The total resident population of each origin in 2001</td>
</tr>
<tr>
<td>OTYPE</td>
<td>The type of local authority each origin is in, such as Scottish Council Area (SCA) or London Borough (LB)</td>
</tr>
<tr>
<td>DREGION</td>
<td>The name of each origin's standard Government Office Region (GOR)</td>
</tr>
<tr>
<td>DWARD_CODE</td>
<td>The standardised, six character ward code for each destination</td>
</tr>
<tr>
<td>DWARD_NAME</td>
<td>The ward name for each destination</td>
</tr>
<tr>
<td>DLA_CODE</td>
<td>The standardised, four character code for each destination's local authority</td>
</tr>
<tr>
<td>DLA_NAME</td>
<td>The name of each destination's local authority</td>
</tr>
<tr>
<td>OCRNTRY</td>
<td>The name of the UK country each destination is located in</td>
</tr>
<tr>
<td>DPOP2001</td>
<td>The total resident population of each destination in 2001</td>
</tr>
<tr>
<td>DTYPE</td>
<td>The type of local authority each destination is in, such as Scottish Council Area (SCA) or London Borough (LB)</td>
</tr>
<tr>
<td>DREGION</td>
<td>The name of each destination's standard Government Office Region (GOR)</td>
</tr>
<tr>
<td>DISTANCE</td>
<td>The distance, measured in kilometres, from the origin's geographic ward centroid to the destination's geographic ward centroid</td>
</tr>
</tbody>
</table>
migration in the UK at an aggregate level, of a kind not possible to display with flow lines.

By being more selective in terms of what to geovisualise, therefore, it has been possible to gain a better understanding of residential mobility processes at a national level. However, in order to understand how migration processes for individual cities may operate, it is necessary to focus on one origin or destination at a time. Given the fact that each flow line has been coded with a range of geographic information relating to the local authority and region of each origin and destination, it is possible to focus on single cities and explore in more detail their spatial interactions across the UK. This process represents the next step in exploiting the richness of the approach.
4.2. Migration flows in detail: Manchester and Glasgow

When working with a vast geographic dataset, the role of the analyst is to put into practice Boulding’s ‘orderly loss of information’ principle (Boulding, 1970, p. 2) so that the main messages are able to be clearly articulated. This necessitates the implementation of some kind of filtered decision making process whereupon background ‘noise’ is removed in order that the important details can be understood with clarity (Holland & Plane, 2001). In order to demonstrate this, two cities in the UK have been selected for further geovisualisation. Given its relative geographic centrality, large population base, and diversity of connections, Manchester provides a useful exemplar. For similar reasons, Glasgow was also selected but it has the additional benefit that its population mobility patterns are less likely to be affected by the in-built data disclosure mechanism (SCAM) since destinations in Scotland are not subject to this process. The remainder of this section will focus on these two cities (based on their local authority boundaries) in order to highlight such issues as asymmetry of migration, origin and destination combinations, spatial scale, flow magnitude, and migration distance.

The geovisualisations in Figs. 5–8 depict a pattern of movement for each city which illustrates a wide migration field covering nearly all areas of the UK. In each map, reciprocal flows (where there were flows in both directions between origins and destination pairs) are displayed separately. Therefore, in each map unique inflows or outflows to and from each city can be viewed and compared to what might be considered ‘core’ reciprocal flows. In the case of Manchester in particular there is a much more obvious pattern of urban to rural migration, particularly involving long-distance moves to the north of Scotland and other more peripheral parts of the UK. Such patterns are of course consistent with the findings of others in relation to population mobility (e.g. Champion & Coombes, 2007; Fielding, 1992) but they are visibly demonstrated here. Also obvious in the Manchester case are reciprocal links with London, inter-urban linkages, and a more voluminous number of origins than destinations. The migration field for Glasgow is similarly diverse with strong connections between the Scottish islands, London and Belfast. The other notable feature of this spatial pattern is the extent to which inflows (Fig. 7) are clearly more voluminous than outflows (Fig. 8). However, given the way the SCAM process works to obfuscate data on origins in England but not destinations in Scotland, this can be viewed as the visual expression of a peculiarity of data disclosure rather than a distinct pattern of geographical mobility.

Although the approach here is predicated upon demonstrating the intrinsic value of geovisualising spatial interaction data, it is also possible to report more precise statistics relating to the flows associated with each city. Table 2 provides details, for Manchester and Glasgow, of the number of persons moving along reciprocal links and unique links. In the Manchester case, we observe an overall net gain of 1380 persons between 2000 and 2001, whilst the number of flow origins is greater by 1049. More interestingly, however, for inflows the number of people moving between reciprocal links represents 69.8% of in-movers but 81.7% of out-movers. This imbalance is reflected by the fact that 30.2% of migrants came from unique origins, whereas only 18.3% moved out to unique destinations. The numeric attributes of movement for Glasgow are considerably more balanced, with a net gain of 45 persons through migration, a reciprocal inflow of 77.3% of the total and a reciprocal outflow of 78.9%. Additionally, the absolute number of reciprocal links for Glasgow is much higher, and the number of unique links much lower, than in the Manchester case. Even from a more cursory analysis, then, it seems clear that Manchester and Glasgow play distinct roles as population redistributors even though the net change in population is rather low (Plane & Mullanigan, 1997; Roseman & McHugh, 1982). The purpose here, however, is not to pontificate over the mechanisms which might cause such differential movement but rather to demonstrate the capabilities of the geovisualisation approach and its associated geodatabase.

A further, final, demonstration of the added value of the approach is provided in Table 3, where the number and percentage of migrants moving to and from each city has been extracted for 10 km intervals up to 100 km. This provides an overview of the distance decay characteristics associated with each city. Both Manchester and Glasgow have the vast majority of inflows and outflows within 10 km of the ward of origin or destination and in both cases there is also a higher percentage moving out within 10 km. For Manchester inflows, the trend of distance decay is reversed at 30–40 km when there is an increase in the number of migrants and again at 40–50 km. For outflows, this pattern occurs at 50–60 km and then declines. The Glasgow case shows more variability with an increase in inflows at 30–40 km and again at 60–70 km, whereas the pattern of distance decay for outflows is more consistent except at 60–70 km where there is a marked increase. Additionally, both cities attract a greater percentage of migrants from 100 km or more than they send out in return. Once again, the point here has not been to provide an in-depth analysis of inter-urban migration patterns for these cities but to demonstrate that the geovisualisation approach is laid upon a solid foundation of query-capable data. Nonetheless, the results from Table 3 are consistent with spatial patterns we might expect from Manchester and Glasgow (see Champion & Coombes, 2007). The former is a part of a major polycentric urban region which includes proximate cities such Liverpool and Leeds, between 40 km and 60 km distant. The latter is also part of a polycentric urban region (Bailey & Turok, 2001), but one dominated by itself and Edinburgh, approximately 70 km to the east.

The spatial patterns displayed so far are particularly useful in presenting an overview of population mobility patterns for individual cities and for providing an overview of patterns across the UK more broadly. However, in order to more fully develop the approach it is necessary to conduct further experiments into the representation of spatial interaction data. The next section therefore presents some additional methods of displaying UK migration data in order that knowledge might be more easily acquired.

### Table 2

Migration magnitudes and origin-destination linkages.

<table>
<thead>
<tr>
<th></th>
<th>Manchester</th>
<th>Glasgow</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Migrants</td>
<td>O-D links</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>%</td>
</tr>
<tr>
<td>Total Inflow</td>
<td>64,336</td>
<td>100.0</td>
</tr>
<tr>
<td>Total Outflow</td>
<td>62,936</td>
<td>100.0</td>
</tr>
<tr>
<td>Reciprocal Inflow</td>
<td>44,894</td>
<td>69.8</td>
</tr>
<tr>
<td>Reciprocal Outflow</td>
<td>51,462</td>
<td>81.7</td>
</tr>
<tr>
<td>Unique Inflows</td>
<td>19,442</td>
<td>30.2</td>
</tr>
<tr>
<td>Unique Outflows</td>
<td>11,494</td>
<td>18.3</td>
</tr>
</tbody>
</table>

The numeric attributes of movement for Glasgow are considerably more balanced, with a net gain of 45 persons through migration, a reciprocal inflow of 77.3% of the total and a reciprocal outflow of 78.9%. Additionally, the absolute number of reciprocal links for Glasgow is much higher, and the number of unique links much lower, than in the Manchester case. Even from a more cursory analysis, then, it seems clear that Manchester and Glasgow play distinct roles as population redistributors even though the net change in population is rather low (Plane & Mullanigan, 1997; Roseman & McHugh, 1982).
4.3. Further experiments in flow mapping

The tradition of representation in flow mapping remains very much a static, linear one (for an exception, see ONS, 2008). In part, this is a reflection of the realities of representing a three dimensional world in two dimensions, but it is also a comment on the lack of technical advancement in a field characterised by such progress. Therefore, some more experimental methods are presented here. Using a combination of vector and raster representation, the intention has been to maximise the communicative power of each geovisualisation by striking a balance between loss of data and clear depiction. The purpose here, once again, is not to provide a comprehensive account of migration patterns but rather to demonstrate the information-generating capabilities to be derived from spatial interaction data. The approaches specified below are illustrative of a more refined, orderly loss of information which

Fig. 2. Flows of 50 km or less and 12 or more persons.
produces the kind of visual knowledge that raw data alone cannot generate. They remain imperfect, but offer different solutions to the representation of spatial interaction data which hitherto remains something of an 'orphan step-child' within the world of GIS.

In Fig. 9 flow lines for inflows to Manchester have been converted to points (with the destination points removed for clarity) and presented in graduated symbol format. Since individual origin wards may actually link to a number of different wards in Manchester, these have been summarised by ward. The resulting display therefore gives a total figure for origin wards of all people moving from there to the district of Manchester between 2000 and 2001. The focus is on the main urban area of northern England and from this we can gain a better understanding of the migration field associated with the city. It is clear that the largest contributions to Manchester’s inflows are those wards in close proximity, but there is also evidence of stronger inter-urban links with Liverpool, Leeds and Sheffield in particular. At one glance, then, we can observe a higher degree of connection within a wider urban system and a more localised process of population mobility from areas just outside the district of Manchester itself. Such a representational device allows us to simultaneously understand which individual wards are most closely connected with the district of Manchester. This combination of spatial scales is therefore able to communicate a large volume of data in a simple format.

The next step in developing the approach is illustrated in Fig. 10, where migration patterns for the wider urban system in northern England are shown. Intra-ward movements here are illustrated by means of graduated point symbols, inter-ward moves of 20 or more are indicated by connecting lines of standard width, and the general migration pattern for all movements has been summarised through the creation of a line density raster surface (0.5 km resolution) for all moves in northern England. The raster surface quite accurately identifies the major functional urban areas (but without the associated visual ‘clutter’ of the full dataset), whereas the connecting flow lines are more indicative of sub-regional functional entities within the major urban areas, such as Merseyside or Greater Manchester. By adding in the intra-ward point data, we are also able to gain an understanding of which areas have particularly high internal population churn, with West Yorkshire most prominent here.

Moving down in spatial scale to the city level, Fig. 11 takes a similar approach for the city of Glasgow in order to illustrate intra-urban, rather than inter-urban connections. In this case, the representation of internal ward migration by graduated point symbols and of general migration trends by a line density raster (0.25 km resolution) is able to provide a greater degree of differentiation between locations, with several of Glasgow’s more peripheral wards experiencing high levels of intra-ward migration. The migration density surface is also able to identify which parts of the city have the greatest degree of population mobility. In addition, the connecting inter-ward lines have this time been aggregated to show gross flows (two-way flows summed), illustrating the absolute level of connectivity between wards. Of particular note here is the high degree of connectivity between a grouping
of wards just to the west of the city centre, in an area highly pop-
ulated by students and therefore more closely associated with res-
idential mobility (Cadwallader, 1992). In both Fig. 10 and Fig. 11,
then, it is possible to understand more about general patterns of
movement for areas at different spatial scales, to illustrate specific
linkages between places, and to identify where the highest levels of
intra-ward mobility exist.

In Fig. 12, a combination of raster and vector representation
has been used in conjunction with district level data, for compar-
ison with the ward-level examples above. The focus here is once
again on an urban area in northern England, but this time net
migration flows by district have been used to create a smooth ras-
ter migration surface at a resolution of 1 km. Areas of high net in-
migration and areas of net out-migration are displayed in con-

Fig. 4. UK migration 2000–2001 as a line density raster.
Major flows of people (200 or more) are displayed as flow lines here for the same area. One of the issues which this geovisualisation attempts to address is the representation of flow directionality. Even with the relatively limited number of flow lines on the map, the use of directional arrows would complicate the display. Therefore, the underlying raster image is informative as to the spatial relationships between places. For example, the contrasting example of Bradford (net out-migration) and Leeds (net in-migration) allows us to understand more about the likely spatial interaction between the two districts in terms of migration. In this instance, there is a net flow from Bradford to Leeds of 431, with 2016 moving from Bradford to Manchester inflows, 2000–2001.
to Leeds and 2447 moving from Leeds to Bradford. The reason for displaying absolute flows in this case is to emphasise the level of spatial interaction between places, rather than identifying the direction of the relationship.

These more experimental geovisualisations are an attempt at reconciling clarity and complexity in order to provide information about spatial interaction in the UK migration system between 2000 and 2001. Although imperfect, the loss of information has been carefully managed and what they do show is more important than what is left out. This is the key to what has been attempted here but further development in this area is still needed from the GIS user community at large. It remains to be seen if we can rise to meet this challenge in a data rich environment such as ours.
5. The value added of a geovisualisation approach to spatial interaction data

5.1. Towards ‘spatial interaction information’?

The foregoing analysis and discussion is firmly rooted in a geographical tradition of geovisualisation which has contributed much to public understandings about the ways in which space is organised and operates. However, this tradition has been less well-placed when it comes to the representation of geographical movement, despite its centrality to the study of geography (e.g. Alvarez & Mossay, 2006; Bontje & Latten, 2005; Edmonston, 2006; Jianfa, 1996; Renkow & Hoover, 2000). In part, this is a remnant of a more Cartesian era but in general it is illustrative of the
fact that the subject matters of geographical movement and connection, on the one hand, and geovisualisation, on the other, have not yet been suitably reconciled. The approach outlined here represents an attempt at reconciliation and a move towards more innovative attempts at generating spatial interaction information. The reasons for this are not purely academic; they are founded upon the belief that the understanding of geographic space is an important aspect of decision making, and that geovisualisation is the most appropriate vehicle through which this can be achieved (Davis, 2001). Attempting to unlock the potential of spatial interaction data in this piece of research can therefore be seen to represent not only a dissemination exercise but, hopefully, one of education and, perhaps, spatial enlightenment.

Fig. 8. Glasgow outflows, 2000–2001.
5.2. Increasing the utility of spatial interaction geovisualisations?

When one considers how to increase the number of potential end users of geographic data, the obvious outlet is some form of online interface. This also has the potential for allowing the static approaches described above to move towards dynamic representations. With the rapid advance of web-based geospatial applications, including Google Maps, Google Earth, Microsoft Virtual Earth, and a whole range of more specific task-oriented tools, the Internet has become a frontier in the race to virtually colonise the globe. It has also brought to the fore the implications of attempting to communicate to a very diverse group of users (Edsall, 2007). These issues, and others relating to the rapid diffusion of spatial data, have recently been discussed in the context of Web 2.0 (CASA/NCeSS, 2008) and the full ramifications are unlikely to be known for several years, but we are at least at the stage where large spatial datasets and interaction matrices can be visualised and disseminated online as a means to increasing spatial awareness. In order to illustrate a simple yet effective method for viewing and publishing the geovisualisations produced here, a combination of ArcGIS, a freely available extension, and Google's geospatial technology have been utilised. The results of this work can be seen on the dedicated project website (Rae, 2009); a summary of method and content is provided here for information.

### Table 3

Migration magnitudes and distance bands.

<table>
<thead>
<tr>
<th>Distance Band (km)</th>
<th>Manchester Inflow</th>
<th>Manchester Outflow</th>
<th>Glasgow Inflow</th>
<th>Glasgow Outflow</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>%</td>
<td>Total</td>
<td>%</td>
</tr>
<tr>
<td>0–10</td>
<td>43,544</td>
<td>67.7</td>
<td>46,087</td>
<td>73.2</td>
</tr>
<tr>
<td>10–20</td>
<td>3056</td>
<td>4.8</td>
<td>3641</td>
<td>5.8</td>
</tr>
<tr>
<td>20–30</td>
<td>1084</td>
<td>1.7</td>
<td>1107</td>
<td>1.8</td>
</tr>
<tr>
<td>30–40</td>
<td>1202</td>
<td>1.9</td>
<td>637</td>
<td>1.0</td>
</tr>
<tr>
<td>40–50</td>
<td>1629</td>
<td>2.5</td>
<td>943</td>
<td>1.5</td>
</tr>
<tr>
<td>50–60</td>
<td>1439</td>
<td>2.2</td>
<td>816</td>
<td>1.3</td>
</tr>
<tr>
<td>60–70</td>
<td>871</td>
<td>1.4</td>
<td>553</td>
<td>0.9</td>
</tr>
<tr>
<td>70–80</td>
<td>572</td>
<td>0.9</td>
<td>384</td>
<td>0.6</td>
</tr>
<tr>
<td>80–90</td>
<td>391</td>
<td>0.6</td>
<td>396</td>
<td>0.6</td>
</tr>
<tr>
<td>90–100</td>
<td>645</td>
<td>1.0</td>
<td>457</td>
<td>0.7</td>
</tr>
<tr>
<td>&gt;100</td>
<td>9903</td>
<td>15.4</td>
<td>7935</td>
<td>12.6</td>
</tr>
</tbody>
</table>

**Fig. 9.** Graduated symbols: Manchester inflows.
Fig. 10. Urban network migration: North of England.

Fig. 11. Intra-city migration: Glasgow.
5.3. Principles for the orderly loss of information

Before considering some future challenges and opportunities, it is useful here to consider that the value added of a geovisualisation approach is in fact closely related to the application of principles for the ‘orderly loss of information’ (Boulding, 1970). Rather than define a prescriptive list, however, the contention is that the skill of the analyst must be to apply general principles in a logical and systematic way; appropriate to the data being presented. First and foremost here is expansive inclusion; that is, the inclusion of every possible data element at the outset, in addition to any related data that may add value (e.g. see Table 1). This will allow the analyst a greater degree of selectivity when decisions are being made about what to display and also facilitate improved query capability; though it will of course make geocomputation more demanding. In the case of mapping a large migration matrix, this is certainly the case. Any loss of information in geovisualisation should be as a result of analytical decision making and not by omission at the outset. The second principle that we can apply is one of iterative loss, so that the key trends and spatial structures can be identified via a series of repetitive experiments whereby the data displayed is iteratively reduced. In the case of flow mapping, this may involve opting not to display small flows in order to draw attention to the more important high volume flows or exploring the impact of using different migration thresholds (e.g. Fig. 1 and 2). In such a process, then, there is an inevitable balance to be struck between information and omission. A third general principle we can apply is related to the skill of the analyst and can be defined as the need to derive simplicity from complexity. The level of difficulty here is very closely related to the overall size and complexity of the dataset, but in an era of data-rich GIS this is a significant challenge and one that must be taken seriously. In this paper, Fig. 4 represents the most obvious implementation of this principle. Finally, we may think of all the above under the rubric...
of optimal compromise, since we are simultaneously trying to optimise the effectiveness of display whilst making compromises in relation to what is either included or omitted (e.g. Fig. 10 and 11). In each of the geovisualisations presented here, and those on the related web site, these principles have been applied. More work remains to be done in this area, of course, but through the implementation of such principles it is possible to add value to data that typically remains unseen and under-used.

6. Conclusions

6.1. Enhancing decision support through geovisualisation

The availability and use of the tools described above can be seen as the culmination of what Tobler has advocated for many years; that is, the orderly and intelligible representation of continuous flow patterns that would otherwise remain unseen. This echoes principles laid down by Longley et al. (2005, p. 290) that ‘a major goal of GIS usage is often to make sense of the huge archive resources that are available, without creating information overload’ It is hoped, ultimately, that the approach adopted here can fulfils the four principal purposes of geovisualisation, identified as exploration, synthesis, presentation, and analysis (Longley et al., 2005, p. 295). Using the inter-ward migration table from the 2001 UK census as an example, it has been possible to demonstrate how a spatial interaction data geovisualisation approach can answer questions relating to the connectedness of places, the disconnectness of others and, at a larger scale, identify important urban spatial structures that depict geographical movement associated with household mobility. The contention here is that the added value to decision making for this kind of data does not in any way constitute a new model of representation; rather it conforms to the existing method but with a form of data hitherto under-represented and little understood. Given its centrality to our understanding of place, and the fundamental importance of spatial interaction data to the study of human activity, it would seem that geovisualisation in this field has been something of a slow starter. If we are to properly understand such key themes as labour market dynamics, urban–rural interaction, household mobility flows, business relocation patterns and the like, it is necessary to be able to properly visualise them.

6.2. Four areas now requiring attention

The future for representations of spatial interaction data ought to follow a similar path to that being forged for other types of data (e.g. CASA, 2008) but for spatial interaction data in particular there are four main areas that require attention. First, now that the means to represent dynamic spatial processes are available to the wider GIS user community, they ought to be more formally embraced by software developers and vendors to the extent that spatial interaction data geovisualisation approach can answer questions relating to the connectedness of places, the disconnectness of others and, at a larger scale, identify important urban spatial structures that depict geographical movement associated with household mobility. The contention here is that the added value to decision making for this kind of data does not in any way constitute a new model of representation; rather it conforms to the existing method but with a form of data hitherto under-represented and little understood. Given its centrality to our understanding of place, and the fundamental importance of spatial interaction data to the study of human activity, it would seem that geovisualisation in this field has been something of a slow starter. If we are to properly understand such key themes as labour market dynamics, urban–rural interaction, household mobility flows, business relocation patterns and the like, it is necessary to be able to properly visualise them.

6.3. Challenges and opportunities: some final thoughts

Developing a method of data display which can produce new information, knowledge and wisdom is the centre-piece of what has been attempted here. In the present era of data-rich and computationally-rich GIS, revisiting and building upon the topic of ‘movement mapping’ (Tobler, 2003) has presented us with challenges and opportunities. The challenges are to continue to develop new ways of representing spatially dynamic processes which are fundamental to the workings of society and our experiences of space, to develop user-friendly dissemination platforms which allow the greatest number of users to assimilate large volumes of information currently locked within large interaction datasets, and to make sure the results are fed back into the formulation, development and evaluation of policy at all spatial and governmental levels. This final challenge is also the biggest opportunity, since the huge variety of datasets available extends far beyond national census outputs and the technical capacity to adopt such approaches already exists in most cases. Other significant opportunities relate to the potential for engendering a new level of spatial understanding among a wider community of non-specialists, capturing the dynamism of spatial interaction through time, and stimulating the creation of new ‘mental maps’ of the spatial structure of urban areas which permeate the public consciousness in the same way that many other geographic representations have (e.g. Dorling, 2007). Not only do we need to bridge the gap between data and knowledge, then, we also need to bridge the gap between research and practice. Geovisualisation can contribute much to the former, whilst the latter will involve a more fundamental geospatial culture change, the kind of which is presently being witnessed in the burgeoning user community of ‘GIS 2.0’ technologies such as Google Maps and KML. The end result of this piece of research therefore also represents a starting point on the road to understanding; one which seems to be long overdue and one for which the foundations were laid long ago.

References
