Parallel computation for accessibility based planning support

Jianquan Cheng[*1], Jianguan Tu[†2] and Liangxiu Han[‡3]

[1]School of Science and the Environment, Manchester Metropolitan University (MMU)
[2] International School of Software, Wuhan University, China
[3] School of Computing, Mathematics and Digital Technology, MMU

Summary

Involving stakeholders in decision-making at workshops requires a computer system to respond quickly to various scenarios proposed by participants. Due to the increasing complexity of urban systems, such planning support often faces the challenges of "big data" and poor computational performance. This paper proposes a new approach using parallel processing techniques (i.e. MPI Message Passing Interface) to support workshop participants in interactively building planning scenarios and visualising outputs of job accessibility across Greater Manchester. MPI-based parallel algorithms have been run on a cluster of computers for reducing computational time cost. The results and performances are critically evaluated and recommendations for future work provided.

KEYWORDS: Parallel computation, MPI, workshop based public participation, job accessibility, Greater Manchester.

1. Introduction

Accessibility can be defined as the potential of opportunities for interaction or the ease of reaching preferred places. Accessibility planning, focusing on promoting social inclusion of disadvantaged groups and improving their access to opportunities including employment, education and health care (in the UK) or on integrated transport and land use planning (in the Netherlands), involves the tasks or stages of identifying issues, visioning (strategic design), generating alternative solutions and evaluating the alternatives. Engaging or involving stakeholders in these stages through online public participation or at workshops, requires a computer system to respond quickly to a wide range of scenarios proposed by diverse participants.

Due to the increasing complexity of urban systems, such planning support often faces the challenges of "big data" (high volume, high velocity and highly variable) (Batty, 2013) and poor computational performance dominating mainstream GIS software. Accordingly, there is an increasing demand for methodological solutions to enable efficient visualisation of large data sets and fast computation of complicated spatial problems.

Parallel computation (e.g. Mapreduce, MPI) has been extensively applied for solving complex spatial problems (Yin et al, 2012), but not for accessibility based planning support particularly when facing the challenge of big data.

The next section introduces the study area, data sources, methods of measuring job accessibility, and

[*] J.cheng@mmu.ac.uk
[†] Tujg1973@gmail.com
[‡] L.han@mmu.ac.uk

MPI approach for parallelizing analysis algorithms (e.g. network analysis and accessibility measurement) over a cluster of computers. Section 3 presents results from a test of workshop-oriented scenario building using a case study of Greater Manchester. The paper finishes with general conclusions, evaluation of performance and recommendation for future work.

2. Data and methods

2.1 Data sources

The study area is Greater Manchester (GM). Considering edge effect of spatial analysis and requirement of job accessibility measurements (e.g. competitions), the spatial extent of this application has been expanded to cover a large area of England (Figure 1) as people living in other counties may commute to the GM region. The data sets include OS-provided ITN data, commuting data from 2011 census via Cider and demographic data from 2011 census via ONS.

A car network data set is built from the road network shape file converted from the ITN data. The total number of jobs at output area level are aggregated from the commuting data with destination sites in the study area. The total working population aged 17-65 at output area level is joined with the job data into a point layer. The total number of nodes on the network is 1.5 million and the total number of points is 90,000 for Origin (residence) and Destination (work) each (Figure 1). The matrix of travelling time between pairs of 90,000 points is 30G when represented as an integer data type. Obviously, this is a typical example of 'Big Data' application in terms of high data volume.

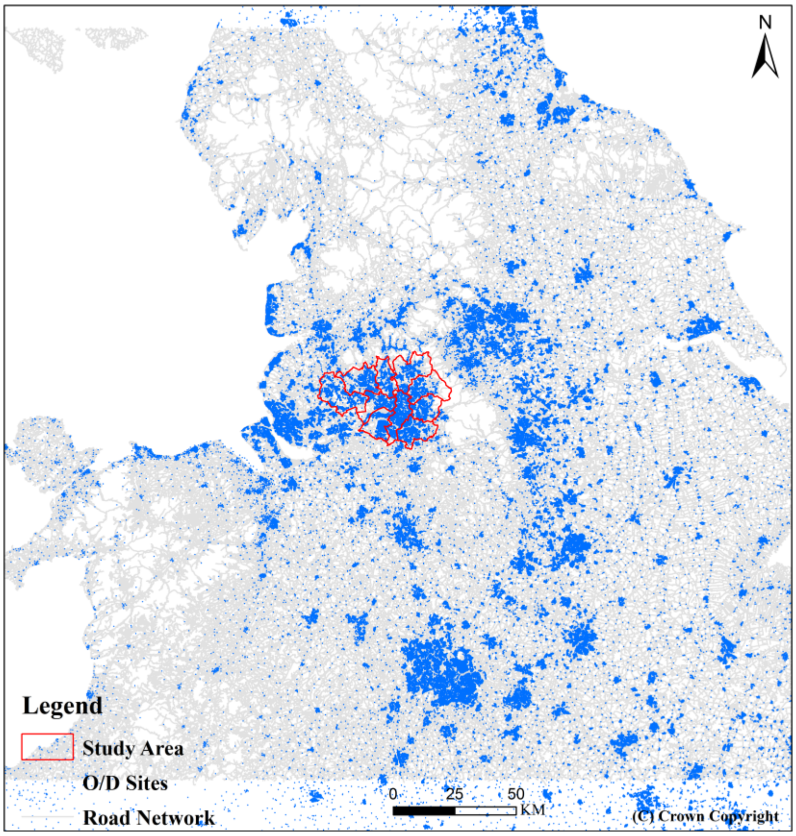

Figure 1 Study area and spatial data coverage

2.2 Measurements of job accessibility

Job accessibility, as the interface between transport, workers and jobs systems, is thus very much dependent on the degree of their interactions, including both spatial and non-spatial interactions (Cheng and Bertolini, 2013). The spatial interaction between workers and jobs systems results in spatial dimensions of accessibility such as competition between workers or between employers. However, non-spatial interactions, which are often neglected in the literature, result from the varied degree of match or imbalance between the demand and supply sides. To accurately measure job accessibility, distance decay, competitions on the demand and supply sides, and diversity need to be incorporated into measurement of job accessibility (Cheng et al. 2013; Cheng et al. 2007). For planning support, place or location based measurement is adopted as spatial separation is the main concern. To overcome the hard-to-interpret issue of gravity-based method, which is a relative measurement, an absolute measurement – job opportunity (potential number of jobs) is proposed by considering distance decay and two-way competitions into Equations 1-6. It is assumed that the total number of residence sites, job sites and job types (similarly worker types) are *m*, *n* and *s* respectively.

$$W_i = \sum_{k=1}^{s} W_{ik} \tag{1}$$

$$E_j = \sum_{k=1}^{s} E_{jk} \tag{2}$$

$$f(t_{ab}) = e^{-\beta(k)\times t_{ab}} \tag{3}$$

$$P^k{}_{jz} = \frac{E_{jk} \times f(t_{zj})}{\sum_{u=1}^{n} E_{mk} \times f(t_{zm})} \tag{4}$$

$$O_{ik} = \sum_{j=1}^{n}\sum_{k=1}^{s} E_{jk}(i) = \sum_{j=1}^{n}\sum_{k=1}^{s} \frac{E_{jk} \times P^k{}_{ji} \times W_{ik} \times f(t_{ij})}{\sum_{z=1}^{m} P^k{}_{jz} \times W_{zk} \times f(t_{zj})} \tag{5}$$

$$O_i = \sum_{k=1}^{s} O_{ik} \tag{6}$$

Where W_i denotes the total working-age population at site *i*, among which W_{ik} being the population for worker type *k*. E_j denotes the total number of jobs at site *j*, among which E_{jk} being the population for job type *k*. A negative exponential function in Equation 3 is used to quantify travel time friction at urban level and t_{ab} is the traveling time by car from residence site *a* to work place *b*. Equation 4 indicates competition for workers between employers and Equation 5 integrates the two competitions into the measurement. In Equation 6, O_{ik} is the job opportunity allocated from job type *k* to residence site *i* so O_i is the aggregated overall job opportunity for the site *i*. The algorithm complexity in Equations 1-6 is O ($m * n^2$), which is much higher than other measurements.

2.3 Parallel computation

The big-data challenges in this study include quick visualisation of car network data set and fast computations of travelling time between pairs of origin and destination and job accessibility (Equations 1-6). The latter is the focus of this paper as there have been extensive studies on the former including GPU techniques. The cluster of computers used is composed of 192 processors – 12 nodes and 16 processors at each node. Each processor (Intel ® Xeon® CPU E5-2620 v2 @ 2.10GHz) has up to 64G RAM.

The calculation of quickest route t_{ab} in Equation 3 is implemented by the traditional Dijkstra algorithm that was compared with Moore algorithm. A circle with a radius of 60 km is used to limit the search space, which is also applied to the calculation of job accessibility based on the assumption that workers may travel by train instead of by car if the driving distance is longer than 60 km. Two parallel computation strategies (Figure 2), based on a parallel approach (Han et al, 2011), are designed to calculate t_{ab} and their selection is based on the data volumes of network and O/D points. The calculated matrix of t_{ab}, totalling about 30G in storage space, is distributed to each processor with an equal number of OD points.

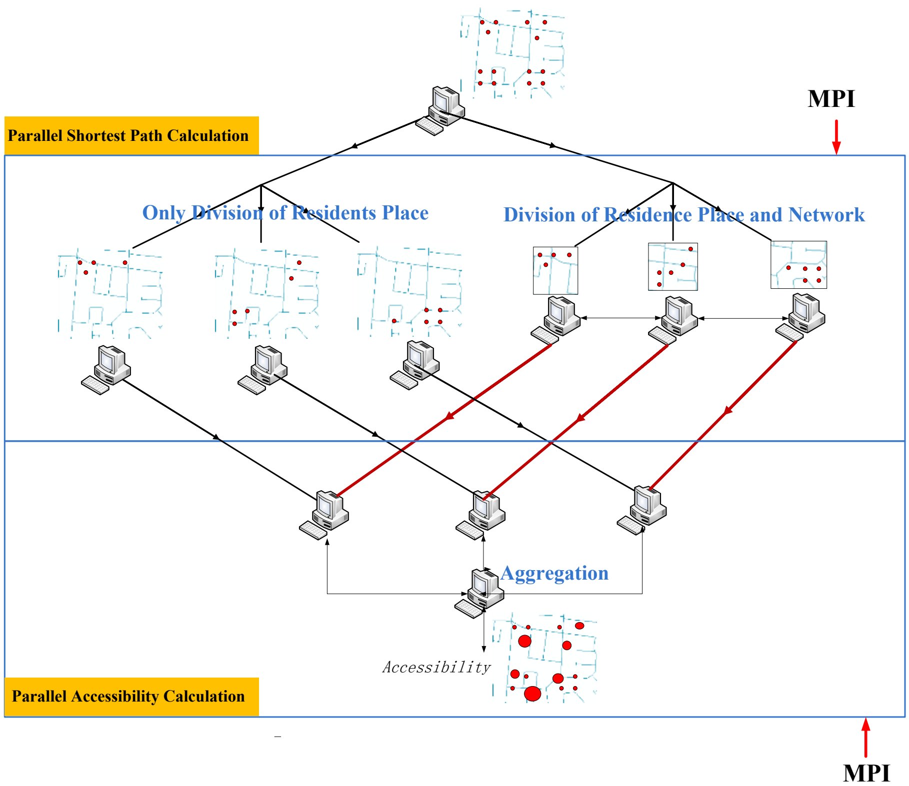

Figure 2 Two strategies of parallelization

The parallel computation of accessibility measurement involves four tasks of allocating the decomposed calculations in Equations 4-5 to each processor and then aggregating the results into a master processor using the MPI technique. For example, Task 1 is focused on competition for workers between employers in Equation 4, which was calculated simultaneously by all processors. Task 2 tends to summarize the calculated results from each processor and scatter it to each processor again using the MPI function. Task 3 is focused on integration of two competitions as shown in Equation 5, calculated simultaneously by all processors. Task 4 is to gather all the results using the MPI function.

3. Results

Due to unavailability of job type data, only distance decay and competitions are considered in the measurement. Several days are needed to work out the job accessibility in the study area on a single PC (Inter (R) Core (TM) i5 CPU 4 Cores) but the work has been massively reduced to 11 minutes over the cluster (see Figure 3). The kernel density shown in Figure 3 is easy for communication with stakeholders as the calculated job opportunity is an absolute value. Next, Manchester Airport is taken as an example of workshop based planning support, in which participants attempt to build a scenario: a new spatial pattern of job opportunity after 10,000 jobs are added to the airport, stimulated by the airport expansion plan. Figure 4 (right) shows the increased job opportunity around the zoomed airport area. The scenario building over the cluster only takes 15 minutes, which is within the acceptable range of workshop practice. Other scenarios such as changes of road network and traffic) will be built and visualised in the same way and with same effects.

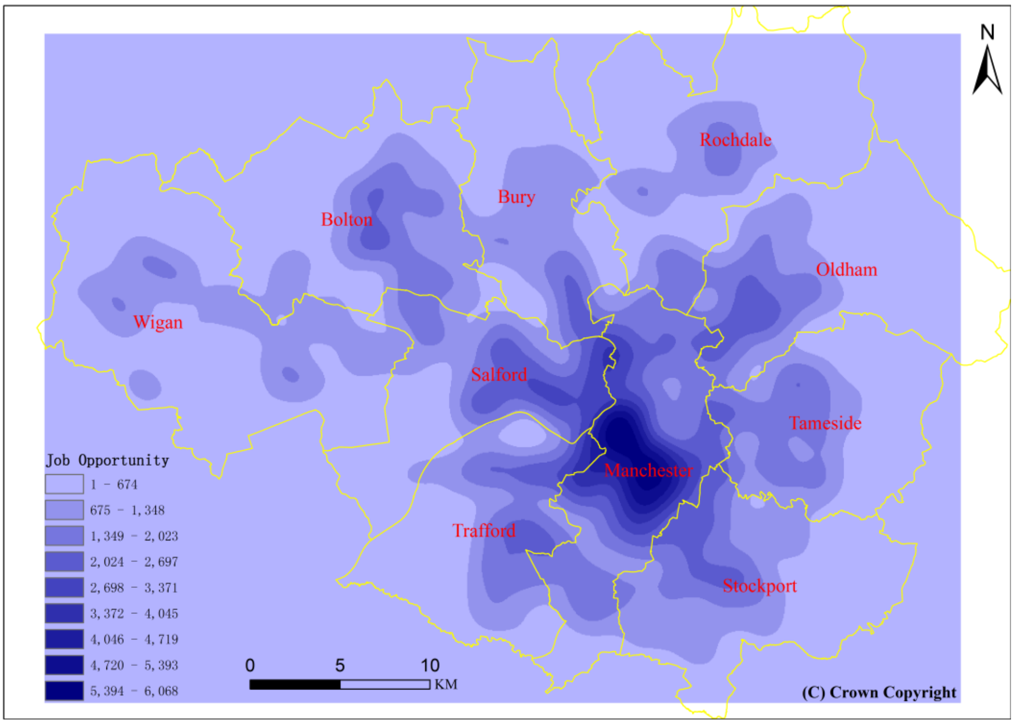

Figure 3 Spatial pattern of calculated job opportunity across the Greater Manchester region

Figure 4 Scenario one: addition of 10,000 jobs at the Manchester Airport and the resulting change

4. Discussion and conclusion

This paper has demonstrated that parallel computation can provide methodological solutions to “big-data” challenges faced by the accessibility based planning support. The methods can be also applied to other complex spatial problem solving. In the future, a multiple-mode transport system, temporal elements and diversity should be further considered.

Acknowledgements

Funded by the School of Science and the Environment, MMU (staff research incentivisation award). Thanks to Dr Nick Gould for commenting on the draft.

Biography

J Cheng is a senior lecturer in GIS and urban planning at MMU, whose research interests include GIS for urban applications and the urbanisation of China. J Tu is an associate professor in GIS at Wuhan University, China, whose interests are focused on GIS application system development. L Han is a reader in computational intelligence at MMU, with interests in parallel computation and big data analytics for a variety of applications.

References

Batty M (2013). Big data, smart cities and city planning, *Dialogues in Human Geography* 3(3), 274–279.

Yin L, Shaw S, Wang D, Carr E, Berry M, Gross L, Comiskey, E J (2012). A framework of integrating GIS and parallel computing for spatial control problems – a case study of wildfire control, *International Journal of Geographical Information Science*, 26(4), 621-641.

Cheng J and Berterloni L (2013). Measuring urban job accessibility with competitions, distance decay and diversity, *Journal of Transport Geography*, 30,100-109.

Cheng J, Berterloni L, le Clercq F, Kapoen L (2013). Understanding urban networks: comparing a node-, a density- and an accessibility-based view, *Cities: International Journal of Urban Policy and Planning*, 31,165-176.

Cheng J, Berterloni L, le Clercq F. (2007). Measuring sustainable accessibility. *Transportation Research Record: Journal of the Transportation Research Board*, 2017, 16-25.

Han L, Liew C S, van Hemert J, Atkinson M P. (2011) A generic parallel processing model for facilitating data mining and data integration. *Journal of Parallel Computing*, 37(1), 157-171.