

Big Data computation for workshop-based planning support

Jianguang Tu

International School of Software
Wuhan University
Wuhan, P.R.China
Tujg1973@gmail.com

Jianquan Cheng *

School of Science and the Environment
Liangxiu Han
School of Computing, Mathematics and Digital Technology
Manchester Metropolitan University (MMU)
Manchester, UK
J.cheng@mmu.ac.uk and L.Han@mmu.ac.uk

Abstract—Involving stakeholders in decision-making at planning 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 visualizing outputs of job accessibility across the Greater Manchester. MPI-based parallel algorithms have been run on a cluster of computers for reducing computational time cost. The tested results and computational performances are critically evaluated and recommendations for future work provided. Particularly this paper also addresses several key research questions related to a theoretical framework of applying big data analytics for urban planning support.

Keywords- Big Data, GIS, MPI, Parallel computation, Planning support system.

I. INTRODUCTION

Planning is a future-oriented activity, strongly conditioned by the past and present. Spatial planning means the methods influencing the future distribution of people and activities at various spatial scales, ranging from national, regional, urban to local or neighborhood levels. Reasonable and scientific spatial planning is a pre-requisite for achieving sustainable city and community. Planners have always sought tools to enhance their analytical, problem-solving and decision-making capabilities. Planning support can be any geospatial tools or ICT environment to support whole or partial stages of any professional spatial planning tasks [1]. The popular planning support systems (PSS) include *What-if* [2], *UrbanSim* [3] and *CommunityViz*, each of which with varied strengths and weaknesses in terms of planning stage and target supported, spatial scale, data requirements, model structure, system function and software development model.

Due to the increasing complexity of urban systems, such planning support systems often face the challenges of “big data” (high volume, high velocity and highly variable) [4] and poor computational performance. Accordingly, there is an increasing demand for methodological solutions to enable efficient visualization of large data sets and fast computation of complicated spatial problems. Parallel computation techniques (e.g. MapReduce and Message Passing Interface - MPI) have been extensively applied for solving complex spatial problems [5], but not for planning support particularly when facing the challenge of big data. The next section explains the theoretical frameworks of planning support system and job accessibility. Section three introduces the study area, data sources, methods of measuring job accessibility, and MPI approach for parallelizing analysis algorithms (e.g. network analysis and accessibility measurement) over a cluster of computers. Section four 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, recommendation for future work and questions for applying big data for planning support.

II. PLANNING SUPPORT SYSTEM

In geospatial terms, any S (Spatial) PSS system can be generalized as having five interrelated components: spatial planning process, data, model, system, and user (Fig.1). Firstly, the planning ‘process’ describes which stage, which scale, which type of planning the PSS should support. Planning processes vary from country to country, from period to period, from neighborhood to urban region, and from land use planning to transport planning. Secondly, ‘data’ covers which data sets should be collected and integrated, in order to support a specific planning process. Data in particular vary with scale (spatial, temporal and decision-making) and level (aggregate, disaggregate and individual). Thirdly, as the core element of a PSS, a ‘model’ refers to which analytical functions the PSS should be

equipped with. The model varies with the objective of planning support (e.g. assessment, prediction and simulation) and is either data-driven, knowledge-driven or method-driven. Fourthly, a ‘system’ describes which environment, platform or interface the PSS should promise to offer for the communication between the PSS and users. The system needs to integrate the modules of both data and model and serves the planning process. Finally, a ‘user’ is composed of diverse actors – stakeholders, planners, decision-makers, residents, developers and managers, who contribute their inputs (e.g. knowledge, opinions and comments) to different stages and tasks of planning.

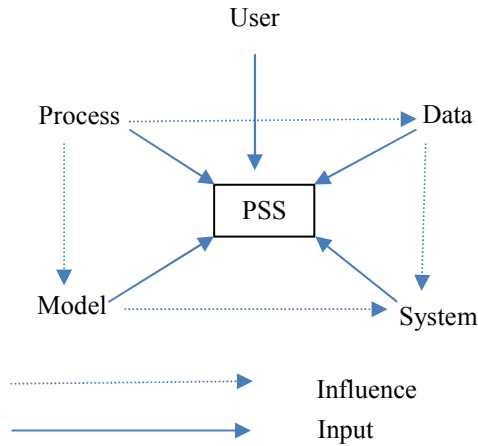


Fig. 1 Five components of a PSS

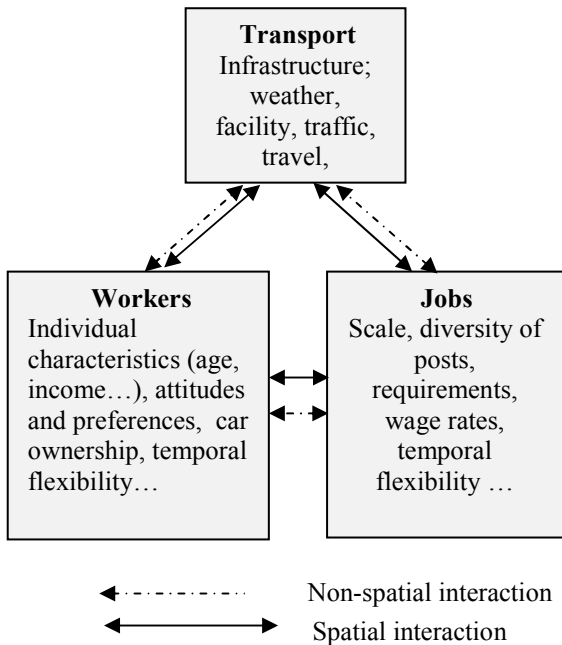


Fig 2. A conceptual framework of job accessibility

In this paper, accessibility-based planning support is taken as an example. Accessibility can be defined as the potential of opportunities for interaction or the ease of reaching preferred places. Job accessibility involving the spatial and no-spatial interactions between transport, workers and jobs systems (Fig.2) [6] not only relates to commuting but also has impacts on spatial inequity. 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. According to the theoretical framework of planning support (Fig.1), a successful planning support system needs to understand the requirements from the five components and solve the challenges from each of them. As a preliminary study, this paper is only focused on the system and model components.

III. DATA AND METHODS

A. Data Sources and Challenges

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 (Fig.3) as people living in other counties may commute to the GM region.

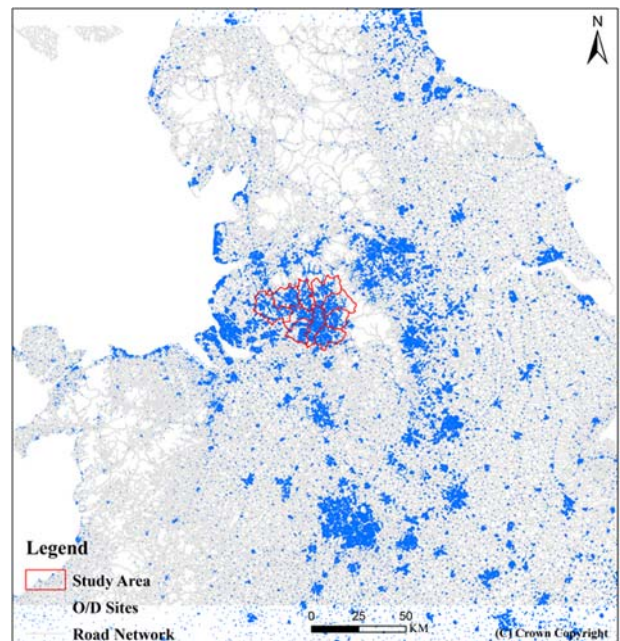


Fig.3. Study area and spatial data coverage

The data sets include the ordnance survey (OS) provided ITN (Integrated Transport Network) data, commuting data from 2011 census via Cider, and demographic data from 2011 census via Office for National Statistics (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 3). The matrix of travel 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.

B. 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 [6]. The spatial interactions between workers and jobs systems result in spatial dimensions of accessibility such as competitions 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 [7, 8]. 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.

$$W_i = \sum_{k=1}^s W_{ik} \quad (1)$$

$$E_j = \sum_{k=1}^s E_{jk} \quad (2)$$

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

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

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

$$O_i = \sum_{k=1}^s O_{ik} \quad (6)$$

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 travel 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 $O(m \times n^2)$, which is much higher than other measurements.

C. Parallel Computation

The big-data challenges in this study include quick visualization of car network data set and fast computations of travel 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 (Fig.4), based on a parallel approach [9], are designed to calculate t_{ab} and their selection is based on the data volumes of network and origin and destination points. The calculated matrix of t_{ab} , totally about 30G in storage space, is distributed to each processor with an equal number of origin and destination points. 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.

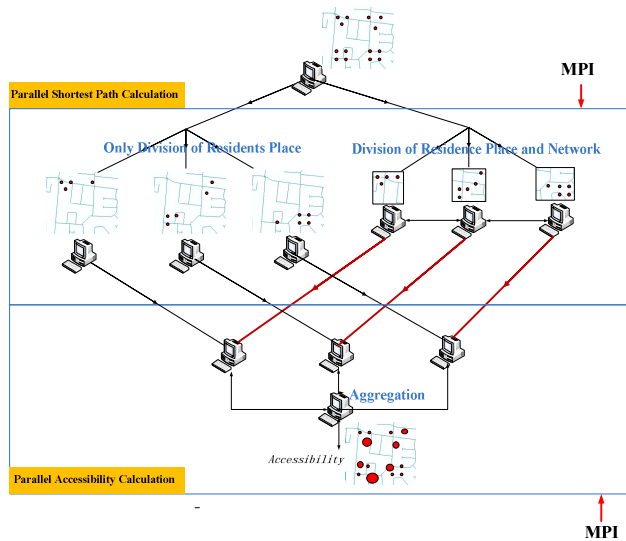


Fig.4. Two strategies of parallelization

IV. RESULTS

Due to unavailability of job type data, only distance decay and competitions are considered in the measurement of job accessibility. A couple of days are needed to work out the job accessibility in the study area on a single PC (Intel (R) Core (TM) i5 CPU 4 Cores) but the work has been massively reduced to 11 minutes over the cluster (see Fig.5). The job opportunity (equation 6) and its kernel density shown in Figure 6 are easy for communication with stakeholders as the calculated job opportunity is an absolute value. Next, Manchester Airport (Fig. 6) 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.

Fig. 7 shows the increased job opportunity across the city and 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 visualized in the same way and with same effects.

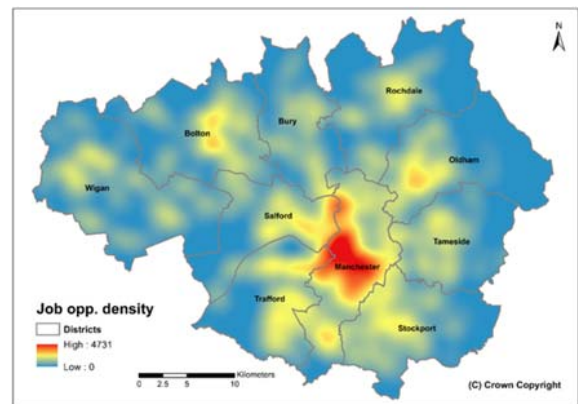
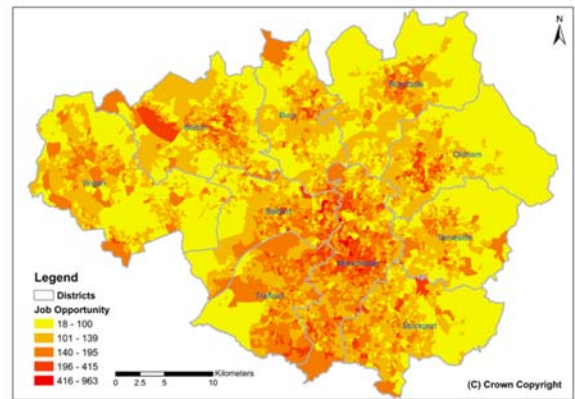


Fig.5. Spatial distribution (upper) and density (lower) of job opportunity calculated



Fig.6. Manchester Airport and output area units

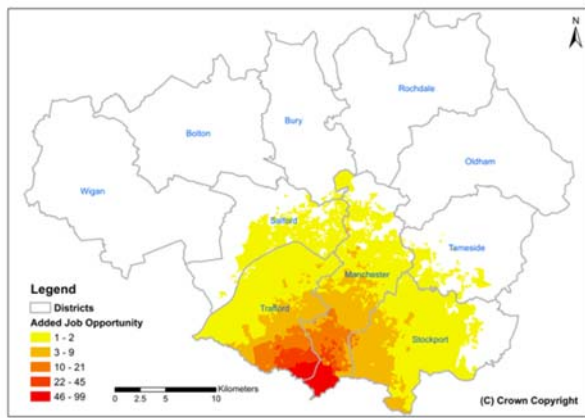
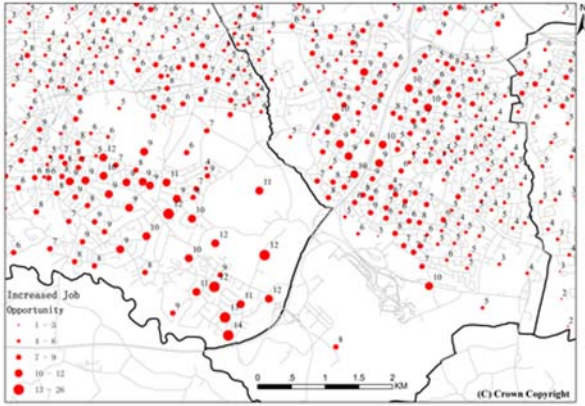


Fig.7. Outcomes of scenario for planning support: added job opportunity (upper) and zoomed areas around the airport (lower)

V. DISCUSSION AND CONCLUSION

This paper has demonstrated that parallel computation can provide methodological solutions to “big-data” challenges faced by the workshop based planning support, particularly in extensive computational time. 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 into the measurement of job accessibility. However, this paper only examined the computational solutions to planning support, which is a small part of all aspects (see Fig 1). Other forms of planning support such as more scenarios, application system development, geo-simulation (job searching behavior), visualization (at building level), and Web-based public participation should be further explored in the future. Big data analytics is providing great opportunities for planning support but the following questions should be answered theoretically, which will be further explored in the future:

- 1) How do we define and distinguish big and small planning processes?
- 2) What are the differences between big and small data sets for planning?
- 3) What are big and small models for planning analysis?
- 4) What are big and small systems for planning applications?
- 5) What are big and small users for planning participation?

ACKNOWLEDGMENT

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

REFERENCES

- [1] S. Geertman and J. Stillwell, “Planning Support Systems: an Inventory of Current Practice”, *Computers, Environment and Urban Systems*, vol. 28, 2004, pp.291-310.
- [2] R.E. Klosterman, “The What if? Collaborative Planning Support System”, *Environment and Planning B: Planning and Design*, vol. 26, 1999, pp. 393-408.
- [3] P. Waddell, “UrbanSim: Modeling Urban Development for Land Use, Transportation and Environmental Planning”. *Journal of the American Planning Association* vol.68(3), 2002, pp.297-314.
- [4] M. Batty, “Big data, Smart Cities and City Planning”, *Dialogues in Human Geography*, vol.3(3), 2013, pp.274–279.
- [5] L. Yin, S. Shaw, D. Wang, E. Carr, M. Berry, L. Gross, E. J. Comiskey, “ A Framework of Integrating GIS and Parallel Computing for Spatial Control Problems – a Case Study of Wild-fire Control”, *International Journal of Geographical In-formation Science*, vol.26(4), 2012, pp. 621-641.
- [6] J. Cheng and L. Berterloni, “ Measuring Urban Job Accessibility with Competitions, Distance Decay and Diversity”, *Journal of Transport Geography*, vol.30, 2013, pp.100-109.
- [7] J. Cheng, L. Berterloni, F. le Clercq, L. Kapoen, “Understanding Urban Networks: Comparing a Node-, a Density- and an Accessibility-based View”, *Cities: International Journal of Urban Policy and Planning*, vol.31, 2013, pp.165-176.
- [8] J. Cheng, L. Berterloni, F. le Clercq, “Measuring Sustainable Accessibility”. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2017, 2007, pp.16-25.
- [9] L. Han, C.S. Liew, J. van Hemert, M.P. Atkinson, “A Generic Parallel Processing Model for Facilitating Data Mining and Data Integration”. *Journal of Parallel Computing*, vol.37(1), 2012, pp.157-171.