Clustering Within Timetabling Conflict Graphs

Camille Beyrouthy, Edmund K. Burke
Computer Science Dept. University of Nottingham, UK, {cbb,ekb}@cs.nott.ac.uk

Barry McCollum, Paul McMullan
Queens University of Belfast, Belfast, UK, {b.mccollum,p.mccullan}@qub.ac.uk

Dario Landa-Silva, Andrew J. Parkes[1]
Computer Science Dept. University of Nottingham, UK, {jds,ajp}@cs.nott.ac.uk

A key concept in timetabling problems is that of the conflict graph with edges representing pairs of events that are not allowed to occur at the same time. Usually, the only information presented about such graphs is their density. However, intuitively, it seems likely that such graphs are structured, and most likely have some clustering. In analysing the structure of social networks or of the world-wide web it is common to use various measures. Amongst these is the "clustering coefficient". We propose using this coefficient in order to analyse timetabling conflict graphs, and give results showing that on some common benchmarks the graphs are indeed clustered by this measure.

1 Introduction and Context

In previous papers [1, 2, 3] we have studied the issue of space planning within academic institutions [5]. The problem is that currently teaching space is poorly utilised. In many institutions, teaching rooms are used only half the time, and even when used they are often only half full. Since building and maintaining rooms is expensive (the second highest institutional budgetary consideration after staff), it is not surprising that institutions would like to rectify this situation. On the other hand, excess teaching space is often requested in order to satisfy institutional timetabling requirements. This leads to the question of precisely how to manage the balance; making best use of minimal space whilst still satisfying demand. In addition, projected demand for teaching space must be considered when deciding upon future space requirements. We intend to address this by supporting institutional decision making with the following methodology. Firstly, analyse and quantitatively classify the current student enrollment and course structures. Secondly, in order to generate the test cases for analysis, create a simulator to generate course structures and student enrollments in a meaningful way. Finally, using the simulator together with appropriate course-timetabling software, run simulations under various proposed scenarios for space changes, and develop a methodology for evaluation and comparison of these scenarios. This outlined methodology is a form of "simulation optimisation" [6]. A crucial part of the process is having reasonable confidence that the simulator is realistic compared to real-world scenarios. The output of this research strives to ensure that the system is validated. The research will also explore the sensitivity of the final decisions to the assumptions underlying the simulator, to ensure that they are applicable.

In practice, this means that the simulator will be designed: (i) based on properties and patterns observable in existing instances, and (ii) validated against such patterns. However, there are few ways in order to compare simulated and real instances. Currently, the only property typically measured of a timetabling instance is the density, d, of the conflict graph. However the density is far too "blunt" a tool: graphs with similar densities might have very different structures. This suggests that better measures are needed in order to characterise and exploit the properties of conflict graphs. We aim to identify a suite of properties to measure in a timetabling instance. Such a suite will be used to ensure that the simulator uses instances that have realistic structures. Here, we start building such a suite by using standard concepts from the network analysis literature, specifically, the "clustering coefficient" of the conflict graph.

Many papers have studied the graphs resulting from social networks and the world-wide web, e.g. see [7]. One of the techniques used in such analysis is the "clustering coefficient" defined as follows.

[1]Contact Author. (Authors listed alphabetically)

Name	Exams (n)	edges (e)	Dens.(d) (%)	Clust. Coeff. c (%)
hec-s-92	81	2823	42	67.3
sta-f-83	139	611	14.4	85.8
yor-f-83	181	941	28.9	57.5
ute-s-92	184	2750	8.5	53.3
ear-f-83	190	1125	26.7	62.9
tre-s-92	261	4360	5.8	45.3
lse-f-91	381	2726	6.3	56.6
kfu-s-93	461	5349	5.6	56.6
rye-s-93	486	11483	7.5	60.2
car-f-92	543	18419	13.8	45.6
uta-s-92	622	21267	12.6	40.4
car-s-91	682	16925	12.8	40.8
pur-s-93	2419	30032	2.9	36.5

Table 1: Sizes and densities of the conflict graphs generated by the Carter instances, together with their clustering coefficients.

Suppose that the degree of node i is k_i - then there are potentially $k_i(k_i-1)/2$ edges between the neighbours of i. Let c_i be the local density of the graph between the neighbours of node i.

$$c_i = \frac{\text{num edges between neighbours of } i}{k_i(k_i-1)/2} \tag{1}$$

The overall "clustering coefficient", c, is defined as the mean value (with respect to the n nodes of the entire graph) of the clustering c_i

$$c = \frac{1}{n}\sum_{i=1}^{n} c_i \tag{2}$$

In a random graph, the edges are selected randomly and independently with probability p [4], hence the expected density of the local neighbourhood is p, the same as the overall density. We will say that the graph is clustered if the clustering coefficient, c, is higher than the overall density, d.

2 Empirical Clustering Coefficients

In the context of the conflict graph in timetabling problems, it is natural to expect that if an event A conflicts with events B and C, then the chances of B and C conflicting with each other is higher than the overall (average) density. This corresponds to expecting the conflict graph to be clustered. Table 1 gives the clustering coefficients of the standard *real-world* Carter instances of Exam Timetabling[2], and confirms that they are indeed clustered.

For the purposes of the "International Timetabling Competition (TTComp)"[3], *artificial* course timetabling problems were "designed by Ben Paechter for the Metaheuristics Network"[4]. Sixty more instances were also made publically available later and are grouped into the sets: "small", "medium" and "large". Figure 1 is a plot of clustering against density for the Carter and TTComp instances. It shows that the TTComp instances seem to fall into the same region as the Carter instances.[5] Note

[2]See `ftp://ftp.mie.utoronto.ca/pub/carter/testprob/` and `http://www.cs.nott.ac.uk/~rxq/data.htm`

[3]`http://www.idsia.ch/Files/ttcomp2002/`

[4]`http://www.idsia.ch/Files/ttcomp2002/IC_Problem/node1.html`

[5]One might object that the Carter instances are for *exam* timetabling and the TTComp for *course* timetabling. However, we do expect that course and exam timetabling will have related structures, particularly in the conflict matrix. For example, on the basis that students taking exams presumably also took the associated course. However, this issue will also be under future study.

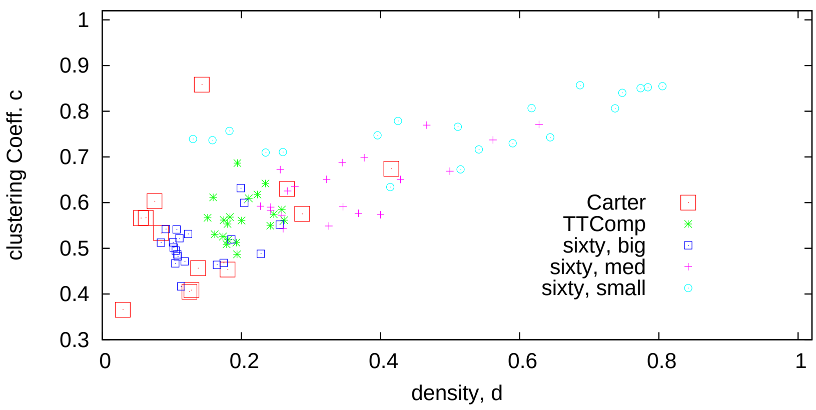

Figure 1: Each point corresponds to the conflict graph from a separate timetabling instance, plotted by their clustering coefficient, c, against their density, d.

that the 'small' artificial instances seem to have unrealistic densities. However, it is reassuring to observe that the 'large' instances do fall into the same region of the (d, c) plane as the real Carter instances. We believe this supports the argument that the TTComp instances are a reasonable test for solvers. We observe that larger problems generally have both smaller density and clustering. However, whilst the density does not seem to have a lower limit, the clustering does not drop below about 40%. That is, even when there are few conflicts they still tend to cluster significantly, with the neighbourhood of nodes having a density of about 40%. Besides being a test of the validity of artificial instances, this might also have interesting implications for the design of solvers.

We have seen that, as might be expected, timetabling conflict graphs are clustered and that the artificial instances used so far exhibit similar clustering. It is likely that the clustering will significantly affect important properties of the conflict graphs - for example, their chromatic numbers. Also, algorithms might be improved if they were to directly measure and exploit the clustering of the graphs. Current work is investigating such links and ways to exploit measured clustering. Finally, we remark that other domains such as scheduling and rostering generate ("conflict") graphs and these might well also exhibit properties that could be revealed (and exploited) by their clustering coefficients. Our findings help us understand the structure of conflict graphs and so are an important step towards improving teaching space utilisation because timetabling directly impacts this issue.

References

[1] C. Beyrouthy, E. K. Burke, B. McCollum, P. McMullan, J. D. Landa-Silva, and A. J. Parkes. The teaching space allocation problem with splitting. Proceedings of PATAT 2006.

[2] C. Beyrouthy, E. K. Burke, B. McCollum, P. McMullan, J. D. Landa-Silva, and A. J. Parkes. Towards improving the utilisation of university teaching space. Technical Report. School of Computer Science & IT, University of Nottingham, 2006.

[3] C. Beyrouthy, E. K. Burke, B. McCollum, P. McMullan, J. D. Landa-Silva, and A. J. Parkes. Understanding the role of UFOs within space exploitation. Extended abstract. Proceedings of PATAT 2006.

[4] B. Bollobas. *Random Graphs*. Academic Press, London, England, 1985.

[5] B. McCollum and P. McMullan. The cornerstone of effective management and planning of space. Technical report, Realtime Solutions Ltd, Jan 2004.

[6] J. Swisher, P. Hyden, S. Jacobson, and L. Scruben. A survey of simulation optimization techniques and procedures. In J. Joines, R. Barton, K. Kang, and P. Fishwick, editors, *Proceedings of the Winter Simulation Conference*, 2000.

[7] D. J. Watts and S. H. Strogatz. Collective dynamics 'small-world' networks. *Nature*, 393:440–442, June 1998.