Enrollment generators, clustering and chromatic numbers

Camille Beyrouthy · **Edmund K. Burke** ·
Barry McCollum · **Paul McMullan** · **Andrew J. Parkes**

Proceedings of PATAT 2008. Montreal, Canada, August 2008.

Abstract We study the timetable conflict graphs produced by an artificial generator of student enrollments. We find correlations of their chromatic number with their density and clustering coefficient. The work gives evidence that the clustering coefficient is a useful measure of a graph.

1 Introduction

A task with large financial implications for Universities is the planning and management of teaching space. There is evidence that teaching space is currently poorly utilised (HEFCE 1999), and we are developing methods with the aim of improving this situation (Beyrouthy et al 2006, 2007c,b). However, significant barriers in such development are (i) the lack of realistic data instances, and (ii) the lack of good understanding of the nature of the timetabling problems that arise in practice, and for which the space provisions need to be targetted.

For good space planning, we believe it is advantageous to be able to simulate many future scenarios, but this cannot be done well without the ability to create realistic scenarios, tailored to a particular institution. Such creation inevitably requires a good understanding of the structure of problems that arise, as well as a good ability to solve them close to optimality. A lot of work in the timetabling community has been directed at the solvers. In this project we are also studying the structures of the problems themselves.

We reported some initial studies in Beyrouthy et al (2007a) of timetabling conflict graphs: vertices represent events, and edges correspond to conflicts between events preventing them taking place at the same time. We introduced their study from the perspective of the clustering coefficient, c, of the graphs. The clustering coefficient has achieved significant usage in the study of networks (Watts and Strogatz 1998). Roughly speaking, it measures the probability that for any node in the graph, any two of its neighbours are also neighbours of each other.

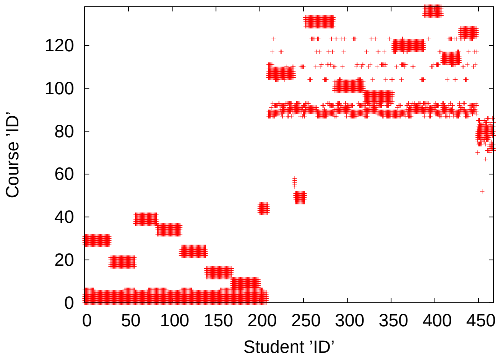
Specifically, the clustering coefficient, c_i of a node, i (of degree at least two) is the density of induced subgraph given by the set of nodes that are adjacent to i. That is, c_i is the probability that

---

Camille Beyrouthy, Edmund K. Burke, Andrew J. Parkes (contact author)
School of Computer Science, University of Nottingham, Nottingham, NG8 1BB, UK.
E-mail: { cbb, ekb , ajp }@cs.nott.ac.uk

Barry McCollum, Paul McMullan
School of Computer Science, Queen's University, Belfast, BT7 1NN, UK.
E-mail: { b.mccollum, p.p.mcmullan }@qub.ac.uk

Fig. 1 “Bi-Clustered” scatter plot of enrollments from a real instance; “sta-f-83” of the Toronto Benchmarks.



two distinct nodes adjacent to node i are also adjacent to each other. The clustering coefficient, c, of the graph is then the average of the coefficients for the nodes.[1]

This clustering naturally complements the usual density, d, which can be thought of as the probability that two randomly selected nodes are neighbours of each other. We briefly report here on further studies related to these “(density,clustering)”, (d,c), measures of a graph.

2 Student Enrollment Generator

Our study is based on developing methods to create graphs artificially but with realistic properties. Conflict graphs are likely to have significant large sets of tightly-interacting events because of courses being grouped together by year of student entry, or curriculum, etc. Hence, at the least, the values for the clustering and the density of artificial instances should be reasonably close to those we expect in a realistic timetabling problem. For example, Figure 1 shows the enrollments in an instance ““sta-f-83” from the Toronto benchmarks[2] introduced by Carter et al (1996) The ordering of the students and courses on the axes was selected so as to reveal a significant block structure by performing a simplified form of bi-clustering. (Bi-clustering is a standard problem of permuting the rows and columns of a matrix so to as reveal hidden block structures: it is used, for example, within gene-expression analysis (Madeira and Oliveira 2004).)

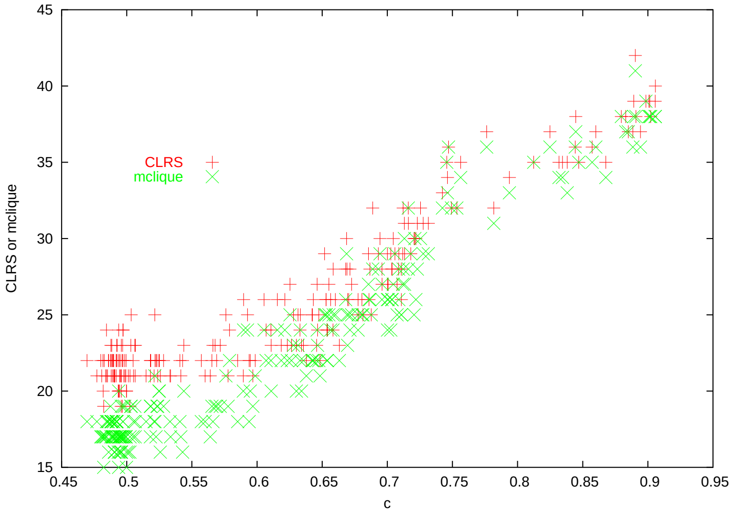
Accordingly, we designed a generator of student enrollments to mimic such blocks. Currently the generator mimics the blocks, but not yet the sub-blocks. The generator takes as part of its input a set of course sizes, for example, from a real instance (though in future work we intend to also generate these). It also has many parameters that control properties such as: the number of blocks; the maximum enrollment per student; and the amount of overlap between the blocks. Details are given in Beyrouthy (2008) where it is extensively used to study the issue of “Partial Inheritance” in sectioning (see also Beyrouthy et al (2008)). That is, it was used to study the extent to which conflicts between courses are resolved when the course is taught in sections at different times.

---

[1] The exact definitions can depend on details of how nodes of degree 0 and 1 are handled, however, in our studies, such nodes are rare, and so the differences are not important here.

[2] ftp://ftp.mie.utoronto.ca/pub/carter/testprob/

Fig. 2 "Slice plot". Scatter plot of results for many instances of conflict graphs at a fixed density. The x-axis is the clustering coefficient of each graph. The y-axis gives the number of colours, and the size of the max clique found for each graph instance.



3 Chromatic Properties

Core problems within timetabling are graph colouring and max-clique determination. Hence, it is natural to study the chromatic and clique numbers of the conflict graphs arising from the generated student enrollments.

Chromatic numbers are already well-known to (generally) increase with density. So we focussed on the effects of the clustering coefficient, by produced "slice plots" to study the variation of chromatic and clique numbers as a function of clustering at a fixed density.

Specifically, based on a fixed set of course sizes, many instances of enrollments were generated by giving a wide variety of parameters to the enrolment generator. Each instance was converted to a conflict graph. Then any graphs with a density not within a small range centered on a target density were discarded. For the remaining graphs, we measured the clustering coefficient, the best-found number of colours, and the best-found clique-size. Specifically, the clique and chromatic numbers are bounded using best results obtained by publicly available software. For the colouring we used the implementation of tabu search of Culberson[3]. For the cliques we used an implementation[4] of Reactive Local Search (Battiti and Protasi 2001).

An example of such results is given in Figure 2, based upon the course sizes from the instance "yor-f-83" of the Toronto benchmarks.

Notice that the gap between the colours needed and the max clique size is generally fairly small, and this gives a fairly tight bound on the chromatic number. This gap is small enough that we can make two significant observations:

1. Density alone does not act as a good predictor of chromatic number – the instances have essentially the same density but still have a wide range of chromatic numbers,
2. Density and clustering together do act as a good (statistical) predictor of chromatic number. Given a fixed (d,c) then the spread of the chromatic numbers obtained is fairly small compared to the mean value.

---

[3] Available from http://www.cs.ualberta.ca/~joe/Coloring/Colorsrc/index.html

[4] We used version 1.2 from http://rtm.science.unitn.it/ (but no longer available there).

Furthermore, we found that the resulting chromatic number was often fairly close to that of the original instance whose course sizes were used in order to generate the artificial enrollments. This provides evidence that the generator is capturing at least some of the salient features of the conflict graph.

4 Conclusion

We provide further evidence that the clustering coefficient of the conflict graph is a measure that should be exploited.

Acknowledgments: Andrew J. Parkes has been supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant GR/T26115/01.

References

Battiti R, Protasi M (2001) Reactive local search for the maximum clique problem. *Algorithmica* 29(4):610–637.

Beyrouthy C (2008) Models, solution methods and threshold behaviour for the teaching space allocation problem. PhD thesis, School of Computer Science, University of Nottingham

Beyrouthy C, Burke EK, Landa-Silva JD, McCollum B, McMullan P, Parkes AJ (2006) Understanding the role of UFOs within space exploitation. In: *Proceedings of the 6th International Conference on the Practice and Theory of Automated Timetabling (PATAT 2006)*, Brno, Czech Republic, pp 359–362.

Beyrouthy C, Burke EK, Landa-Silva D, McCollum B, McMullan P, Parkes AJ (2007a) Clustering within timetabling conflict graphs. In: Baptiste P, Kendall G, Munier A, Sourd F (eds) *Proceedings of the 3rd Multidisciplinary International Scheduling Conference: Theory and Applications (MISTA 2007)*, Paris, France, pp 553–556.

Beyrouthy C, Burke EK, Landa-Silva D, McCollum B, McMullan P, Parkes AJ (2007b) The teaching space allocation problem with splitting. In: Burke EK, Rudova H (eds) *Revised Selected papers from the 6th International Conference on the Practice and Theory of Automated Timetabling (PATAT 2006)*, Brno, Czech Republic, Lecture Notes in Computer Science, vol 3867, Springer-Verlag, pp 228–247.

Beyrouthy C, Burke EK, Landa-Silva JD, McCollum B, McMullan P, Parkes AJ (2007c) Towards improving the utilisation of university teaching space. *Journal of the Operational Research Society* URL http://dx.doi.org/10.1057/palgrave.jors.2602523, to appear.

Beyrouthy C, Burke EK, McCollum B, McMullan P, Parkes AJ (2008) Conflict inheritance in sectioning and space planning. In: Burke EK, Gendreau M (eds) *Proceedings of the 7th International Conference on the Practice and Theory of Automated Timetabling (PATAT 2008)*, Montreal, Canada.

Carter MW, Laporte G, Lee SY (1996) Examination timetabling: Algorithmic strategies and applications. *Journal of the Operational Research Society* 47(3):373–383.

HEFCE (1999) Estates management statistics project. Tech. rep., Higher Education Funding Council for England, report 99/18. http://www.hefce.ac.uk/pubs/hefce/1999/99_18.htm.

Madeira SC, Oliveira AL (2004) Biclustering algorithms for biological data analysis: A survey. *IEEE Transactions on Computational Biology and Bioinformatics* 1(1):24–45.

Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393:440–442.