How to find the best approximation results – a follow-up to Garey and Johnson*

Pierluigi Crescenzi† Viggo Kann‡

Abstract

A compendium of NP optimization problems, containing the best approximation results known for each problem, is available on the world wide web at

`http://www.nada.kth.se/~viggo/problemlist/`

In this paper we describe the compendium, and specify how the compendium is consultable as well as modifiable on the web. We also give statistics for the use of the compendium.

1 Introduction

The last few years have seen an extremely prolific research activity in the field of design and analysis of approximation algorithms. Actually, the notion of approximation algorithms have been considered since the very beginning of the theory of NP-completeness as a way of coping with the difficulty of solving NP-hard combinatorial optimization problems. For an account of the development of the field, see [10].

The number of both positive and negative results regarding the approximability properties of NP-hard combinatorial problems has continuously increased with an "exponential" growth rate. And it does not seem to decline. Currently, there is basically no conference or workshop in the field of algorithms and complexity that does not include in its proceedings at least two or three papers regarding the design of approximation algorithms. As a consequence, the number of problems for which approximation algorithms and/or non-approximability results have been obtained has drastically increased in the last seven years. Moreover, the degree of approximability of several important problems (mostly contained in the appendix of the book by Garey and Johnson [6]) has changed very rapidly. One single example: MAXIMUM 2-SATISFIABILITY started from the 2-approximation algorithm of Johnson [9] and in a few years reached the incredible 1.0741-approximation algorithm of Feige and Goemans [5] (which is almost the best possible performance ratio obtainable for this problem [7]).

Following this rapid development is thus becoming a hard task. This is the main motivation for collecting the existing approximation results into a compendium available to all researchers working (or starting to work) in this field. This paper describes such a compendium, and specifies how the compendium is consultable (and modifiable) on the world wide web.

*Published in ACM SIGACT News, December 1998.

†Dipartimento di Sistemi ed Informatica, Università di Firenze, Via Cesare Lombroso 6/17, 50134 Firenze, Italy, `piluc@dsi.unifi.it`

‡Department of Numerical Analysis and Computing Science, Royal Institute of Technology, SE-100 44 Stockholm, Sweden, `viggo@nada.kth.se`

Copyright © Pierluigi Crescenzi and Viggo Kann, 1998

2 A compendium of NP optimization problems

In 1993 the two authors of this paper started to collect approximation results. We soon noticed that the amount of approximability results was enormous, and therefore we had to limit ourselves to considering polynomial time approximation results for NP optimization problems that are not solvable optimally in polynomial time. This means that we considered neither approximation algorithms for #P-hard or PSPACE-hard problems, nor fast approximation algorithms for polynomially solvable problems such as MAXIMUM MULTICOMMODITY FLOW. We of course collected the best upper and lower bounds of approximation, but we did not try to find the fastest approximation algorithm giving the best approximation upper bound – any polynomial time algorithm sufficed.

With these restrictions, in 1994 we made the first version of the compendium available as a Postscript file via anonymous ftp, and announced it in the news group comp.theory. The response from the research community was fortunately enough extensive, and we received both corrections, updated results and results for new problems. Since then we have made several revisions of the compendium, and the current version has number eight. In Section 6 we explain how you can help us to improve the compendium.

We noted early that a Postscript file is not the ideal medium for the compendium. A researcher is often only interested in looking up the results for a single problem. At the end of 1994 we therefore constructed a web version of the compendium, containing exactly the same text as the Postscript file, but with hypertext links between the different sections and problems, making it a lot easier to use. Recently, we have also added links from the bibliography to papers that are available electronically. In Section 4 some statistics for the use are presented.

3 How is the compendium structured?

We have chosen to structure the problems in the same way as Garey and Johnson did [6], that is systematically in twelve categories according to subject matter. The first five categories are divided into subcategories. The following table shows how many problems there are in each category in the August 1998 version of the compendium, compared to how many there are in [6].

Code	Name of category	Number of problems in	
		G&J [6]	compendium
GT	Graph theory	65	55
ND	Network design	51	65
SP	Sets and partitions	21	14
SR	Storage and retrieval	36	10
SS	Sequencing and scheduling	22	20
MP	Mathematical programming	13	17
AN	Algebra and number theory	18	1
GP	Games and puzzles	15	2
LO	Logic	19	13
AL	Automata and language theory	21	5
PO	Program optimization	10	2
MS	Miscellaneous	19	17

For many categories the numbers are about the same. This might be surprising considering that only some of the NP-complete problems in [6] have corresponding optimization problems. The explanation is of course that several new NP-complete problems have been studied since 1979: an updated list of NP-complete problems would be much larger than the original one.

It is interesting that, in some categories, there are much fewer problems in the compendium than in [6]. The reason could either be that there are only a few optimization problems in these categories or that no approximability results have been shown for the optimization problems in these classes. The latter case suggests that more research is needed for these categories.

The current version of the compendium thus contains more than 200 problems. However, under the same basic problem, several variations are also included. A typical entry consists of eight parts: the first four parts are mandatory while the rest are optional.

1. The problem name that also specifies the goal of the problem.
2. The definition of the instances of the problem.
3. The definition of the feasible solutions of the problem.
4. The definition of the measure of a feasible solution.
5. A ‘good news’ part that contains the best approximation positive result (upper bound) for the problem.
6. A ‘bad news’ part that contains the worst approximation negative result (lower bound) for the problem.
7. A section of additional comments. In this section approximability results for variations of the problem are mentioned.
8. A reference to the ‘closest’ problem appearing in the list published in [6].

The following is an example of what an entry looks like.

GT7. MINIMUM EDGE COLORING

INSTANCE: Graph $G = \langle V, E \rangle$.
SOLUTION: A coloring of E, that is, a partition of E into disjoint sets $E_1, E_2, \ldots, E_k$ such that, for $1 \leq i \leq k$, no two edges in E_i share a common endpoint in G.
MEASURE: Cardinality of the coloring, i.e., the number of disjoint sets E_i.

Good News: Approximable within 4/3, and even approximable with an absolute error guarantee of 1 [14].
Bad News: Not approximable within $4/3-\varepsilon$ for any $\varepsilon > 0$ [8].
Comment: Also called *Minimum Chromatic Index*. APX-intermediate unless the polynomial-hierarchy collapses [4]. On multigraphs the problem is approximable within $1.1+(0.8/opt)$ [11]. The maximization variation in which the input is extended with a positive integer k, and the problem is to find the maximum number of consistent vertices over all edge-colorings with k colors, is approximable within $e/(e-1)$ [3], but does not admit a PTAS [12].
Garey and Johnson: OPEN5

4 How is the compendium used today?

The web version of the compendium on URL `http://www.nada.kth.se/~viggo/problemlist/` was between June 1997 and May 1998 used about 30 times each day, and in each session about 8 web pages were accessed. During the year more than 4000 users from 70 different countries around the world accessed the web compendium. This means that the compendium is used not only by researchers in the field, and probably it is used by non-researchers. 600 persons used it more that 10 times during the year.

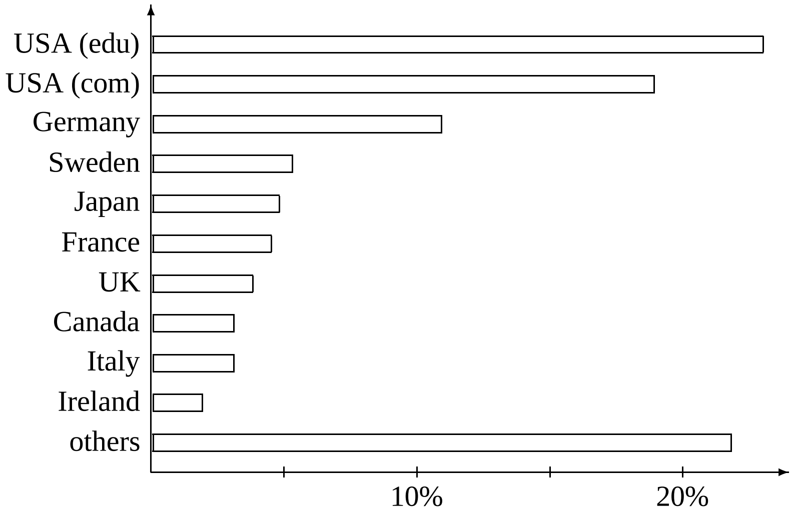

Figure 1: Use of the compendium by country.

Table 1: Most popular problems and the number of times they were looked up in a year.

GT1	MIN VERTEX COVER	891	SS1	MAX CONSTRAINED SEQUENCING...	370
ND1	MIN k-SPANNING TREE	514	GT23	MAX INDEPENDENT SET	365
GT5	MIN GRAPH COLORING	511	GP2	MIN TRAVEL ROBOT LOCALIZATION	347
ND32	MIN TRAVELING SALESPERSON	465	GP1	MIN GRAPH MOTION PLANNING	345
MP3	MAX PACKING IP	461	GT6	MAX ACHROMATIC NUMBER	338
GT2	MIN DOMINATING SET	439	ND33	MIN METRIC TSP	336
GT44	MIN LINEAR ARRANGEMENT	434	LO1	MAX SATISFIABILITY	321
SP4	MIN SET COVER	423	SR1	MIN BIN PACKING	320
GT22	MAX CLIQUE	419	ND2	MIN DEGREE SPANNING TREE	312
ND8	MIN STEINER TREE	401	LO11	MIN EQUIVALENCE DELETION	308

Figure 1 shows how much the compendium has been used from different countries.

There are links to the compendium from 400 domains. The most used links (50%) were from web indexes such as Altavista, InfoSeek, Yahoo, and Lycos. Except for the web indexes most used links were from www.nada.kth.se (domain of the compendium), www.ing.unlp.edu.ar (TSPBIB), and www.cs.pitt.edu (algorithm web page by Kirk Pruhs).

Table 1 shows the problems in the compendium that are most popular, that is, have been looked up most times. This might give an indication of which problems are hot. In the list we find several of the most basic NP-hard problems, but interestingly also MINIMUM LINEAR ARRANGEMENT and MINIMUM TRAVEL ROBOT LOCALIZATION. The least popular problems are shown in table 2.

5 Two peculiar statistics

Besides being a tool to look for the best known approximation results regarding specific problems, the compendium can also be used to perform statistics of different kinds. In this section, we give two examples of such statistics.

Table 2: Least popular problems.

ND58	MIN DIAMETERS DECOMPOSITION	62	MP11	MIN UNSATISFYING LINEAR SUBSYSTEM	76
ND59	MAX k-FACILITY DISPERSION	65	MP17	MIN BLOCK-ANGULAR CONVEX PROG.	77
SR8	MAX COMMON POINT SET	73	ND62	MIN k-SWITCHING NETWORK	78
MS7	MIN k-LINK PATH IN A POLYGON	73	MP12	MAX HYPERPLANE CONSISTENCY	81
ND65	MIN SEPARATING SUBDIVISION	75	MS8	MIN SIZE ULTRAMETRIC TREE	81

Table 3: The distribution of performance ratios.

Performance ratio	**Number of problems**
FPTAS	7
PTAS	12
APX, $\leq 4/3$	10
APX, $\leq 3/2$	9
APX, < 2	18
APX, 2	25
APX, > 2	20
polylog	38
poly	26

The oldest results The oldest result contained in the compendium is a 4/3-approximation algorithm for MINIMUM EDGE COLORING due to [14]. Actually, this algorithm is optimal since in [8] it is shown that, for any $\varepsilon > 0$, no $(4/3-\varepsilon)$-approximation algorithm exists for this problem. It is worth pointing out that in neither of the above two references the notion of approximation algorithm is explicitly used.

The second oldest result, instead, is due to [9] which is widely considered the starting point of the theory of approximation algorithm (indeed, the compendium contains *four* results due to this paper). In particular, the above reference contains a $(1+\ln n)$-approximation algorithm for MINIMUM SET COVER: 25 years later it has been proved that this algorithm is optimal [13]!

Distribution of performance ratios Another interesting statistic is the number of problems that admit a specific performance ratio. The distribution of the performance ratios within the compendium is shown in Table 3 (note that the total number of problems in the table does not coincide with the total number of problems in the compendium since, for some problems, there are no good news at all). It is surprising that so many problems admit a 2-approximation algorithm: moreover, for most of them the algorithm is not known to be optimal.

6 The future of the compendium

The above numbers show that the compendium seems to be considered very useful by researchers and encourage us to continue maintaining it. Updating and completing the compendium, however, is an endless work, so there will never be a final version. The compendium will be part of a new book on approximation [2], but the web version will continue to evolve after the book is printed.

It is impossible for the authors to keep up with all new approximation results presented at conferences and published in journals without help from others. In the future we will trust that any researcher publishing a new result that would fit in the compendium will report it to us. In order to facilitate this communication we have created four web forms for

1. entering results for a problem that already is in the compendium,
2. entering results for a new problem,
3. updating bibliography entries,
4. reporting errors.

The forms are available from the home page of the compendium and are designed to be easy to use.

In order to guarantee that no important results are missed we will from November 1998 engage three researchers as subeditors. Each subeditor takes care of certain sections of the compendium. The three subeditors are

- Magnus Halldórsson, *Graph Theory*: Covering and Partitioning, Subgraphs and Supergraphs, *Sets and Partitions*.
- Marek Karpinski, *Graph Theory*: Vertex Ordering, *Network Design*: Cuts and Connectivity.
- Gerhard Woeginger, *Sequencing and Scheduling*.

7 Conclusion

In this paper we have briefly described a compendium of NP optimization problems that is available on the web. We believe that such a compendium will turn out to be very useful whenever someone has to deal with the approximate solution of an NP-hard optimization problem. Indeed, as stated in [1], "the first step in proving an inapproximability result for a given problem is to check whether it is already known to be inapproximable." The compendium is then the right starting point to perform this test (actually, this is true also for positive results). Moreover, even if the problem at hand is not contained in the compendium, it is likely that it contains other problems that can be reduced to this problem.

We have also described how the compendium can now be updated directly on the web by means of four forms that have been designed to be easy to use. We hope that in this way researchers will be stimulated to help us in maintaining the compendium as up-to-date as possible: the sooner they report a result the sooner it will be publicly known!

References

[1] Arora, S., and Lund, C. (1997), *Hardness of approximations,* In Hochbaum, D. S., ed., *Approximation Algorithms for NP-hard Problems,* chapter 10, PWS Publishing Company, Boston.

[2] Ausiello, G., Crescenzi, P., Gambosi, G., Kann, V., Marchetti Spaccamela, A., and Protasi, M. (1999), *Approximate solution of NP-hard optimization problems,* Springer-Verlag, To appear.

[3] Bertsimas, D., Teo, C-P., and Vohra, R. (1996), "On dependent randomized rounding algorithms", *Proc. 5th Int. Conf. on Integer Prog. and Combinatorial Optimization,* Lecture Notes in Comput. Sci. 1084, Springer-Verlag, 330–344.

[4] Crescenzi, P., Kann, V., Silvestri, R., and Trevisan, L. (1995), "Structure in approximation classes", *Proc. 1st Ann. Int. Conf. on Computing and Combinatorics*, Lecture Notes in Comput. Sci. 959, Springer-Verlag, 539–548.

[5] Feige, U., and Goemans, M. X. (1995), "Approximating the value of two prover proof systems, with applications to MAX 2SAT and MAX DICUT", *Proc. 3rd Israel Symp. on Theory of Computing and Systems*, IEEE Computer Society, 182–189.

[6] Garey, M. R., and Johnson, D. S. (1979), *Computers and Intractability: a guide to the theory of NP-completeness*, W. H. Freeman and Company, San Francisco.

[7] Håstad, J. (1997), "Some optimal inapproximability results", *Proc. 29th Ann. ACM Symp. on Theory of Comp.*, ACM, 1–10.

[8] Holyer, I. (1981), "The NP-completeness of edge-coloring", *SIAM J. Comp.* **10**, 718–720.

[9] Johnson, D. S. (1974), "Approximation algorithms for combinatorial problems", *J. Comput. System Sci.* **9**, 256–278.

[10] Kann, V., and Panconesi, A. (1997), "Hardness of approximation", In Dell'Amico, M., Maffioli, F., and Martello, S., ed., *Annotated Bibliographies in Combinatorial Optimization*, chapter 2, Wiley, 13–30.

[11] Nishizeki, T., and Kashiwagi, K. (1990), "On the 1.1 edge-coloring of multigraphs", *SIAM J. Disc. Math.* **3**, 391–410.

[12] Petrank, E. (1994), "The hardness of approximation: gap location", *Computational Complexity* **4**, 133–157.

[13] Raz, R., and Safra, S. (1997), "A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP", *Proc. 29th Ann. ACM Symp. on Theory of Comp.*, ACM, 475–484.

[14] Vizing, V. G. (1964), "On an estimate of the chromatic class of a p-graph", *Diskret. Analiz.* **3**, 23–30.