Search in a Small World

Toby Walsh
Department of Computer Science
University of Strathclyde
Glasgow G1 1XL
Scotland
tw@cs.strath.ac.uk

Abstract

In a graph with a "small world" topology, nodes are highly clustered yet the path length between them is small. Such a topology can make search problems very difficult since local decisions quickly propagate globally. We show that graphs associated with many different search problems have a small world topology, and that the cost of solving such search problems can have a heavy-tailed distribution. The strategy of randomization and restarts appears to eliminate these heavy tails. A novel restart schedule in which the cutoff bound is increased geometrically appears particularly effective.

1 Introduction

Graphs that occur in many biological, social and man-made systems are often neither completely regular nor completely random, but have instead a "small world" topology in which nodes are highly clustered yet the path length between them is small [Watts and Strogatz, 1998]. By comparison, random graphs with a similar number of nodes and edges have short path lengths but little clustering, whilst regular graphs like lattices tend to have high clustering but large path lengths. A small world topology can have a significant impact on properties of the graph. For instance, if you are introduced to someone at a party in a small world, you can usually find a short chain of mutual acquaintances connecting you together.

One reason for the occurrence of small world graphs is that it only takes a few short cuts between neighbourhood cliques to turn a large world (in which the average path length between nodes is large) to a small world (in which the average path length is small). Watts and Strogatz have shown that a social graph (the collaboration graph of actors in feature films), a biological graph (the neural network of the nematode worm *C. elegans*) and a man-made graph (the electrical power grid of the western United States) all have a small world topology. In a simple model of an infectious disease. they demonstrate that disease spreads much more easily and quickly in a small world. A small world topology may therefore have a significant impact on the behavior of dynamical systems. How do they affect search problems?

2 Testing for a small world

To formalize the notion of a small world, Watts and Strogatz define the clustering coefficient and the characteristic path length. The path length is the number of edges in the shortest path between two nodes. The characteristic path length, L is the path length averaged over all pairs of nodes. The clustering coefficient is a measure of the cliqueness of the local neighbourhoods. For a node with k neighbours, then at most $k(k-1)/2$ edges can exist between them (this occurs if they form a k-clique). The clustering of a node is the fraction of these allowable edges that occur. The clustering coefficient, C is the average clustering over all the nodes in the graph.

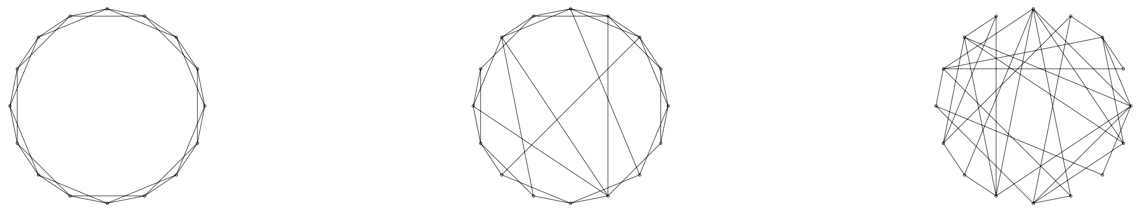
Watts and Strogatz define a small world graph as one in which $L \gtrsim L_{rand}$ and $C \gg C_{rand}$ where L_{rand} and C_{rand} are the characteristic path length and clustering coefficient of a random graph with the same number of nodes n and edges e. Rather than this simple *qualitative* test, it might be useful to have a *quantitative* measure of "small worldliness". We can then compare the topology of different graphs. To this end, we define the proximity ratio μ as the ratio of C/L normalized by C_{rand}/L_{rand}. In graphs with a small world topology, the proximity ratio $\mu \gg 1$. By comparison, the proximity ratio μ is unity in random graphs, and small in regular graphs like lattices. In table 1, we show that the proximity ratio, μ is large in those graphs studied in [Watts and Strogatz, 1998] with a small world topology.

	L	L_{rand}	C	C_{rand}	μ
film actors	3.65	2.99	0.79	0.00027	2396
power grid	18.7	12.4	0.080	0.005	10.61
C. elegans	2.65	2.25	0.28	0.05	4.755

Table 1. Characteristic path lengths, clustering coefficients and proximity ratios for graphs studied in [Watts and Strogatz, 1998] with a small world topology.

3 Modeling a small world

Watts and Strogatz propose a model for small world graphs. Starting from a regular graph, they introduce disorder into the graph by randomly rewiring each edge with probability p. If $p = 0$ then the graph is completely regular and ordered. If $p = 1$ then the graph is completely random and disordered. Intermediate values of p give graphs that are neither completely regular nor completely disordered. Watts and Strogatz start from a ring



$p = 0$
ring lattice

$0 > p > 1$
small world

$p = 1$
random graph

Figure 1. Random rewiring of a regular ring lattice. We start a ring of n nodes, each connected to the k nearest neighbours. With probability p, we randomly rewire each edge. For clarity, $n = 16$ and $k = 4$ in the above example. However, larger n and k are used in the rest of this paper.

lattice with n nodes and k nearest neighbours. They observe similar qualitative behavior with other initial regular graphs and with other mechanisms for introducing disorder. A rewired edge is reconnected to a node chosen uniformly at random from the lattice. If rewiring an edge would create a duplicate edge, they leave it unchanged.

To focus on large, sparse graphs, they demand that $n \gg k \gg \ln(n) \gg 1$, where $k \gg \ln(n)$ ensures that the graphs remain connected. For such graphs, $C \approx 3/4$ and $L \approx n/2k \gg 1$ for $p = 0$, and $C \approx k/n \ll 1$ and $L \approx \ln(n)/\ln(k)$ for $p = 1$. Note that the proximity ratio $\mu \approx 3\ln(n)/2\ln(k)$ for $p = 0$ (and this is small as $k \gg \ln(n)$ and hence $\ln(k) \approx \ln(n)$) and $\mu = 1$ by definition for $p = 1$. That is, graphs do not have a small world topology for $p = 0$ and $p = 1$. As p increases from 0, the characteristic path length drops sharply since a few long-range edges introduce short cuts into the graph. These short cuts have little effect on the clustering coefficient. As a consequence the proximity ratio rises rapidly and the graph develops a small world topology. As p approaches 1, the neighbourhood clustering start to break down, and the short cuts no longer have a dramatic effect at linking up nodes. The clustering coefficient and the proximity ratio therefore drop, and the graph loses its small world topology. These topological changes are clearly visible in Figure 2.

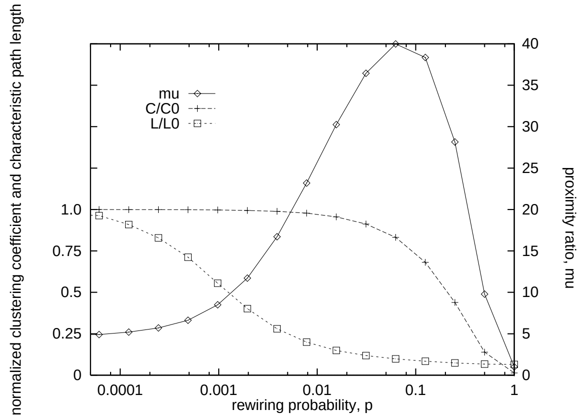


Figure 2. Characteristic path length, clustering coefficient (left axis, normalized by the values for a regular lattice) and proximity ratio (right axis) for a randomly rewired ring lattice. As in [Watts and Strogatz, 1998], we use $n = 1000$ and $k = 10$. We vary $\log_2(p)$ from -15 to 0 in steps of 1, and generate 100 graphs at each value of p. A logarithmic horizontal scale helps to identify the interval in which the characteristic path length drops rapidly, the clustering coefficient remains almost constant, and the proximity ratio, μ peaks.

	L	L_{rand}	C	C_{rand}	μ
fpsol2i.1	1.677	1.915	0.906	0.0949	10.902
zeroini.1	1.479	1.815	0.883	0.185	5.857
mulsoli.1	1.586	1.799	0.887	0.203	4.956

Table 2. Characteristic path lengths and clustering coefficients for some of the DIMACS graph coloring benchmarks.

4 Search problems

There are many search problems in AI that involve graphs (for example, the constraint graph in a constraint satisfaction problem, and the adjacency graph in a Hamiltonian circuit problem). Do such graphs have a small world topology? Does this have an impact on the hardness of solving problems? If so, can we design algorithms to take advantage of the topology?

4.1 Graph coloring

One search problem directly affected by the structure of an underlying graph is graph coloring. We therefore tested the topology of some graph coloring problems from the DIMACS benchmark library. We focused on the register allocation problems as these are based on real code. Table 2 demonstrates that these problems have large clustering coefficients like regular graphs, yet small characteristic path lengths like random graphs. They therefore have a small world topology. We observed similar results with other problems from the DIMACS benchmark library.

4.2 Time tabling

Many time-tabling problems can be naturally modelled as graph coloring problems. We therefore tested some real world time-tabling problems from the Industrial Engineering archive at the University of Toronto. Table 3 demonstrates that sparse problems in this dataset have large clustering coefficients like regular graphs, but small characteristic path lengths like random graphs. They therefore have a small world topology. By comparison, dense problems from this dataset have nodes of large degree which are less clustered. Such graphs therefore have less of a small world topology. We conjecture that graphs will often start with a sparse small world topology but will become more like dense random graphs as edges added "saturate" the structure.

sparse graphs	L	L_{rand}	C	C_{rand}	μ
LSE	2.484	2.149	0.575	0.0625	7.960
U Toronto	2.419	2.161	0.534	0.0855	5.579
St Andrews	1.872	1.808	0.867	0.196	4.272

dense graphs	L	L_{rand}	C	C_{rand}	μ
Earl Haig Col.	1.814	1.729	0.639	0.271	2.247
York Mill Col.	1.749	1.697	0.578	0.303	1.851
Ecole H Etudes	1.587	1.583	0.680	0.417	1.627

Table 3. Characteristic path lengths and clustering coefficients for time-tabling benchmarks from the University of Toronto archive. Sparse problems have an edge density of less than 15%. Dense graphs have an edge density of more than 25%.

4.3 Quasigroup problems

A quasigroup is a Latin square, a m by m multiplication table in which each entry appears just once in each row or column. Quasigroups model a variety of practical problems like tournament scheduling and designing drug tests. AI search techniques have been used to answer some open questions in finite mathematics about the existence (or non-existence) of quasigroups with particular properties [Fujita *et al.*, 1993]. More recently, Gomes and Selman have proposed a class of quasigroup problems as a challenging benchmark for constraint satisfaction algorithms [Gomes and Selman, 1997].

An order m quasigroup problem can be represented as a constraint satisfaction problem with m^2 variables, each with a domain of size m. The constraint graph for such a problem (see figure 3) consists of $2m$ cliques, one for each row and column, with each clique being of size m. Calculation shows that, for large m, such constraint graphs have a small world topology. As any pair

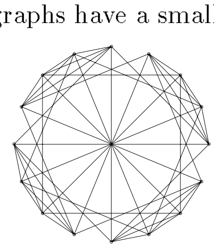


Figure 3. The constraint graph of a m by m quasigroup problem for $m = 4$. The graph has m^2 nodes and $m^2(m-1)$ edges. The edges form $2m$ cliques, each of size m.

of entries in a quasigroup either directly constrain each other or indirectly through at most one intermediate entry, the characteristic path length, $L \approx 2$ is small. A random graph with the same number of edges and nodes also has a characteristic path length $L_{rand} \approx 2$. As each variable has $2m - 1$ neighbours and these form 2 m-cliques, the clustering coefficient, $C \approx 1/2$ is large. A random graph with the same number of edges and nodes has a smaller clustering coefficient $C_{rand} \approx 2/m$. As $L \approx L_{rand}$ and $C \gg C_{rand}$ for large m, the constraint graph has a small world topology, with a proximity ratio $\mu \approx m/4$. Computation confirms these calculations (see table 4).

m	L	L_{rand}	C	C_{rand}	$m/4$	μ
20	1.905	1.929	0.486	0.0952	5	5.169
16	1.882	1.908	0.483	0.118	4	4.150
12	1.846	1.874	0.476	0.154	3	3.138

Table 4. Characteristic path lengths and clustering coefficients for m by m quasigroup problems.

5 Search cost

Graphs with a small world topology demonstrate that local properties (i.e. clustering) can be bad predictors for global property (i.e. characteristic path length). Unfortunately, heuristics often use local properties to guide the search for a (global) solution. For example, the Brelaz heuristic colors the node with the least available colors which is connected to the most uncolored nodes. Because of this mismatch between local and global properties, a small world topology may mislead heuristics and make search problems hard to solve. To test this thesis, we colored graphs generated according to Watts and Strogatz's model using an algorithm due to Mike Trick. which is based upon Brelaz's DSATUR algorithm [Brelaz, 1979]. To ensure that problems were of a manageable size for this algorithm, we used graphs with $n = 100$ and $k = 8$. For these graphs, the proximity ratio peaks around $\log_2(p) \approx -4$ similar to Figure 2. Whilst most graphs of this size can be colored without too much search, one graph was not solved in 10^{10} nodes and more than a week of CPU time. We therefore imposed a search cutoff at 10^8 nodes. On a 133 MHz Pentium, this is approximately 1 hour of computation. As we distributed our experiments over a variety of networked computers, we do not report runtimes but use nodes visited in the backtracking search tree. On any given machine, runtimes are roughly proportional to the number of nodes searched.

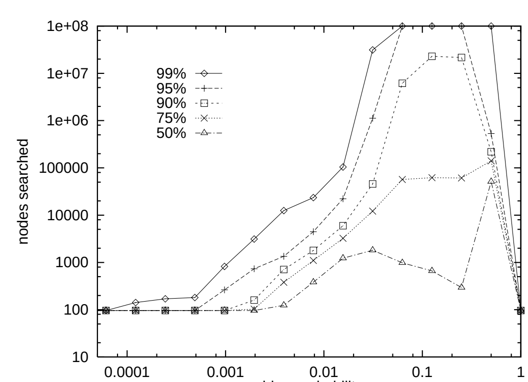


Figure 4. Cost to color graphs generated according to Watts and Strogatz's model with $n = 100$, $k = 8$ and $\log_2(p)$ varied from -15 to 0 in steps of 1.

In Figure 4, we plot the search cost against the rewiring probability, p. As p increases from 0 and graphs develop a small world topology, the search cost in the higher percentiles rises rapidly. However, as p approaches 1 and graphs lose their small world topology, the search cost in the higher percentiles falls back. For graphs with a small world topology, most problems took less than 10^5 nodes to color. However, 1% took more than 10^8 nodes. We suspect that problems with a small world topology can be difficult to color since local decisions quickly propagate globally.

6 Heavy-tailed distributions

Exceptionally hard problems like this have been observed in other search problems, though usually with less frequency. For example, Gent and Walsh found that a few satisfiability problems from the constant probability model were orders of magnitude harder than the median [Gent and Walsh, 1994]. Grant and Smith also found a few exceptionally hard problems in a random model of binary constraint satisfaction problems [Grant and Smith, 1995]. More recently, Gomes and Selman have observed similar behavior in quasigroup completion problems (whose constraint graph, we recall, has a small world topology) [Gomes *et al.*, 1997] and in tournament scheduling, planning and circuit synthesis problems [Gomes *et al.*, 1998]. They show that such behavior can often be modeled by a "heavy-tailed" distribution of the Pareto-Lévy form. In such a distribution, the tail has the form,

$$Pr(X > x) \sim C.x^{-\alpha}$$

where $Pr(X > x)$ is the probability that the variable X exceeds some value x, and $\alpha > 0$ is a constant called the "index of stability". If $\alpha < 2$ then any moment of X of order less than α is finite but higher orders are infinite[1]. For example, if $\alpha = 1.4$, then X has finite mean but infinite variance.

To test for such heavy-tails in the distribution of search costs, we plot the nodes searched against the probability that the search takes more than this number of nodes using log scales. A heavy-tailed distribution of the Pareto-Lëvy form gives a straight line with gradient $-\alpha$. In Figure 5, we plot the distribution of search costs at a variety of different values of $\log_2(p)$. This figure shows that for graphs with a small world topology (i.e. $\log_2(p) = -4$), we can model the distribution of search costs by a heavy-tailed distribution of the Pareto-Lëvy form. The gradient suggests that the index of stability, $\alpha < -1/2$. That is, the model of the distribution of search costs has infinite mean and variance. By comparison, for random graphs (i.e. $\log 2(p) = 0$) and more structured graphs (i.e. $\log_2(p) = -8$), search costs in the tail of the distribution drop more rapidly.

As in [Gomes *et al.*, 1997; 1998], we observe similar heavy-tail behavior even when solving a single problem instance. We picked the first problem from the sample

[1]Backtracking algorithms like DSATUR have an upper bound on their running time that is exponential in the problem size. The mean or variance in their running time cannot therefore be infinite. However, for large problem instances, the upper bound may be so astronomically large that we can model it as if it were infinite.

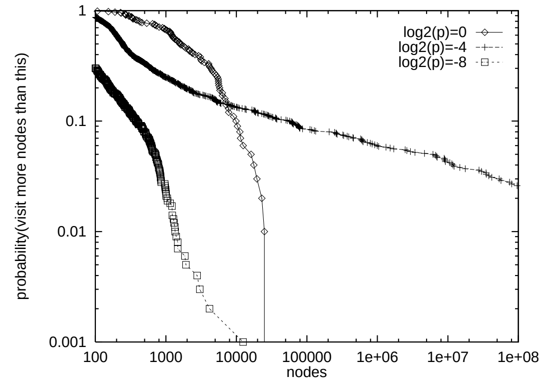


Figure 5. Log-log plot of distribution of search costs to color graphs generated according to Watts and Strogatz's model with $n = 100$, $k = 8$ and $\log_2(p) = 0, -4, -8$.

generated at $\log_2(p) \approx -4$ which took more than 10^8 nodes to solve. We re-solved this problem 1000 times, randomizing the order of the nodes in the graph so that the Brelaz heuristic makes different branching decisions. We again observe heavy-tailed behavior, with 77% of presentations of this problem taking less than 300 nodes to solve, but the rest taking more than 10^7 nodes.

A small world topology appears therefore to have a significant impact on the cost to solve graph problems. In particular, it often introduces a heavy-tail into the distribution of search costs. We conjecture that a small world topology will have a similar effect on other types of search problems involving graphs. For example, we predict that the distribution of costs for finding a Hamiltonian circuit in a graph with a small world topology will often display heavy-tailed behavior. As a second example, we anticipate that exceptionally hard problems will be more common in constraint satisfaction problems whose constraint graph has a small world topology than those with a purely random topology. Gomes, Selman and Kautz's state that "Further studies are needed to determine what characteristics of combinatorial search problems lead to heavy-tailed behavior" page 435, [Gomes *et al.*, 1998]. Our experiments suggest that a small world topology may be one answer to this question.

7 Randomization and restarts

To combat heavy-tailed distributions like this, Gomes, Selman and Kautz propose the *RRR* strategy of randomization and rapid restart [Gomes *et al.*, 1998]. This strategy prevents a search procedure from being trapped in the long tail to the right of the median, and exploits any heavy-tail to the left of the distribution (i.e. occasional runs that succeed very quickly). Does such a strategy help when searching graphs with a small world topology? Does it eliminate this heavy-tailed behavior? In Figure 6, we plot the search cost using the *RRR* strategy against the cuttoff parameter on the problem at $\log_2(p) = -4$ mentioned in the previous section. This problem has a small world topology and took the chronological backtracking search procedure over 10^8 nodes to solve. With a cutoff limit of between 200 to 400 nodes,

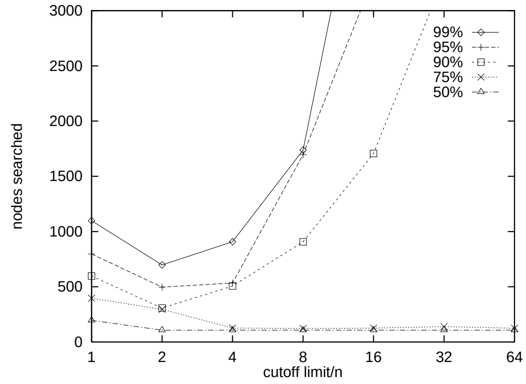


Figure 6. The effect of randomization and rapid restarts on the cost to color a graph generated according to Watts and Strogatz's model with $n = 100$, $k = 8$ and $\log_2(p) = -4$. Search is restarted after a fixed number of nodes have been visited.

the *RRR* strategy eliminates the heavy-tail in the search cost distribution.

To determine a good cutoff value for the restart strategy, Gomes *et al.* suggest "... a trial-and-error process, where one experiments with various cutoff values, starting at relatively low values ..." [Gomes *et al.*, 1998]. To help automate this process, we suggest the *RGR* strategy of randomization and *geometric* restarts. Each new trial has a cutoff value which is a constant factor, r larger than the previous. This strategy profits from the success rate being high when the cutoff value is close to optimal. Increasing the cutoff value geometrically ensures that we get close to the optimal value within a few restarts. We then hope to solve the problem within a few more restarts, before the cutoff value has increased too far from optimum. Figure 7 shows that the *RGR* strategy is even better at reducing search than the *RRR* strategy on this problem instance. The *RGR* strategy was also less sensitive to the setting of the cutoff parameters than the *RRR* strategy. We obtained similar improvements on other problem instances. It would be interesting to see if the *RGR* strategy is effective on other search problems like planning and scheduling.

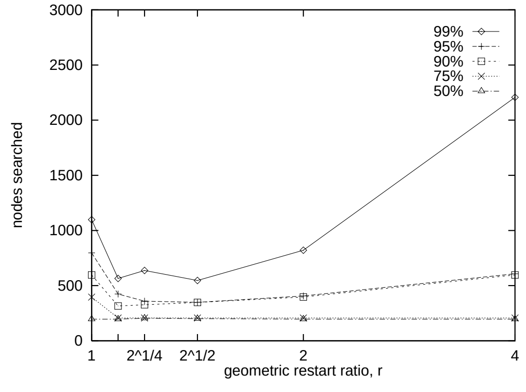


Figure 7. The effect of randomization and geometric restarts on the cost to color a graph generated according to Watts and Strogatz's model with $n = 100$, $k = 8$ and $\log_2(p) = -4$. On the ith restart, search is cutoff after $n.r^i$ nodes have been visited. We use the same problem and the same y scales as in Figure 6 to aid comparison.

8 Approximate entropy

Hogg notes that real constraint satisfaction problems often have variables that are more clustered than in random problems [Hogg, 1998]. He uses a notion of "approximate entropy" to distinguish between problems drawn from a clustered ensemble and those from a random ensemble. Can approximate entropy act as a replacement measure for the clustering coefficient?

The approximate entropy is a measure of the similarity between substructures in a problem. It therefore depends on the representation used. As in [Hogg, 1998], we consider the substructures of a graph to be subgraphs, and compare them up to isomorphism. Consider the d distinct subgraphs of size m. If f_i is the frequency of the ith distinct subgraph, and $\Phi^m = -\sum_{i=1}^{d} f_i \log f_i$ then the approximate entropy for subgraphs of size m is defined by,

$$S^m \quad =_{\text{def}} \quad \Phi^{m+1} - \Phi^m \qquad (1)$$

This measures the log-likelihood that two subgraphs of size $m + 1$ are isomorphic given that they contain subgraphs of size m that are isomorphic.

As there are only two distinct subgraphs of size 2 up to isomorphism (the 2 node subgraph with an edge, and the one without) $S^1 = \Phi^2 = -f\log(f) - (1-f)\log(1-f)$ where $f = 2e/n(n-1)$. Hence, S^1 depends only on n and e, and not on the topology of the graph. We will need to consider S^m for $m \geq 2$ to distinguish random graphs from clustered ones. Fortunately, the number of non-isomorphic subgraphs grows rapidly with m, so it is usually adequate to consider S^2 or S^3. In Figure 8, we plot S^1, S^2 and S^3 for randomly rewired ring lattices generated by Watts and Strogatz's model. As we take logarithms to base 2, the approximate entropy is measured in bits. The approximate entropy for subgraphs up to size 3 shows no obvious correlation with the rewiring probability, p (and therefore the clustering coefficient). Unfortunately, computing the approximate entropy for larger subgraphs is not computationally practical as it involves considering all subgraphs of size 5 (or, at least, a representitive sample of them).

9 Related work

Grant and Smith studied binary constraint satisfaction problems with a "braided" constraint graph similar to a ring lattice [Grant and Smith, 1995]. In this problem class, exceptionally hard problems occur more frequently and with greater vigor than in a purely random ensemble. They suggest that, as problems met in practice are likely to have constraint graphs with this sort of structure, such problems deserve further investigation. Our research supports and refines these observations.

To generate graphs with more realistic structures, Hogg has proposed a clustered ensemble based on grouping the nodes into a tree-like structure [Hogg, 1996]. In a random ensemble, each graph with n nodes and e edges is equally likely. In Hogg's clustered ensemble, an ultrametric distance between the n nodes is defined by grouping them into a binary tree and measuring the distance

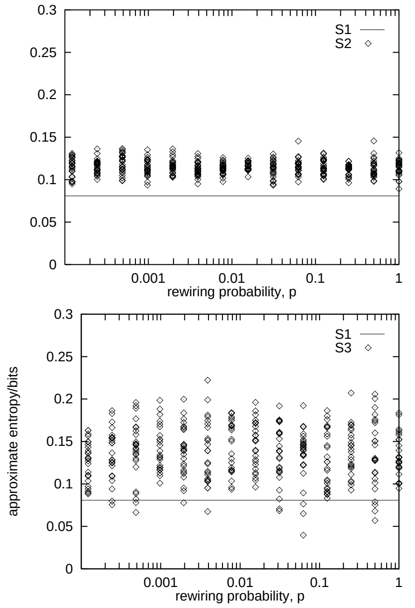


Figure 8. Approximate entropy for randomly rewired ring lattices generated according to Watts and Strogatz's model. As in Figure 2, $n = 1000$ and $k = 10$ and edges in the ring lattice are rewired with probability p. We vary $\log_2(p)$ from -13 to 0 in steps of 1, and generate 25 graphs at each value of p.

up this tree to a common ancestor. A pair of nodes at ultrametric distance d is joined by an edge with relative probability p^d. If $p = 1$, graphs are purely random. If $p < 1$, graphs have a hierarchical clustering as edges are more likely between nearby nodes. These graphs tend to have characteristic path lengths and clustering coefficients similar to random graphs. Although nodes at the top of the tree tend to be more clustered than in random graphs, nodes lower down tend to be less clustered. We therefore need a more refined notion than the clustering coefficient to distinguish their topology from that of random graphs.

10 Conclusions

In a graph with a small world topology, nodes are highly clustered yet the path length between them is small. Such a topology can make search problems very difficult since local decisions quickly propagate globally. To provide a quantitative measure of the extent to which a graph has this topology, we have proposed the proximity ratio, μ. This is the ratio of the clustering coefficient and the characteristic path length, normalized by the values for a random graph with the same number of edges and numbers. Using this measure, we have shown that many graphs associated with search problems have a small world topology, and that the cost of solving such search problems can have a heavy-tailed distribution. As in other studies, randomization and restarts appear to eliminate these heavy tails.

What general lessons can be learnt from this study? First, search problems met in practice may be neither completely structured nor completely random. Since algorithms optimized for purely random problems may perform poorly on problems that contain both structure and randomness, it may be useful to benchmark with problem generators that introduce both structure and randomness. Second, simple topological features can have a large impact on the cost of solving search problems. In particular, search problems involving graphs with a small world topology can sometimes be very difficult to solve. It may therefore be useful to optimize algorithm performance for such topological features. Third, randomization and restarts is again an effective strategy to tackle heavy-tailed distributions. The *RGR* strategy in which the cutoff value increases geometrically shows considerable promise. And finally, it really does seem that we live in a small world.

References

[Brelaz, 1979] D. Brelaz. New methods to color the vertices of a graph. *Communications of ACM*, 22:251–256, 1979.

[Fujita *et al.*, 1993] Masayuki Fujita, John Slaney, and Frank Bennett. Automatic generation of some results in finite algebra. In *Proc. of the 13th IJCAI*, pages 52–57, 1993.

[Gent and Walsh, 1994] Ian P. Gent and Toby Walsh. Easy problems are sometimes hard. *Artificial Intelligence*, pages 335–345, 1994.

[Gomes and Selman, 1997] C. Gomes and B. Selman. Problem structure in the presence of perturbations. In *Proc. of the 14th National Conference on AI*, pages 221–226, 1997.

[Gomes *et al.*, 1997] C. Gomes, B. Selman, and N. Crato. Heavy-tailed distributions in combinatorial search. In G. Smolka, editor, *Proc. of 3rd International Conference on Principles and Practice of Constraint Programming (CP97)*, pages 121–135. Springer, 1997.

[Gomes *et al.*, 1998] C. Gomes, B. Selman, and H. Kautz. Boosting combinatorial search through randomization. In *Proc. of 15th National Conference on Artificial Intelligence*, pages 431–437, 1998.

[Grant and Smith, 1995] S. Grant and B.M. Smith. Where the *Exceptionally* Hard Problems Are. In *Proc. of the CP-95 workshop on Really Hard Problems*, 1995.

[Hogg, 1996] T. Hogg. Refining the phase transition in combinatorial search. *Artificial Intelligence*, 81(1–2):127–154, 1996.

[Hogg, 1998] T. Hogg. Which search problems are random? In *Proc. of the 15th National Conference on Artificial Intelligence*, pages 438–443, 1998.

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