Empirical Evidence for an Asymptotic Discontinuity in the Backbone of the 3-Coloring Phase Transition ⋆

Joseph Culberson[1] and Ian P. Gent[2]

[1] Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada, T6G 2H1. joe@cs.ualberta.ca

[2] School of Mathematical and Computational Science, University of St Andrews, St Andrews, Fife KY16 9SS, United Kingdom. ipg@dcs.st-and.ac.uk

Report APES-16-1999

September 1999

Abstract. By adding just one edge at a time, we can for a given permutation of edges determine exactly which edge causes a graph to become uncolorable. In addition, we track the development of frozen pairs, that is pairs which must be the same color under all three colorings of the graph. We call this measurement process the frozen development process. This set of pairs shows an explosive jump just at the threshold. By identifying pairs of vertices that are frozen, we create a collapsed graph, which has the same set of 3-colorings. Corresponding to the jump in frozen pairs, there is an exceedingly rapid collapse of this graph near the phase transition. It is this sudden freezing that causes the threshold. The evidence supports the contention that there is a sudden (discontinuous in the limit) change in the frozenness with respect to 3-coloring.
However, we also argue that if the measure were to be taken with respect to minimization of violated edges, then this sudden drop would mostly disappear. The evidence for this comes from measures showing large (on average) sets of edges such that removing any one of them from the threshold graph would result in a distinct set of 3-colorings. This could explain the discrepancy between the jump we observe and the lack thereof in certain statistical mechanical analyses of the 3-coloring phase transition.

Available from http://www.cs.strath.ac.uk/~apes/apesreports.html

⋆ Joseph Culberson's work was supported in part by an NSERC grant OGP8053 and an EPSRC visiting fellowship, GR/M54605. We thank members of APES for their help. We are members of the APES research group, www.apes.cs.strath.ac.uk.

1 Introduction

A phase transition has been identified for many NP-complete problems and is frequently correlated with a high frequency of hard instances. This contrasts with randomly chosen instances from other regions of the problem space, where most often such instances are easy. Recently the techniques of statistical mechanics has been applied to the analysis of this transition, and has yielded insights into the nature of the region and its relation to hardness. (See [3] for an overview and many references).

One of the more recent efforts has been the identification of the nature of the order parameter at the transition. In some problems (k-SAT, $k > 2$) there is evidence that this parameter is discontinuous [5, 7]. These problems typically have a high frequency of hard instances. On the other hand problems with a continuous parameter tend not to have a high frequency of hard instances. This has been explored most thoroughly in the case of $2 + P$-SAT [6, 7]. For the Hamiltonian Cycle problem, the order parameter may also be continuous [9] which corresponds to a very low frequency of hard instances at the transition [8].

For SAT one measure of the order parameter is the *backbone* [7] which is the number of variables that are frozen to a particular value under all satisfying assignments. For k-SAT, $k > 2$, the evidence indicates that this measure jumps from zero to a fixed fraction of the n variables at the transition. Similarly, for Hamiltonian Cycle we can use as variables the number of edges that must appear in all Hamiltonian Cycles of the graph. In contrast to the K-SAT case, only a few edges are frozen in satisfiable instances at the transition.

For k-coloring (we consider $k = 3$ almost exclusively in this document) it is more difficult to define and measure the order parameter, due in large part to the symmetry of the solutions. However, it is clear that if we add an edge to a colorable graph and it then becomes uncolorable, then in the graph without the edge the pair of vertices must have been colored with the same color in every coloring of the graph. Such a pair we refer to as frozen-out for reasons that will become apparent in the next section. We adopt the number of such pairs as our backbone, an approximation to the order parameter for coloring, as it mimics most closely the backbones used in other problems.

What we find is that this backbone shows strong evidence of a discontinuous jump at the transition. This appears to differ from first attempts at a theoretical analysis[9]. In the last section we attempt to provide a reason, based on further empirical analysis.

In a followup report we will show by further experiments that for 3-coloring there is a high frequency of (exponentially) hard instances at the transition, which would correspond to current hypotheses correlating a discontinuity to hard instances.

2 The collapse under 3-coloring

The frozen development process for detecting the threshold for coloring and taking measures of how frozen the coloring is works as follows:[3]

```
Generate a random sequence (permutation) S of all (n choose 2) pairs of vertices.
Let G = (V, E) be an empty graph on n vertices
For 1 ≤ i ≤ (n choose 2)
    Add edge S_i to E
    If G is not-k-colorable
        return G and i.
            # i is the threshold index for this S
            # G is the threshold graph for this S
        Exit
```

By taking a large sample of such sequences we can get an accurate estimate of the threshold.

However, we can gain significantly more information. Basically as each S_i is added, assuming the graph is still k-colorable, we scan forward and determine all those pairs which are colored the same under all colorings. That is, each remaining unfrozen pair (x, y) is tested and recorded as *frozen-out at index* $f_o(x, y) = i$ iff for every coloring c of G $c[x] = c[y]$. The name frozen-out means that this edge cannot be in any k-colorable graph containing all the edges up to index i in sequence S.

To determine whether a pair is frozen-out under all colorings, we temporarily add an edge and see if the graph $\mathrm{G} + \{x, y\}$ is still k-colorable. If not then the pair is frozen-out. After the test the edge is deleted and we move on to the next. When the threshold is found we may add to our statistics the number of edges frozen-out at each i.

From our empirical studies typically, just before the threshold, there is a sequence of maybe 1–5 edges which when added cause on average 16% or more of the pairs to be frozen-out; the typical big jump we might expect.

Similarly, we can measure for each i the number of pairs (x, y) such that for all colorings $c[x] \neq c[y]$.[4] Such a pair is said to be frozen-in. To test whether the pair (x, y) is frozen-in, we merge x to y. A merge requires y to be deleted and then for every edge $\{y, z\}$ that was in G, we add an edge $\{x, z\}$ to $\mathrm{G} \circ (xy)$

[3] Implementation detail: in fact we do a binary search for the threshold which means we only need to do $O(\log n)$ calls to the coloring program instead of $O(n^2)$. Other savings are possible for the rest of the frozen development process, we only present the conceptual framework here. See APES report 13 at http://www.cs.strath.ac.uk/~apes/apesreports.html for a more complete description.

[4] Trivially, the pair S_i is one pair frozen-in since it has just been added as an edge. However, we want to keep edges separate, so the number we compute depends only on pairs with index greater than i.

if it is not already present. If G $\circ$ (xy) is not k-colorable then the pair (x, y) is frozen-in.

Note that frozen-in is not quite the complement of frozen-out. When a pair is frozen-out it cannot appear in a k-colorable graph which is a super graph of the current one. When a pair is frozen-in it means we are free to add it to any k-colorable super-graph of the current graph. Its addition will make no difference to the set of legal colorings of any k-colorable super-graph. Thus, we can skip adding these edges, because they simply do not matter to the set of colorings.[5]

In practice we do not want to stop measurement at the threshold index; after all if some other edge had occurred at this point in the sequence we might be able to keep on going.

In the *full frozen development method* when a pair S_i is added as an edge that renders the graph non-k-colorable, then the edge is deleted and we move on to the next edge. This gives us a way to smoothly extend frozen measures beyond the threshold until all pairs are either an edge, frozen-out, or frozen-in. At that point, the graph is a maximal k-colorable graph.

In figure 1 we show how the number of frozen-out pairs grows as the edge density increases. This is clearly typical of a phase transition, and the sharpness suggests that there will be a discontinuity at $n = \infty$. The range over which these curves are plotted is the entire range over which any difference in the number of frozen-out or frozen-in pairs occurred. For the larger values of n it is of course a tiny fraction of the set of $\binom{n}{2}$ pairs. Bollobás em et al[1] showed that

$$\lim_{n\to\infty} \Pr \text{ (Min Degree} \geq 2) = \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-e^{-2c}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty \end{cases}$$

where $m = n/2(\ln n + \ln\ln n + c_n)$.

For 3-coloring, clearly a pair cannot be frozen-out unless each vertex in the pair has degree at least two. In figure 2 we plot the index of the last edge inducing a frozen pair against the value predicted by the Bollobás *et al* formula and against the natural log of n.

One potential objection to using the number of frozen pairs as a measure of the big jump is due to transitivity. Let Y be the set of vertices which appear in a pair with y which is frozen-out. We define X similarly with respect to vertex x and assume that in the current G the pair (x, y) is not yet frozen. As a consequence X and Y form two equivalence classes with respect to frozenness. Over all vertices, frozen-out defines an equivalence relation that partitions the vertices into classes. Each class is a set of vertices that are all forced to take the same color under every coloring of the graph G. Because coloring is a monotonic property, this partition is a refinement of every partition induced by a k-colorable super-graph of G. *In other words, these sets are independent sets that must remain independent sets in every k-colorable super-graph.* Monotonic

[5] When we test the hardness of coloring, the subject of another paper, all edges must be added. The frozen-in edges could make a big difference to a coloring algorithm, for example by preventing errors high in the search tree.

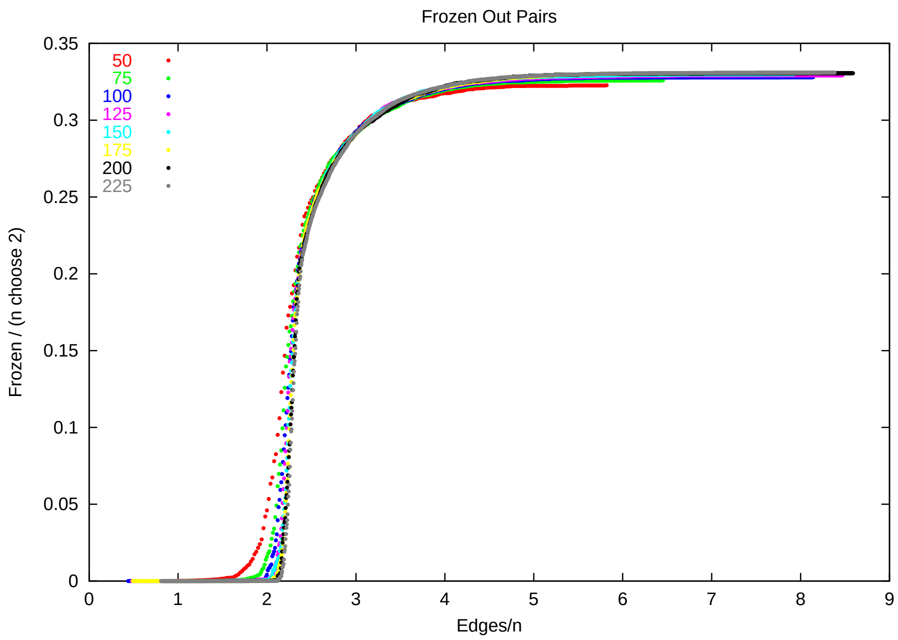

Fig. 1. The Number of frozen-out pairs as a ratio to $\binom{n}{2}$ plotted against the ratio m/n. This is for $k = 3$.

here simply means that if a graph G is k-colorable, then every subgraph of G is also k-colorable. Note that "not-k-colorable" is also monotonic (in the opposite sense).

We call the partition induced by frozen-out the *forced equivalence relation of* G, meaning it indicates the sets of vertices of G that must be colored the same in all k-colorings.

Suppose the next edge is added and the pair (x, y) becomes frozen-out. This means that the classes X and Y must merge, and thus we get $|X||Y|$ new frozen-out pairs as a result. So the question that could be asked is, "is the big jump due to two large classes merging, or the merging of several smaller ones?" A little analysis indicates that the latter must also be happening, but there is another clearer analysis that we investigate next.

After frozen-out and frozen-in are computed for the full frozen development of a sequence, it is an easy matter to start with an empty graph and compute a *collapsed graph* for each index i except when i is frozen-out. To compute the collapsed graph at i, add all non-frozen-out edges up to and including the ith. Then scan forward over the remaining pairs. For every forward pair that is frozen-out perform the merge. For every forward pair that is frozen-in add the edge. Due to transitivity, most of these forward operations will be redundant

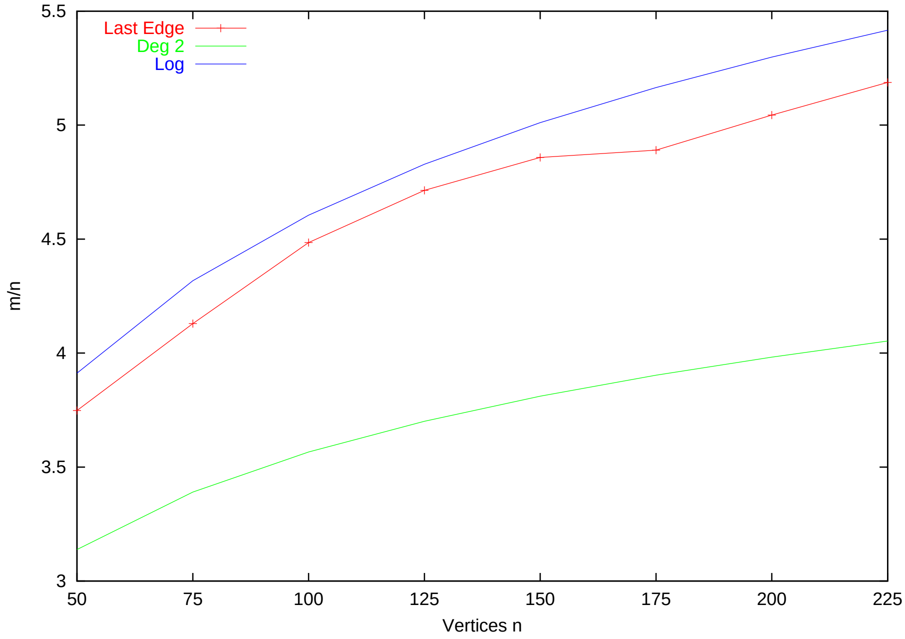

Fig. 2. The index of the last edge inducing a frozen pair compared to the theoretical limit for minimum degree two, and against the natural log. The latter seems to be the best approximation.

and will require no actual work to be done.

Now the vertices in the collapsed graph at index i are class representatives of the elements in the forced equivalence relation. Simply counting the number of vertices in the collapsed graph gives us the number of forced classes. Clearly, for a graph with no edges this number is n, the number in the original graph. For a maximal k-colorable graph, the collapsed graph is a k-clique. So the number of vertices forms a non-increasing sequence from n down to k.

By definition, for every pair of vertices in the collapsed graph not joined by an edge there is a k-coloring of the graph that makes the vertices the same color, and another coloring that colors the pair differently. That is, there are no frozen-out pairs in the collapsed graph. This suggests that the study of the structure of this graph could be of interest.

For this note, we only study the sequence of the number of vertices in the collapsed graph for 3-coloring examples. As one might expect, there is indeed a sudden collapse near the threshold in each sequence. However, if we compute the average number of vertices in the collapsed graph over the sample set, then the drop looks a lot smoother. The average is not a good indicator of what is happening in the individual instances, at least at the scale at which we are examining the data.

We ran 100 samples on at each value of $n = 50, \ldots, 225$ in steps of 25. First we show two sets of samples of 15 graphs, each from the $n = 200$ set, graphs 1-15 and 21-35.

The y-axis represents the number of vertices remaining in the collapsed graph. The x-labels indicate that the x-axis is the sequence index, that is the number of edges that would be added if none were skipped. We see that typically each instance drops rather rapidly somewhere in the range 430 to 470, which when divided by 200 gives a ratio of 2.15 to 2.35 as we might expect. The threshold graph will typically occur very shortly after the big drop given that there are so many frozen-out pairs lying around after the big drop (the threshold pair must be a frozen-out pair).

The marked points on the curves represent the *effective* edges, that is those edges that were not frozen in the sequence when encountered and so were actually added to the graph. We see that only a small number of edges are actually required to cause the catastrophic drop, that is even smaller than the range of sequence indices might indicate.

More to the point, in figure 4 we plot the average number of vertices remaining in the collapsed graph for each index, together with 20 graphs starting at sample 41 ($n = 200$). The average is taken over all 100 samples. Almost all instances exhibit a very narrow range over which they drop from a fairly well defined region at the top to another at the bottom. And all have at least one large drop caused by a single edge. The average really only reflects the percentage of instances that have dropped so far, not how fast they drop.

So how fast do the instances drop? We find for each experiment (i.e. one run on a sequence) the one edge that caused the maximum drop in the number of vertices on that sequence. Call the number dropped by this edge the *maxdrop* of the experiment.

In figure 5 we see that the average maxdrop is from 72% to 44%, a drop of 28% of the vertices on a single edge. This is represented by a near-vertical jump in the average curve.

Under the assumption of a threshold for 3-colorabillity, we expect the interval over which the drops occur to converge (wrt n). In figure 6 we plot the index (as a ratio to n) where the collapse first has fewer than 80% of its original vertices. In contrast, in figure 7 we plot the actual (top) and relative number of edges required to cause the graph to drop from $0.9n$ down to $0.4n$ vertices. This 50% interval seems not only to be converging relative to n but even in absolute terms. In fact, there are specific instances where an edge caused the drop from above 90% to below 40%, causing the measured number of edges to be zero.

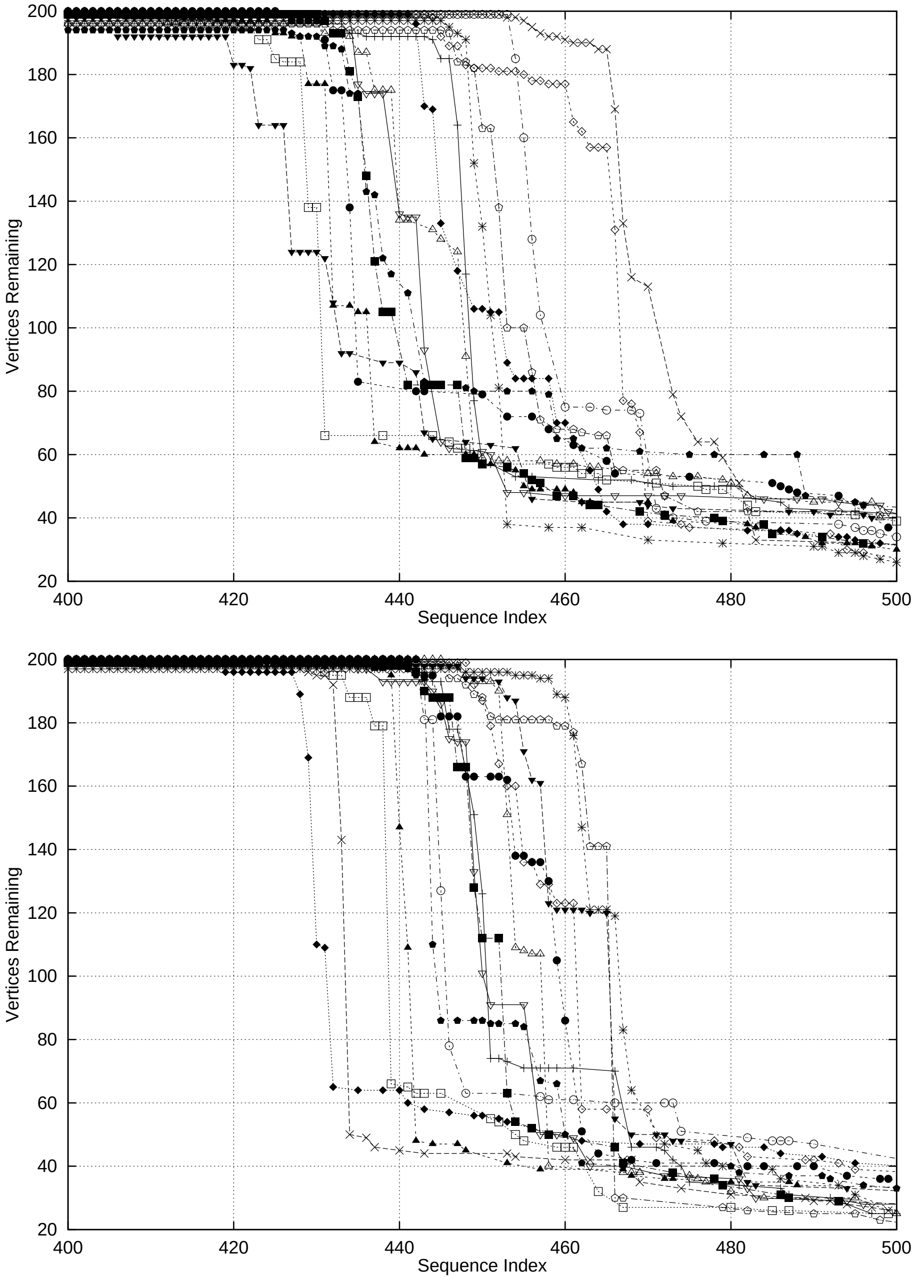

Fig. 3. Typical collapse of 200 vertex instances (20 each). The y-axis represents the number of vertices remaining in the collapsed graph. The x-labels indicate that the x-axis is the sequence index, that is the number of edges that would be added if none were skipped.

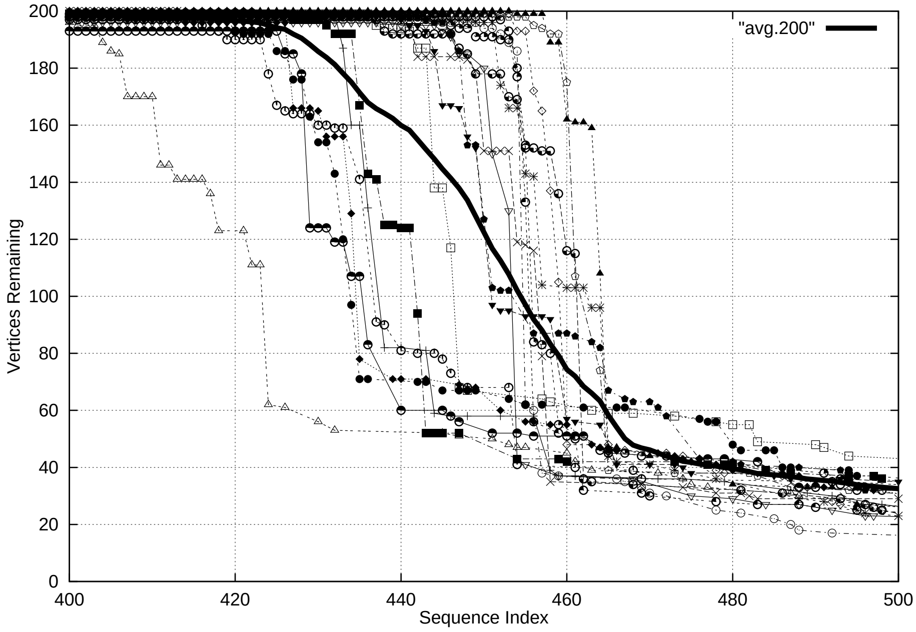

Fig. 4. Another set from $n = 200$ with the average of the entire 100 samples overlaid. Notice that the average does not closely approximate the shape of any of the curves, even the most atypical case at the far left.

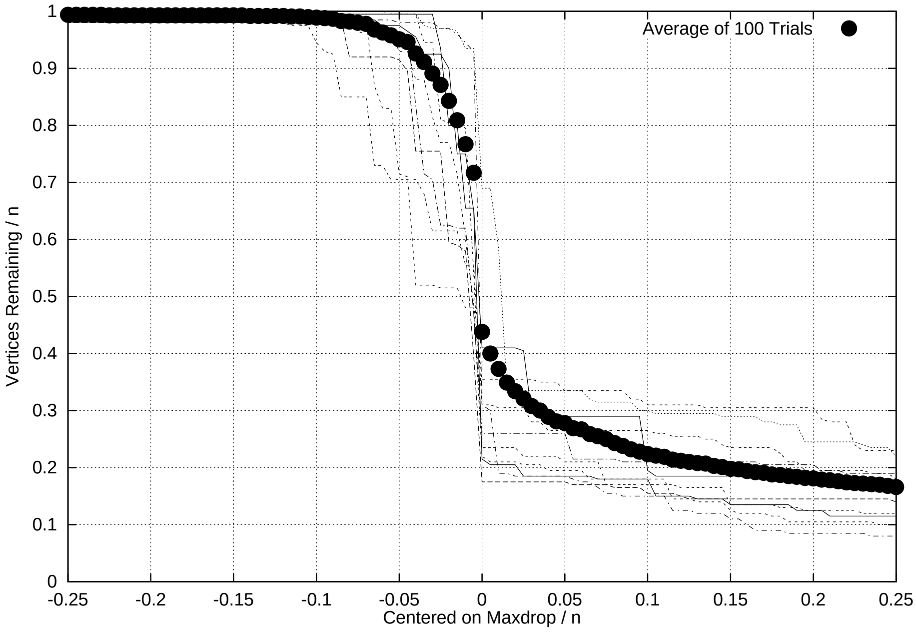

Fig. 5. The first ten experiments from figure 3 and the average of all 100 experiments, but with the samples and average aligned at the index of the maxdrop edge.

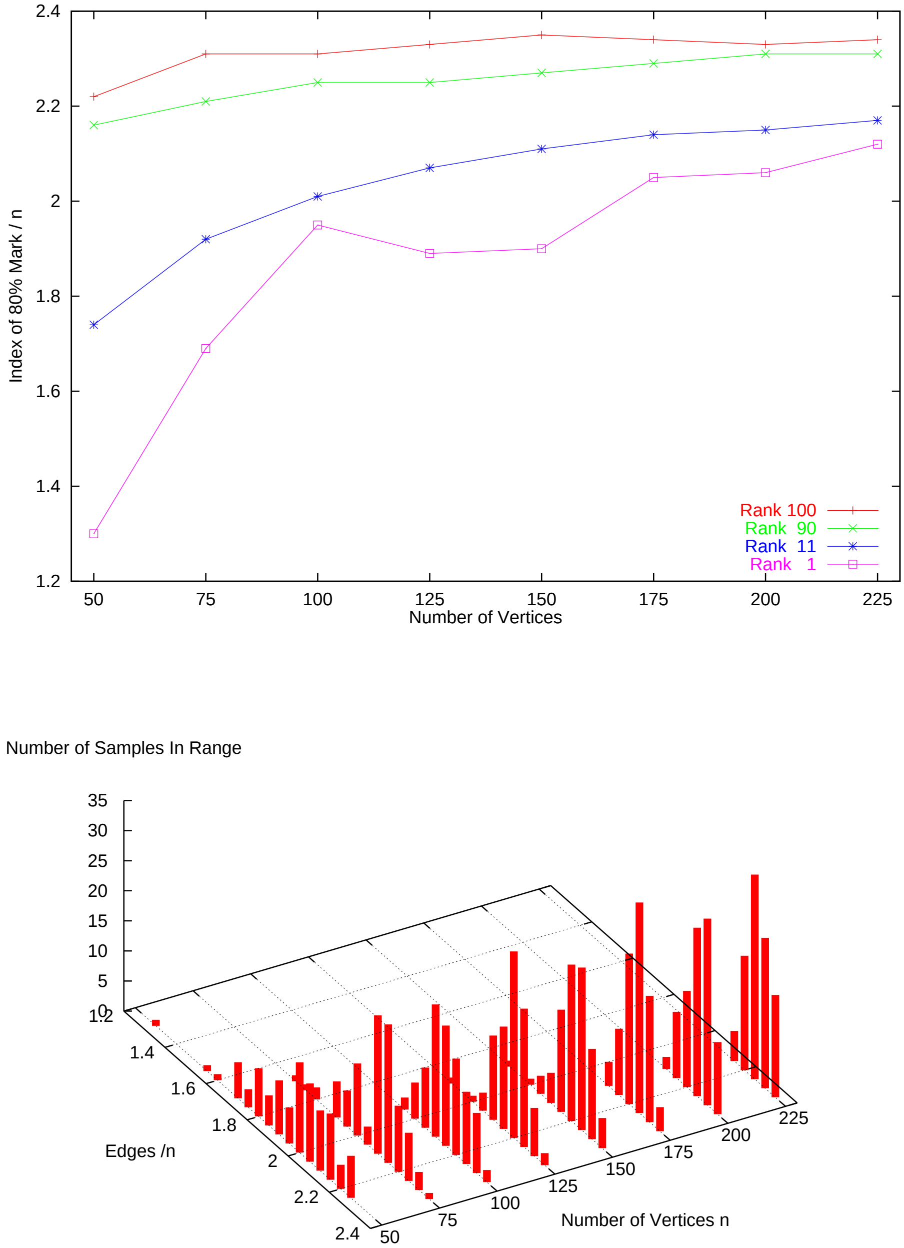

Fig. 6. In the top part of this figure we plot the index (as a ratio to n) where the collapse first has fewer than 80% of its original vertices. The top and bottom lines correspond to the full range over the 100 samples while the center lines bound the 80% of the samples in the middle, a simple minded attempt to remove the high variance extremes. In the bottom, histograms show the distribution of samples with respect to the first index to drop below the 80% point. Note the convergence apparent here, and also note that in no instance did the first drop below 80% exceed $2.35n$ edges.

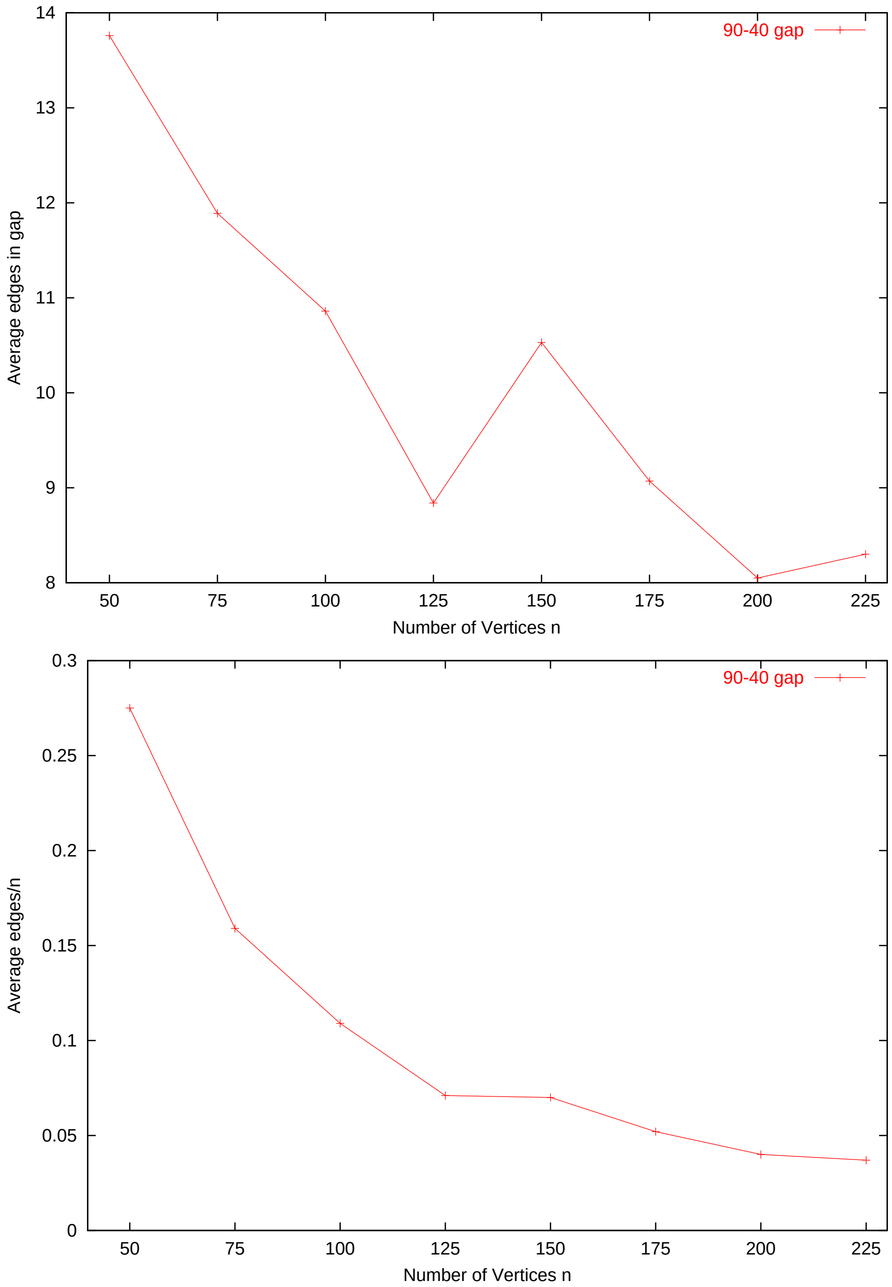

Fig. 7. The average over 100 samples of the number of edges in the gap between the first drop below 90% and the first drop below 40%. In the top the actual number of edges seems to be decreasing, and the bottom shows the ratio of this number to n.

In figure 8 we plot in the form of error bars for each n the average, minimum and maximum maxdrop of the set of samples. This shows that the average of $\approx 28\%$ is very consistent, and the remarkable thing is that even the minimum is consistent, with the smallest fraction at 10% for $n = 50$. That is, in every one of the 800 trials, there was a single edge that caused a fractional drop of at least 10%.

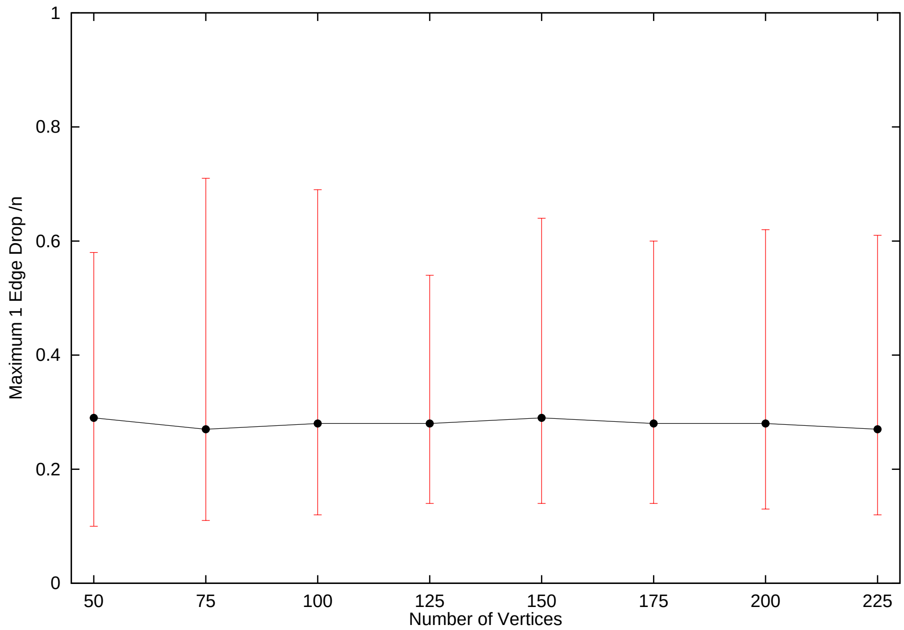

Fig. 8. The average maximum drop on a single edge. The error bars are in fact the minimum maximum drop and the maximum maximum drop over 100 samples.

It is intuitively surprising that a single edge would cause a fixed fraction of vertices to collapse. However, so far the evidence is consistent with this hypothesis. Perhaps for significantly larger n this fraction would be reduced.

3 Measuring the minimum violation and the potential loss of collapse

In graph k-coloring we have defined a threshold with respect to the frozen development process as the edge which when added causes the graph to become non-k-colorable. Let us, for reasons to become clear shortly, refer to this as the

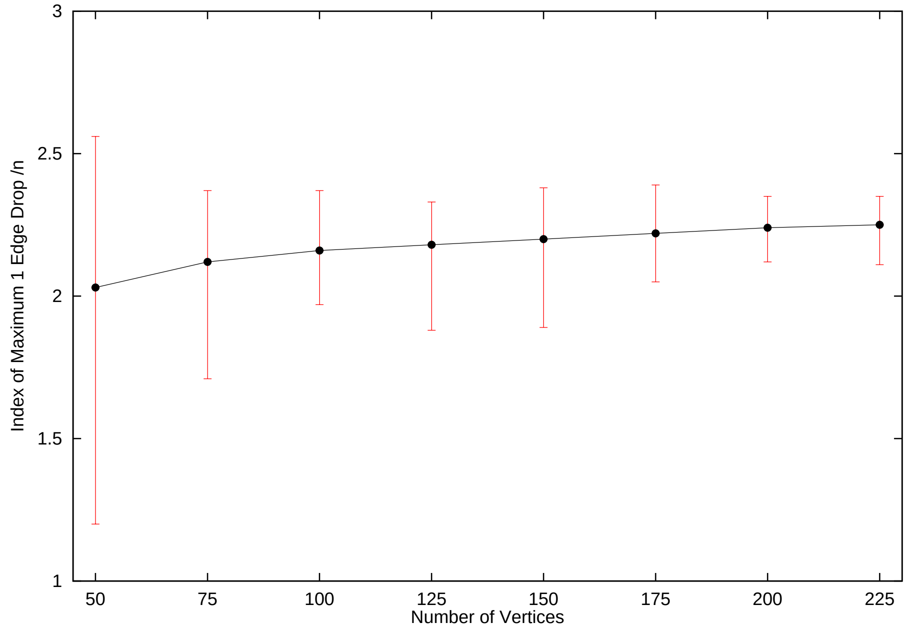

Fig. 9. The average index of the maximum drop edge. As expected the average index of the maxdrop edge appears to be converging towards 2.3. The error bars indicate minimum and maximum index over the entire set.

t_0^k threshold, where the superscript k refers to k-coloring. For the rest of this paper we will assume $k = 3$ and drop the superscript.

The subscript 0 refers to the fact that we will 3-color the graph with no edge violations. When statistical mechanical models are used to study the phase transition typically the order parameter is taken with respect to the minimization of violated edges (corresponding to clauses in SAT)[5] .

In previous work [2] and in this paper so far, we have considered an approximation to the order parameter by measuring the number of frozen pairs; in particular the number of pairs of vertices that are frozen-out (i.e. must be colored the same) under all colorings of the graph and the number of pairs frozen-in. In this note we suggest that we can also take this measure for a sequence of thresholds, $t_0, t_1, \ldots$.

To define this, first we define the (k, v) frozen development process. We begin with a vertex set $V = \{1, 2, \ldots, n\}$. We generate a random permutation of the $\binom{n}{2}$ vertex pairs of vertices $S =< S_1, \ldots S_{\binom{n}{2}} >=< (x_1, y_1), \ldots (x_{\binom{n}{2}}, y_{\binom{n}{2}}) >$.

We define the graph G_m (with respect to S) as the n-vertex graph $G_m = (V, E_m)$ where $E_m = \{S_1, \ldots S_m\}$. ($G_0$ is the empty graph, consisting of V and an empty set of edges.) We say that a graph is (k, v)-colorable if it can be

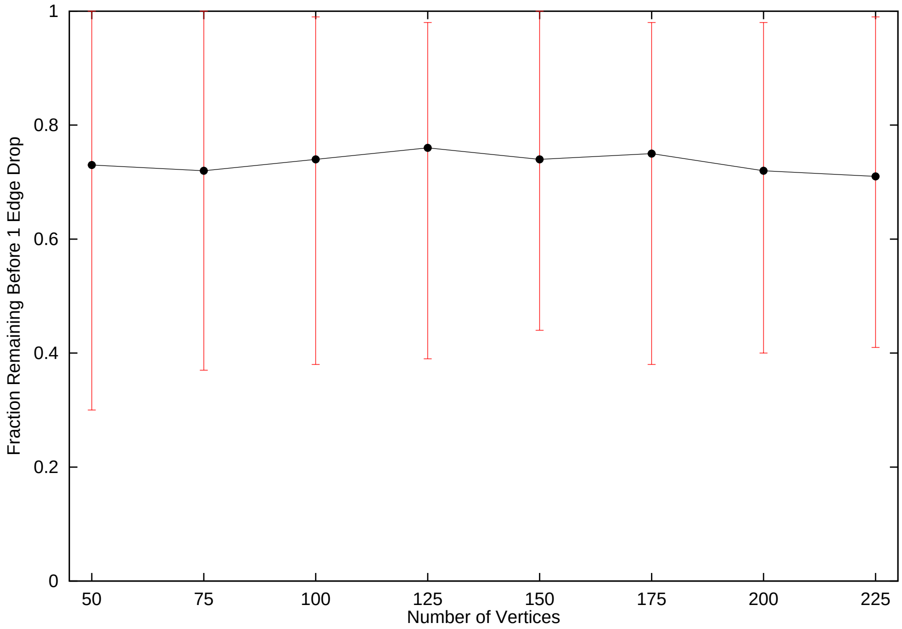

Fig. 10. The minimum, average and maximum of the number of vertices in the graph at the top of the maximum drop. The average here too seems fairly consistent with the 72% observed before for $n = 200$. Observe that the drop on one edge is almost equal to the total drop up to that point on average; i.e. 100-72=28. However, the range can extend anywhere from under 40% to 100%.

k-colored by violating at most v edges[6].

We say that a pair of vertices $S_h = (x_h, y_h)$ *with index* $h > m$ *is* k, v*-frozen-out at* m *under* (k, v)*-colorings if*

1. G_m *is* (k, v)*-colorable,*
   AND
2. *for every* (k, v)*-coloring* c *of* G_m *we have* $c(x_h) = c(y_h)$.

Note that if we set $v = 0$ then this is exactly the same frozen-out we have used previously. Also, note that if a pair is frozen-out at m, then it is frozen for every (k, v)-colorable graph $G_{m'}$ in the sequence where $m' > m$. We say the *index of freezing* of a pair (x_h, y_h) is m if m is the minimum value for which the pair is frozen.

Now we are ready to define the multiple thresholds. We define the (k, v)-threshold as

$$t_v^k = t_v^k(S) = \min\{m : S_m \text{ is frozen the same at } m - 1\} \tag{1}$$

[6] An edge is violated by a coloring c if both end points receive the same color.

We can then define the T_v^k threshold (the average) as

$$T_v^k = \frac{\sum_{S \in \Pi} t_v^k(S)}{\binom{n}{2}!} \tag{2}$$

where Π is the set of all permutations of the set of vertex pairs.

Although in principle we can now use our frozen development process, the cost of doing so will be quite high. One measure we do have for $n \leq 200$ is the size of the threshold set. The *threshold set* is the set of edges in the uncolorable threshold graph such that removing any one of them will make the graph colorable. In other words, it is the set of violated edges in the set of $(3,1)$-colorings of the t_0^3 threshold graph.

Note that on the index before the threshold graph, that is the last 3-colorable graph, any of the approximately $2.3n$ edges may be violated in a $(3,1)$-coloring. However, at the next edge every $(3,1)$-coloring must violate exactly one of the edges in the threshold set.

The data in table 1 show that this is immediately reduced to approximately $0.35n$ on average.

Number of Threshold Edges					
N	Avg	Std	Avg/n	Min	Max
50	16.80	11.62	0.34	1	49
75	26.96	18.68	0.36	1	83
100	31.85	21.12	0.32	1	85
125	41.50	29.92	0.33	3	160
150	51.18	34.63	0.34	5	140
175	70.40	49.41	0.40	1	195
200	69.73	52.73	0.35	2	205

Table 1. The size of threshold sets. We do not have them for 225 yet.

However, an average of $0.35n$ is nevertheless a large increase in the freedom of 3-coloring. That is because each of the edges in the threshold set gives a distinct set of 3-colorings of the graph when it is the one edge violated, since the pair of vertices receive the same color only in those colorings.

We have not yet taken the measurements, but we expect that the set of frozen pairs will be greatly reduced. This is because each separate violation will result in a different set of pairs being frozen the same.

Finally in figure 11 we show the distribution of the size of threshold sets for $n = 200$ over the 100 samples.

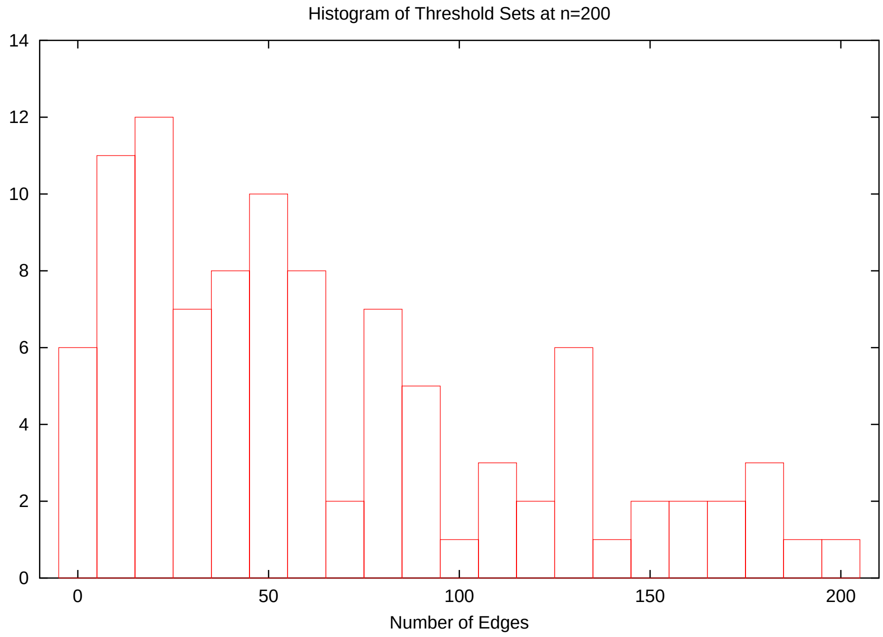

Fig. 11. The distribution of threshold sets at $n = 200$, 100 samples.

4 Conclusions and Future Results

We have defined a backbone for 3-coloring based on pairs of vertices that must receive the same color under all colorings and demonstrated empirically beyond reasonable doubt that this measure exhibits a first order discontinuity at the 3-coloring threshold. Since this measure is based on elements (pairs) which are quadratic in the number of vertices, and since coloring is really a partitioning problem, we also converted this measure to one of counting the number of e-quivalence classes forced by the set of three colorings. This results in a collapse of the graph, which shows a corresponding sharp drop in size.

We then pointed out that if we were taking this measure with respect to a minimum edge violation coloring approximation, this sharp (discontinuous) change might be reduced or eliminated. The empirical evidence supporting this is based on the threshold sets, that is the set of edges such that the removal of any one of them would make the uncolorable graph colorable. At the threshold, this set is on average large, possibly $0.35n$ or larger. This means that allowing one edge to be violated might cause the t_0 threshold to exhibit few or no frozen pairs.

To verify this last conjecture we need to modify our programs, and likely make some efficiency improvements, as the number of colorings needed could be significant.

In a follow on report, nearly complete, we will show that there is a widening range $(\alpha m/n, \alpha' m/n)$ over which the best programs available, including a conversion to SAT, show exponentially (in n) increasing median time. This median growth rate is $\approx 2^{n/25}$ over the range examined. This result agrees with prior conjectures, and is supported by theoretical results on hardness of coloring[4].

References

1. B. Bollobás, T. I. Fenner, and A. M. Frieze. An algorithm for finding Hamilton paths and cycles in random graphs. *Combinatorica*, 7(4):327–341, 1987.
2. J. Culberson and I.P. Gent. Well out of reach: why hard problems are hard. Technical Report APES-13-1999, APES Research Group, 1999. Available from http://apes.cs.strath.ac.uk/apesreports.html.
3. Brian Hayes. Can't get no satisfaction. *American Scientist*, 85:108–112, March 1997. http://www.amsci.org/amsci/issues/Comsci97/compsci9703.html.
4. Colin McDiarmid. Determining the chromatic number of a graph. *SIAM Journal on Computing*, 8(1):1–14, February 1979.
5. R. Monasson and R. Zecchina. Statistical mechanics of the random k-sat model. *Phys. Rev. E*, 56, 1997.
6. R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyansky. Phase transition and search cost in the 2+p-sat problem. In T. Toffoli, M. Biafore, and J. Leao, editors, *Proceedings of PhysComp 96*, Boston, 1996. http://www.lpt.ens.fr/ monasson/ http://www.ictp.trieste.it/ zecchina.
7. R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L Troyansky. Determining computational complexity from characteristic 'phase transitions'. *Nature*, 400:133–137, 1998. http://www.lpt.ens.fr/ monasson/ http://www.ictp.trieste.it/ z ecchina.
8. Basil Vandegriend and Joseph Culberson. The G_{nm} phase transition is not hard for the Hamiltonian cycle problem. *Journal of Artificial Intelligence Research*, 9:219–245, 1998. www.jair.org/contents/v9.html.
9. R. Zecchina. personal communication.