SOME NEW APPLICATIONS OF PROBABILITY METHODS TO COMBINATORIAL ANALYSIS AND GRAPH THEORY

P. Erdös

Probability methods have often been applied successfully to solve various combinatorial problems which in some cases have been (and still are) unassailable by other methods. A systematic discussion of this method can be found in my recent book with J. Spencer.

In the present paper I give some new results obtained by this method. I do not give detailed proofs, but by following the very sketchy outlined instructions, they can easily be supplied by the interested reader who is familiar with the method.

P. Erdos and J. Spencer, Probabilistic Methods in Combinatorics, Acad. Press and Publishing House Hung. Acad. Sci. 1974.

1. Let $G_r(n)$ be a uniform hypergraph of n vertices (i.e. the basic elements of our hypergraph are vertices and r-tuples, for $r = 2$ we obtain the ordinary graphs, see e.g. the excellent recent book of C. Berge, Graphs and Hypergraphs, North Holland and Amer Elsevier 1973).

$K(G)$ will denote the chromatic number of G. Let $f_r(m,k;n)$ be the largest integer for which there is a $G_r(n)$ with $K(G_r(n)) = f_r(m,k;n)$ and such that all m point subgraphs of our $G_r(n)$ have chromatic number not exceeding k. I proved

$$(1) \qquad f_2(m,3;n) > c_1\left(\frac{n}{m}\right)^{1/3}\left(\log \frac{2n}{m}\right)^{-1} .$$

($c_1, c_2, \ldots$ denote absolute constants). (1) has the surprising consequence that for every l there is an $\varepsilon = \varepsilon(l) > 0$ and a graph

PROC. 5TH S-E CONF. COMBINATORICS, GRAPH THEORY, AND COMPUTING, pp. 39-51.

$G_2(n)$ satisfying $K(G_2(n)) \geq l$, yet every subgraph of $[\epsilon n]$ vertices has chromatic number at most three. I have no nontrivial upper bound for $f_2(m;3,n)$ valid for all m. Perhaps

$$f_2(m;3,n) < c_2\left(\frac{n}{m}\right)^{1-\alpha}$$

holds for some α and all m, but I have not even been able to prove that there is an $A(x)$ which tends to infinity as $x \to \infty$ and for which

$$f_2(m,3;n) < c_3 \frac{n}{m}\left(A\left(\frac{n}{m}\right)\right)^{-1}$$

$f_2(m,3;n) \leq 3\frac{n}{m}$ is, of course, trivial but it is not clear how much it can be improved (the case $m = \epsilon n$, $n \to \infty$ seems difficult).

Now we investigate whether (1) can be improved for all m. It is very easy to see that the exponent $\frac{1}{3}$ can not be replaced by a number greater than $\frac{2}{3}$. If $m = 4$ the condition that every subgraph of 4 vertices has chromatic number ≤ 3 simply means that $G_2(n)$ contains no $K(4)$ (a complete graph of 4 vertices) and by a theorem of Szekeres and myself this implies $K(G_2(n)) < cn^{2/3}$.

Results of Faudree, Rousseau, Schelp and myself which are not yet quite complete will, I feel sure imply that (1) can not hold for all m with an exponent greater than $\frac{1}{2}$. At present I can only show that it can not hold for $\frac{2}{3} - \epsilon$. It would be nice if the following would be true: For every $\epsilon > 0$ and $n > n_0(\epsilon)$

$$f_4(m,3;n) > c_4\left(\frac{n}{m}\right)^{\frac{1}{2}-\epsilon}$$

Following G. Dirac we say that the graph G is (vertex) critical if the omission of any vertex decreases its chromatic number. My proof of (1) was based on the obvious fact that every 4-chromatic critical graph has all its vertices of valency ≥ 3, thus if it has n vertices it has at least $\frac{3m}{2}$ edges. Gallai proved that a 4-chromatic critical graph of n vertices has at least $(\frac{3}{2}+\frac{1}{26})\,n$ edges but Dirac showed that this can not be improved to $\frac{3n}{2}+\frac{n-4}{6}<\frac{5n}{3}$. My proof of (1) actually gives that, if every 4-chromatic critical graph of n vertices has at least cn edges, then the exponent $\frac{1}{3}$ can be replaced by $\frac{c-1}{c}$. Thus, by Dirac's result, my method will certainly not give a better exponent than $\frac{2}{5}$. Gallai's result gives (1) with an exponent $\frac{7}{20}$ instead of $\frac{1}{3}$.

The situation is somewhat confused for $f(m,2;n)$ and this, as I now explain, is my fault. Gallai constructed a 4-chromatic graph $G(n)$ the shortest odd circuit of which has length $> [\sqrt{n}]$. In other words, Gallai proved $f_2([\sqrt{n}];2,n) > 3$. Perhaps this result is best possible. I stated that I can prove that, for a sufficiently large constant c,

$$(4) \qquad f_2([c\sqrt{n}],2;n) = 3.$$

Unfortunately I have not been able to reconstruct my 'proof' of (4) and perhaps it was incorrect. In any case (4) has to be considered open at the moment. Gallai and I conjectured that there are constants $c_1^{(l)}$ and $c_2^{(l)}$ so that

$$(5) \qquad f_2([c_1^{(l)}n^{1/l}],2;n) = l+1 \quad \text{but } f_2([c_2 n^{1/l}],2;n) > l+1.$$

It may be that the constants $c_1^{(l)}$ and $c_2^{(l)}$ are independent of l.

We made no progress with the proof of (5) and its proof or disproof may not be easy. On the other hand, I proved that there are constants $c_1^{(l)}$ and $c_2^{(l)}$ so that every $G(n)$ with $K(G(n)) > l$ contains a circuit having fewer than $c_1^{(l)} \log n$ edges, and yet there is a $G(n)$ with $K(G(n)) > l$ all whose circuits have at least $c_2^{(l)} \log n$ edges.

Observe that $f_2(m,2;n)$ behaves very differently from $f_2(m,k;n)$ for $k \geq 3$. This is perhaps not surprising since the critical 3-chromatic graphs are trivial (they are the odd circuits) whereas the critical 4-chromatic graphs can be very complicated. For $r > 2$ the 3-chromatic graphs already have a non-trivial structure (probably too complicated for complete characterisation). Actually for $r > 2$ $f_r(m,2;n)$ does not behave very differently from $f_r(m,k;n)$. The proof used for (1) gives in this case

$$(6) \qquad f_r(m,2;n) > c\left(\frac{n}{m}\right)^{r-2/r-1} .$$

For the proof of (6) one needs in addition the following result of Lovasz and Woodall (Which did not exist when) proved (1)): Every critical 3-chromatic r-graph ($r > 2$) of n vertices contains at least n edges. Woodall also proved that this result is best possible i.e. there are such r-graphs with exactly n edges.

P. Erdős, On circuits and subgraphs of chromatic graphs, Mathematika 9(1962), 170-175.

P. Erdos and G. Szekeres, A combinatorial problem in geometry, Comp. Math. 2(1935), 463-470.

T. Gallai, Kritische Graphen, Publ. Math. Inst. Hung. Acad. Sci. 8(1963), 165-192 Sec in particular pp 172-173 and 186-189.

G. Dirac, A theorem of R.L. Brooks and a conjecture of Hadiriger Proc. London Math. Soc. (3) 7(1957) 161-195.

L. Lovasz Graphs and set systems, Beitrage zur Graphentheorie Conference held at Ilmenau 1968 H Sachs editor) Teubner Leipzig 1968 99-106, D.R. Woodall, Property B and the four colour problem Combinatorics, Proc. Conference Oxford 1972, Institute of Math. and Applications, 322-340.

2. Uniquely colourable graphs. This work was done jointly with Ehud Artzy. A k-chromatic graph is called uniquely colourable if the k independent sets into which our graph can be decomposed are unique. Trivial examples of uniquely colourable graphs are the complete graphs less trivial examples were known but they were relatively few in number. In particularly in was not known if for every k and l there is a uniquely colourable k-chromatic graph of girth l. During my last visit to Israel I discussed these problems with E. Artzy and we noticed that the methods used in my paper quoted below, apply here too. A previous problem of Sauer should be mentioned here which was very helpful. Sauer asked (oral communication) several years ago if for $n > n_0$ there exists a 3-chromatic graph not containing a triangle and having n vertices of each color so that it should not contain an independent set of more than n points. I observed that the probability method relatively easily gives the existence of such graphs and we observed that the same method gives a k-chromatic uniquely colorable graph of arbitrarily large girth.

I give a brief outline of the construction. Let k and l be given, $\varepsilon = \varepsilon(k,l)$ sufficiently small, $n > n_0(\varepsilon,k,l)$ large. Consider all graphs of kn vertices, with k color classes $S_1, \dots S_k$, each S_i has n vertices. S_i and S_j are joined at random by $[n^{1+\varepsilon}]$ edges. The number of these graphs clearly equals

$$\binom{n^2}{[n^{1+\varepsilon}]}^n = A_n .$$

A simple computation (involving only the first moment) shows that, with the exception of $o(A_n)$ of them, these graphs have $o(n)$ circuits of length $\le l$. From each of these circuits we omit an arbitrary edge. The resulting graphs clearly have girth $> l$ and a simple (but slightly more complicated) computation gives that all but $o(A_n)$ of them have no independent sets of size n other than the S_i $(i = 1, \dots k)$. Thus these are k-chromatic uniquely colorable graphs of girth l.

Some further refinements are possible. This method gives that there are absolute constants c_1 and c_2 so that for every l and n there is a uniquely colorable graph $G(n)$ of n vertices, girth l and chromatic number $> c_1 n^{c_2/l}$.

We can further show that all but $o(A_n)$ of our graphs have the property that no set U of its vertices is independent which intersects two S_i's insets of size $> n^{1-\eta}$ where $\eta = \eta(k, l)$ is sufficiently small.

At the recent conference in graph theory in Prague (June 24 - 28, 1974) Müller gave uniquely colorable k-chromatic graphs of girth l for every k and l by a direct construction without using probability arguments.

Chung. C. Wang and Ehud Artzy, Note on uniquely colorable graphs, J. Comb. Theory ser. B 15(1973), 204-206.

P. Erdös, Graph theory and probability I and II Canad. J. Math 11(1959), 34-38 and 13(1961), 346-352.

3. Answering a question of Berman I showed that for sufficiently small $\varepsilon > 0$ and $n > n_0(\varepsilon)$ one can direct the edges of a $K(n,n)$ in such a way that every subgraph of $[n^{2-\varepsilon}]$ edges contains a cyclically directed C_4. The proof is straightforward. One can show that if one directs the n^2 edges of our $K(n,n)$ at random all but $\sigma(2^{n^2})$ of these graphs will have the required property. I have not been able to determine the best possible value of ε. I am quite sure that for every $\eta > 0$ and $n > n_0(\eta)$ one can direct the edges of a $K(n,n)$ in such a way that every subgraph of $[n^{3/2+\eta}]$ edges contains a cyclically directed C_4. Perhaps this already holds for subgraphs of $cn^{3/2}$ edges if c is a sufficiently large absolute constant. Similar results will hold for cyclically directed C_{2r}'s - perhaps for every subgraph of size $c_r n^{1+1/r}$.

The question of Berman led me to the following problem which is of independent interest. Let $G(k)$ be a directed graph of k vertices. $A(n;G(k))$ is the largest integer so that there is a graph $G(n)$ (of n vertices) which can be directed in $A(n;G(k))$ ways so that none should contain a subgraph isomorphic to $G(k)$ (isomorphism here of course means that all the edges are directed as in $G(k)$). Determine or estimate $A(n,G(k))$.

Perhaps the following modification of this problem is more interesting and useful. Let now $G(k)$ be an undirected graph of k vertices and let $f(n;G(k))$ be the largest integer so that there is a $G(n;f(n;G(k))$ [$G(n;l)$ is a graph of n vertices and l edges] which does not contain our $G(k)$ as a subgraph. The function $f(n;G(k))$ is of course well known from the theory of extremal graphs, e.g. (Turan in his pioneering paper determined $f(n,K(k))$. Denote now by $F(n;r,G(k))$ the largest integer for which there is a $G(n)$ whose edges can be coloured by r colors in $F(n,r,G(k))$ ways so that there should not be a monochromatic $G(k)$. Clearly

$$(1) \qquad F(n;r,G(k)) \geq r^{f(n;G(k))}$$

To see (1) observe that by definition of $f(n;G(k))$ there is a $G(n;f(n;G(k))$ which does not contain our $G(k)$ and hence its edges can be coloured arbitrarily. It seems that for very many (perhaps all) graphs $G(k)$, $F(n;r,G(k))$ is not much bigger than $r^{f(n;G(k))}$. In particular Rothschild and I conjectured that for $n > n_0$

$$(2) \qquad F_2(n;C_3) = 2^{[n^2/4]}$$

and more generally for every s and $n > n_0(s)$

$$(3) \qquad F\ (n;K(s)) = 2^{f(n;K(s))}$$

where $f(n;K_s)$ is the number of edges of the well known Turan-graph i.e. the largest graph on n vertices which does not contain a $K(s)$.

(2) clearly does not hold for all n. It is possible that

$$F(n;r,G(k)) < r^{(1+\varepsilon)f(n;G(k))}$$

will hold for all (or "nearly" all) graphs $G(k)$. Clearly analogous problems can be stated for hypergraphs.

4. **Professor Even and others considered the following question:** A graph $G(n)$ is called a rigid circuit graph if every circuit of it contains at least one diagonal. They obtained various algorithms for determining a rigid circuit graph $G(n)$ which contains a given $G(l)$ and has as few edges as possible. Not being good at finding algorithms but being interested in extremal problems, when Even told me of these questions I asked: Determine or estimate the smallest integer $f(n)$ so that for every $G(n)$ one can add $\leq f(n)$ new edges so that the resulting graph should be a rigid circuit graph. I first thought that $f(n) < (1+o(1))\frac{n^2}{8}$, but Even showed by a simple construction that $f(n)$ must be greater than $[\frac{n^2}{4}]$. Then I observed that in fact $f(n) = \frac{n^2}{2}(1+o(1))$. More precisely the methods of Rényi and myself give the following result: Let $\alpha > 0$ be sufficiently small. Consider all the graphs $G(n;t)$ where $t = [n^{2-\alpha}]$. The number of these graphs is clearly $p = \binom{\binom{n}{2}}{t}$. There is an $\alpha' > 0$ so that all but $o(p)$ of these graphs have the following property. Consider all those C_4's of our $G(n;t)$ which do not have any diagonals. Add one of the diagonals to all the C_4's. Then the resulting graph will always have more than $\binom{n}{2} - n^{2-\alpha'}$ edges.

This clearly implies

$$f(n) > \binom{n}{2} - n^{2-\varepsilon} \tag{1}$$

for a certain $\varepsilon > 0$. I suspect but can not prove that this method might give $f(n) > \binom{n}{2} - n^{3/2+\varepsilon}$.

The exact determination or better estimation of $f(n)$ seems to me to be an interesting problem. I have no non trivial upper bound for $f(n)$

and can not even prove

$$f(n) < \frac{n^2}{2} - cn$$

for every $c > 0$ and $n > n_0(c)$.

Even informed me that the following problem has been considered: Let G be a planar graph of n vertices. Find the smallest rigid circuit graph containing G(smallest of course means having the fewest number of edges). If I remember correctly examples are known of planar graphs G for which every rigid circuit graph containing G must have at least $n^{3/2}$ edges but no non trivial upper bounds are known.

Several modifications of the problem of estimating $f(n)$ seem to have some interest, here I only state one of them. Let $f_1(n)$ be the smallest integer with the following property: To every $G(n)$ we can add $\leq f_1(n)$ edges so that the resulting new graph should contain at least one diagonal of every C_4 of our G. In fact in (1) we really proved $f_1(n) > \binom{n}{2} - n^{2-\varepsilon}$. It is not difficult to show

$$f_1(n) < \binom{n}{2} - cn^{3/2}.$$

Clearly $f_1(n) \leq f(n)$. An intermediary function $f_2(n)$ can be defined as follows: $f_2(n)$ is the smallest integer so that to every $G(n)$ we can add $\leq f_2(n)$ edges so that in the resulting graph every C_4 should have a diagonal. Clearly $f(n) \geq f_2(n) \geq f(n)$. I am certain that for $n > n_0$ the inequalities are strict, perhaps $f_1(n) > \binom{n}{2} - n^{3/2+\varepsilon}$ for every $\varepsilon > 0$ and $n > n_0(\varepsilon)$.

5. Hamiltonian circuits of random graphs. Rényi and I conjectured that if c is sufficiently large then all but $o\left(\binom{\binom{n}{2}}{t}\right)$ graphs $G(n;t)$, $t = [cn \log n]$ are Hamiltonian.

M. Wright proved this with $t = [cn^{3/2}]$. Recently Posa proved our conjecture and by a refinement of his method Komlos and Szemeredi proved that the result holds with $t = [(1/2+\varepsilon)n \log n]$.

Perhaps the following two problems of Rényi and myself are of interest: For what values of t is it true that if we know that $G(n;t)$ is connected then with probability tending to 1 as n tends to infinity, it is Hamiltonian?

Put $\binom{\binom{n}{2}}{[cn]} = T(n,c)$. Is it true that there is a function $f(c)$ so that all but $o(T(n,c))$ graphs $G(n;[cn])$ have their longest circuit of size $(1+o(1))f(c)n$? We know that $f(c) = 0$ for $c \le \frac{1}{2}$ and believe that $f(c)$ is continuous strictly increasing for $\frac{1}{2} < c < \infty$, further

$$\lim_{c=\infty} f(c) = 1.$$

Spencer and I have the following further conjectures. Rényi and I proved that if $t = [\frac{1}{2}n \log n + \frac{1}{2}n \log\log n + cn]$ then the probability that $G(n;t)$ has all its vertices of valency ≥ 2 is $1-e^{-e^{-2c}}$. We hope that the probability that it is Hamiltonian given by the same expression.

The results of Rényi and myself on linear factors of random graphs give our conjecture some support.

Spencer and I have various further conjectures on this subject we had no time to think much about them thus we are not sure whether they are likely to lead to good results. Let $G(n;t)$ be a random graph. Suppose we know that each of its vertices has valency ≥ 2. For which t can we conclude that the conditional probability of it having a

Hamiltonian circuit tends to 1. We feel sure that this will hold for much smaller values of t than $\frac{1}{2}n\log n(\sigma(n\log n)^{?})$.

Further we formulated the following two problems: Let $G(n;k)$ be a graph of n vertices and k edges which has no Hamiltonian circuit. We add edges to it at random. How many edges do we have to add that with probability tending to 1 the resulting graph should be Hamiltonian? In this generality the problem is trivial $G(n;k)$ may be such that the addition of every further edge makes it Hamiltonian, but it seems that if k is not too large (say $< cn$ or $cn \log n$) then we hope that one will have to add "many" edges (more than $cn^{?}$) to make the graph Hamiltonian with probability tending to 1.

The other problem is perhaps more interesting. Consider all the $\binom{n}{2}!$ numberings of the edges of the complete graph with the integers $1,2,\ldots,\binom{n}{2}$. To each of these enumerations we associate two integers n_1 and n_2 $n \le n_1 \le n_2 < \binom{n}{2}$ so that n_1 is the smallest integer for which the first n_1 edges give a graph each vertex of which has valency ≥ 2 and n_2 is the smallest integer for which the first n_2 edges give a graph with a Hamiltonian cycle. What can be said about the expected value and distribution of $n_2 - n_1$? It may be that $n_2 = n_1$ for almost all of the $\binom{n}{2}!$ permutations. This could be an extremely strong result but it appears very difficult to show.

Clearly both these problems can be formulated for other properties too.

The following further (non-probabilistic) problem is due to B.Bollobá
Determine the smallest integer $h(n)$ for which there is a graph $G(n;h(n))$

which is not Hamiltonian but if one adds any further edge the graph becomes Hamiltonian. As far as we know the problem is still unsolved.

P. Erdös and A. Rènyi, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci. 5(1960), 17 - 67 and On the strength of connectedness of a random graph, Acta Math. Acad. Sci. Hungar. 12(1961), 261-267, see also The Art of Counting p. 574-624.