Almost all graphs with high girth and suitable density have high chromatic number

Deryk Osthus*, Hans Jürgen Prömel and Anusch Taraz
Institut für Informatik, Humboldt-Universität zu Berlin
Unter den Linden 6, 10099 Berlin, Germany
E-mail address: {osthus,proemel,taraz}@informatik.hu-berlin.de

Abstract

Erdős proved that there exist graphs of arbitrarily high girth and arbitrarily high chromatic number. We give a different proof (but also using the probabilistic method) that also yields the following result on the typical asymptotic structure of graphs of high girth: for all $\ell \geq 3$ and $k \in \mathbb{N}$ there exist constants C_1 and C_2 so that almost all graphs on n vertices and m edges whose girth is greater than ℓ have chromatic number at least k, provided that $C_1 n \leq m \leq C_2 n^{\ell/(\ell-1)}$.

1 Introduction and Results

In 1959, Erdős [4] proved that there are graphs of arbitrarily large girth and arbitrarily large chromatic number. (Here the *girth* of a graph G is the length of its shortest cycle and is denoted by girth(G).) His proof is one of the first and most well-known examples of the probabilistic method: he showed that with high probability one can alter a random graph (with suitable edge probability) so that it has no short cycles and no large independent sets. Here we give a proof (also using the probabilistic method) which gives more information about the typical asymptotic structure of graphs of high girth and given density.

Let $\mathcal{F}_{n,m}(C_{\leq \ell})$ denote the set of all graphs with n vertices and m edges which contain no cycle whose length is at most ℓ, (writing $\mathcal{F}_{n,m}(K_3)$ instead of $\mathcal{F}_{n,m}(C_{\leq 3})$). We say that *almost all* graphs in $\mathcal{F}_{n,m}(C_{\leq \ell})$ have some property if the proportion of graphs in $\mathcal{F}_{n,m}(C_{\leq \ell})$ with this property tends to one as n tends to infinity.

Theorem 1. *For all $\ell \geq 3$ and $k \in \mathbb{N}$, there are constants C_1 and C_2 so that almost all graphs in $\mathcal{F}_{n,m}(C_{\leq \ell})$ have chromatic number at least k, provided that $C_1 n \leq m \leq C_2 n^{\ell/(\ell-1)}$.*

Let $G_{n,m}$ denote a graph chosen uniformly at random from the set of graphs with n vertices and m edges. We say that $G_{n,m}$ has some property $\mathcal{Q}$ *almost surely* if the probability that it has $\mathcal{Q}$ tends to one as n tends to infinity. The restriction that $m \geq C_1 n$ in Theorem 1 is clearly necessary, since for $m = o(n)$,

*Graduate school "Algorithmische Diskrete Mathematik", supported by Deutsche Forschungsgemeinschaft grant GRK/3-98

$G_{n,m}$ almost surely contains no cycles at all. For the case $\ell = 3$, it turns out that the restriction that $m \leq C_2 n^{3/2}$ is also close to best possible. Indeed, building on earlier results, in [8] we showed the following. Set

$$t_3 = t_3(n) = \frac{\sqrt{3}}{4} n^{3/2} \sqrt{\log n}$$

and fix any $\varepsilon > 0$. Then if $m \geq (1+\varepsilon)t_3$, almost all graphs in $\mathcal{F}_{n,m}(K_3)$ are bipartite. This threshold is sharp in the sense that if $n/2 \leq m \leq (1-\varepsilon)t_3$, then almost no graph in $\mathcal{F}_{n,m}(K_3)$ is bipartite.

Instead of Theorem 1, we actually prove the following stronger result, which gives a lower bound on the chromatic number of almost all graphs in $\mathcal{F}_{n,m}(C_{\leq \ell})$ in terms of n and m.

Theorem 2. *For all $\ell > 3$ there exist constants d_1, d_2 and d_3 with the following properties. Let*

$$m_0 = d_1 n^{(\ell+2)/(\ell+1)} (\log n)^{2/(\ell+1)}.$$

If $2n \leq m \leq m_0$, then almost all graphs in $\mathcal{F}_{n,m}(C_{\leq \ell})$ have chromatic number at least

$$\frac{m}{2n \log(2m/n)}. \tag{1}$$

If $m_0 \leq m \leq d_2 n^{\ell/(\ell-1)}$, then almost all graphs in $\mathcal{F}_{n,m}(C_{\leq \ell})$ have chromatic number at least

$$d_3 \sqrt{n^\ell / m^{\ell-1}}. \tag{2}$$

We have made no attempt to find the best constants that can be obtained from our proof of Theorem 2. Note that for $m \leq m_0$, the bound is of the same order of magnitude as that which is known for $G_{n,p}$, where $p = m/\binom{n}{2}$ and $G_{n,p}$ is a random graph with n vertices with edge probability p. In fact Łuczak (see e.g. [5]) proved that if $pn \to \infty$ and $p \to 0$, then the chromatic number of $G_{n,p}$ is almost surely

$$(1+o(1)) \frac{pn}{2 \log(pn)}.$$

It seems likely that the chromatic number of almost all graphs in $\mathcal{F}_{n,m}(C_{\leq \ell})$ is $\Theta(\frac{m}{n \log(m/n)})$ whenever $n \ll m \ll n^{\ell/(\ell-1)}$. However, this seems to be significantly more difficult to prove than Theorem 2 even for the triangle-free case.

Related to this is the question of how high the chromatic number of a graph can be if it has n vertices and girth greater than ℓ. Let $f(n,\ell)$ be the maximum chromatic number of such a graph. The proof of Erdős [4] shows that for fixed ℓ, $f(n,\ell) \geq n^{1/\ell + o(1)}$. For the triangle-free case $\ell = 3$ this was improved by Kim [6], who solved a longstanding open question by showing that $f(n,3) \geq \frac{1}{9} n^{1/2} / \sqrt{\log n}$, which (by a result of Ajtai, Komlós, and Szemerédi [1]) is best possible up to the value of the constant factor. It is well known (see e.g. Krivelevich [7, Lemma 6.1] or [9]) that $f(n,\ell) \geq n^{1/(\ell-1)+o(1)}$, which is the best known lower bound for $\ell > 3$. As pointed out to us by one of the referees, an upper bound on $f(n,\ell)$, where $\ell > 3$ is even, may be obtained as follows. For even ℓ, Bondy and Simonovits [3] showed that a C_ℓ-free graph has $\mathcal{O}(n^{1+2/\ell})$ edges. Thus it has an independent set of size $\Omega(n^{1-2/\ell})$. Removing this set and applying induction, it is easily seen that such a graph has chromatic number $\mathcal{O}(n^{2/\ell})$ and thus $f(n,\ell) = \mathcal{O}(n^{2/\ell})$ for even ℓ. This can be improved by a logarithmic factor using the results on independents sets in [1] (see also [2, Lemma XII.15]). The bounds obtained from Theorem 2 are much smaller than the lower bounds mentioned above: they achieve their maximum when $m = m_0 = n^{(\ell+2)/(\ell+1)+o(1)}$, where they imply that almost all graphs in $\mathcal{F}_{n,m}(C_{\leq \ell})$ have chromatic number at least $n^{1/(\ell+1)+o(1)}$.

In the remainder of this note, we prove Theorem 2. Although the proof is not quite as simple as that of the original existence result of Erdős, it turns out to be fairly straightforward. Indeed, for a graph G let $\alpha(G)$ denote the size of a largest independent set of vertices. Since for a graph G on n vertices, we have $\chi(G) \geq n/\alpha(G)$, it suffices to show that almost all graphs in $\mathcal{F}_{n,m}(C_{\leq \ell})$ have no large independent set (where m satisfies the conditions of the theorem). This is done by demonstrating that for suitable choices of parameters, the probability that there is a "large" independent set in $G_{n,m}$ is much smaller than the probability that $G_{n,m}$ has girth greater than ℓ.

2 Proof of Theorem 2

Throughout this section, we set $p = m/\binom{n}{2}$. Using the fact that $\chi(G) \geq n/\alpha(G)$ for any graph G on n vertices, Theorem 2 follows immediately from the following lemma. Throughout, we assume that n is large enough for our estimates to hold and we denote by $G_{n,m}$ a graph chosen uniformly at random from the set of graphs with n vertices and m edges.

Lemma 3. *For all $\ell > 3$ there exist constants c_1, c_2 and c_3 with the following properties. Let*

$$p_0 = c_1 n^{-\ell/(\ell+1)}(\log n)^{2/(\ell+1)}.$$

If $4/n \leq p \leq p_0$, then

$$\mathbb{P}[\,\alpha(G_{n,m}) \geq \frac{4}{p}\log(np) \mid \mathrm{girth}(G_{n,m}) > \ell\,] = o(1). \tag{3}$$

If $p_0 \leq p \leq c_2 n^{-(\ell-2)/(\ell-1)}$, then

$$\mathbb{P}[\,\alpha(G_{n,m}) \geq c_3\sqrt{p^{\ell-1}n^{\ell}} \mid \mathrm{girth}(G_{n,m}) > \ell\,] = o(1). \tag{4}$$

To prove Lemma 3, we shall need Lemma 4 (see also Prömel and Steger [10] and Theorem 3.11 in [5] for similar results), whose proof relies on the FKG-inequality (see e.g. [5]). For $i \geq 3$, let X_i denote the number of i-cycles in $G_{n,m}$. Note that

$$\mathbb{E}[X_i] = (1+o(1))\frac{(n)_i p^i}{2i} = \Theta(m^i/n^i).$$

Lemma 4. *For any $\ell \geq 3$, there are constants c, $c' > 0$ so that if $2n \leq m \leq c' n^{\ell/(\ell-1)}$,*

$$\mathbb{P}[\,\mathrm{girth}(G_{n,m}) > \ell\,] \geq \mathrm{e}^{-c\mathbb{E}[X_\ell]}.$$

Proof. We will make use of the inequality

$$1 - x \geq \mathrm{e}^{-x-x^2} \geq \mathrm{e}^{-2x}, \tag{5}$$

valid for $x \leq 1/2$ (see e.g. page 5 of [2]). Since for $i \geq 3$, the number of i-cycles in the complete graph on n vertices is $\frac{(n)_i}{2i}$, the FKG-inequality implies that

$$\begin{aligned}\mathbb{P}[\,\mathrm{girth}(G_{n,2p}) > \ell\,] &\geq \prod_{i=3}^{\ell}(1-(2p)^i)^{\frac{(n)_i}{2i}} \overset{(5)}{\geq} \prod_{i=3}^{\ell}\mathrm{e}^{-2(2p)^i\frac{(n)_i}{2i}} \\ &\geq \prod_{i=3}^{\ell}\mathrm{e}^{-3\cdot 2^i\,\mathbb{E}[X_i]} \geq \mathrm{e}^{-3\ell\, 2^\ell\,\mathbb{E}[X_\ell]},\end{aligned} \tag{6}$$

where the last line follows since $m \geq 2n$ implies that $\mathbb{E}[X_i] \leq \mathbb{E}[X_\ell]$ for $3 \leq i \leq \ell$.

But since the property of containing no cycle of length at most ℓ is monotone decreasing, we have (denoting by $e(G)$ the number of edges of a graph G and letting $N = \binom{n}{2}$)

$$\begin{aligned}\mathbb{P}[\,\mathrm{girth}(G_{n,2p}) > \ell\,] &\leq \mathbb{P}[\,\mathrm{girth}(G_{n,m}) > \ell\,] + \mathbb{P}[|e(G_{n,2p}) - 2pN| \geq pN] \\ &\leq \mathbb{P}[\,\mathrm{girth}(G_{n,m}) > \ell\,] + \mathrm{e}^{-pN/12},\end{aligned} \tag{7}$$

where the last line follows from standard tail estimates for the binomial distribution (see e.g. Theorem 7(i) in [2]). Thus (6) and (7) imply that

$$\mathbb{P}[\,\mathrm{girth}(G_{n,m}) > \ell\,] \geq \mathrm{e}^{-3\ell\, 2^\ell\,\mathbb{E}[X_\ell]} - \mathrm{e}^{-m/12}.$$

The result now follows immediately by observing that for c' sufficiently small, $m \leq c' n^{\ell/(\ell-1)}$ implies that m is significantly larger than $\mathbb{E}[X_\ell]$. □

We shall also need Pittel's inequality (see page 35 in [2]), which states that if $\mathcal{Q}$ is any property and $0 < p = m/\binom{n}{2} < 1$, then

$$\mathbb{P}[\,G_{n,m} \text{ has } \mathcal{Q}\,] \leq 3\sqrt{m}\,\mathbb{P}[\,G_{n,p} \text{ has } \mathcal{Q}\,]. \tag{8}$$

Proof of Lemma 3. First note that for any $r = r(n)$ with $r \to \infty$,

$$\mathbb{P}[\alpha(G_{n,p}) \geq r] \leq \binom{n}{r}(1-p)^{\binom{r}{2}} \leq (\mathrm{e}n/r)^r \mathrm{e}^{-pr(r-1)/2} = \mathrm{e}^{-(1+o(1))\phi},$$

where for convenience we write

$$\phi = r(pr/2 - \log(n/r)).$$

Then by Lemma 4 and (8), there is a constant $c > \ell$ so that

$$\begin{aligned}\mathbb{P}[\,\alpha(G_{n,m}) \geq r \mid \mathrm{girth}(G_{n,m}) > \ell\,] &\leq \frac{\mathbb{P}[\,\alpha(G_{n,m}) \geq r\,]}{\mathbb{P}[\,\mathrm{girth}(G_{n,m}) > \ell\,]} \\ &\leq 3\sqrt{m}\,\mathbb{P}[\,\alpha(G_{n,p}) \geq r\,]\,\mathrm{e}^{c\mathbb{E}[X_\ell]} \\ &\leq 3\sqrt{m}\,\mathrm{e}^{-(1+o(1))(\phi - c\mathbb{E}[X_\ell])}.\end{aligned}$$

Thus to prove (3), it suffices to prove that if $r = \frac{4}{p}\log(np)$ and $4/n \leq p \leq p_0$ (where c_1 in the definition of p_0 will be determined below), then $\phi \geq 4\log n$ and $\phi/\mathbb{E}[X_\ell] \geq 2c$. Note that our choice of r implies that $\log(1/p) = (1+o(1))\log r$. This in turn implies that $pr/4 = (1+o(1))\log(n/r)$ and thus that

$$\phi = (1+o(1))pr^2/4 = (1+o(1))\frac{4}{p}\left(\log(np)\right)^2 \geq 4\log n,$$

with room to spare. Also

$$\frac{\phi}{\mathbb{E}[X_\ell]} = (1+o(1))\frac{p}{4}\left(\frac{4}{p}\log(np)\right)^2\frac{2\ell}{n^\ell p^\ell} = (1+o(1))\frac{8\ell(\log(np))^2}{n^\ell p^{\ell+1}} \geq 2c,$$

as required. The final inequality holds if we choose c_1 (in the definition of p_0) sufficiently small compared to c.

Inequality (4), where $p_0 \leq p \leq c_2 n^{-(\ell-2)/(\ell-1)}$, is dealt with in a similar way. Indeed, setting $r = c_3\sqrt{p^{\ell-1}n^\ell}$, where c_3 is chosen to be sufficiently large compared to c_1, gives

$$pr/4 \geq \frac{c_3}{4}\sqrt{p_0^{\ell+1}n^\ell} \geq \log(n/r).$$

This in turn implies

$$\phi \geq pr^2/4 = (pn)^{\ell+o(1)} \geq 4\log n.$$

Also, we have

$$\frac{\phi}{\mathbb{E}[X_\ell]} = (1+o(1))\frac{p}{4}\,c_3^2 p^{\ell-1}n^\ell\,\frac{2\ell}{n^\ell p^\ell} = (1+o(1))\frac{c_3^2\ell}{2} \geq 2c,$$

as required, provided we choose c_3 sufficiently large compared to c. □

References

[1] M. Ajtai, J. Komlós, and E. Szemerédi, A note on Ramsey numbers, *J. Combin. Theory Ser. A* **29** (1980), 354–360.

[2] B. Bollobás, *Random Graphs*, Academic Press, London, 1985.

[3] J.A. Bondy and M. Simonovits, Cycles of even length in graphs, *J. Combin. Theory Ser. B* **16** (1974), 97–105.

[4] P. Erdős, Graph theory and probability, *Canad. J. Mathematics* **11** (1959), 34–38.

[5] S. Janson, T. Łuczak, and A. Ruciński, *Random Graphs*, Wiley-Interscience, New York, 2000.

[6] J.H. Kim, The Ramsey number $R(3,t)$ has order of magnitude $t^2/\log t$, *Random Struct. Algorithms* **7** (1995), 173–207.

[7] M. Krivelevich, On the minimal number of edges in color-critical graphs, *Combinatorica* **17** (1997), 401–426.

[8] D. Osthus, H.J. Prömel, and A. Taraz, For which densities are random triangle-free graphs almost surely bipartite?, *Combinatorica*, to appear.

[9] D. Osthus and A. Taraz, Random maximal H-free graphs, *Random Struct. Algorithms* **18** (2001), 61–82.

[10] H.J. Prömel and A. Steger, Counting H-free graphs, *Discr. Math.* **154** (1996), 311–315.