MULTICOLOURED TREES IN RANDOM GRAPHS

Alan Frieze*
Department of Mathematics,
Carnegie Mellon University,
Pittsburgh, U.S.A.

and

Brendan D. McKay
Department of Computer Science,
Australian National University
Canberra, Australia

March 1, 2001

1 INTRODUCTION

Let $G = (V, E)$ be a graph in which the edges are coloured. A set $S \subseteq E$ is said to be *multicoloured* if each edge of S is a different colour. A spanning tree of G is said to be multicoloured if its edge set is. In this paper we study

*Supported by NSF grant CCR-9024935

the existence of a multicoloured spanning tree (MST) in a randomly coloured random graph.

In fact, our main result will concern a randomly coloured graph process. Here $e_1, e_2, \ldots, e_N$ is a random permutation of the edges of the complete graph K_n and so $N = \binom{n}{2}$. Each edge e independently chooses a random colour $c(e)$ from a given set of colours W, $|W| \geq n-1$.

The graph process consists of the sequence of random graphs $G_m, m = 1, 2, \ldots, N$, where $G_m = ([n], E_m)$ and $E_m = \{e_1, e_2, \ldots, e_m\}$. We identify the following events:

$\mathcal{C}_m = \{G_m$ is connected $\}$.

$\mathcal{N}_m = \{|W_m| \geq n-1\}$, where W_m is the set of colours used by E_m.

$\mathcal{MT}_m = \{G_m$ has a multicoloured spanning tree $\}$.

Let $\mathcal{E}_m$ stand for one of the above three sequences of events and let

$$m_{\mathcal{E}} = \min\{m : \mathcal{E}_m \text{ occurs}\},$$

provided such an m exists. Clearly, if $m_{\mathcal{MT}}$ is defined,

$$m_{\mathcal{MT}} \geq \max\{m_{\mathcal{C}}, m_{\mathcal{N}}\},$$

and the main result of the paper is

Theorem 1 *In almost every (a.e.) randomly coloured graph process*

$$m_{\mathcal{MT}} = \max\{m_{\mathcal{C}}, m_{\mathcal{N}}\}.$$

□

To establish the existence of an MST we use a result of Edmonds [2] on the matroid intersection problem. In this scenario M_1, M_2 are matroids over a common ground set E with rank functions r_1, r_2 respectively. Edmonds' general theorem on this problem is

$$\max(|I| : I \text{ is independent in both matroids}) = \min_{\substack{E_1 \cup E_2 = E \\ E_1 \cap E_2 = \emptyset}} (r_1(E_1) + r_2(E_2)). \tag{1}$$

For us M_1 is the cycle matroid of a graph $G = G_m$ and M_2 is the partition matroid associated with the colours. Thus for a set of edges S, $r_1(S) = n - \kappa(S)$ where $\kappa(S)$ is the number of components of the graph $G_S = ([n], S)$ and $r_2(S)$ is the number of distinct colours occurring in S. If $i \in W$ then C_i denotes the set of edges of colour i and for $I \subseteq W$, $C_I = \bigcup_{i \in I} C_i$. We will use Edmonds' theorem as follows:

Theorem 2 *A necessary and sufficient condition for the existence of an MST is that*

$$\kappa(C_I) \leq |W| + 1 - |I| \qquad \textit{for all } I \subseteq W. \tag{2}$$

Proof To see this, w.l.o.g. restrict attention in (1) to E_2 of the form C_J and then take $I = W \setminus J$ in (2). □

2 Proof of Theorem 1

Observe first that if $\omega = \omega(n) \to \infty$ slowly, then in a.e. randomly coloured graph process

$$m_{\mathcal{C}} \geq m_0 = \lfloor \tfrac{1}{2} n (\ln n - \omega) \rfloor \text{ and } m_{\mathcal{N}} \leq m_1 = \lceil n (\ln n + \omega) \rceil.$$

Fix some m in the range $[m_0, m_1]$ and let $w_m = |W_m|$. We define the event

$$\mathcal{A}_k = \{\exists I \subseteq W_m, |I| = k : \kappa(C_I) \geq w_m - |I| + 2\}.$$

We know that if $m \geq \max\{m_{\mathcal{C}}, m_{\mathcal{N}}\}$ and there is no MST then $\mathcal{A}_k$ occurs for some $k \in [3, w_m - 1]$ ($\mathcal{A}_1 \cup \mathcal{A}_2$ cannot occur since the colours of W_m are all used and $\mathcal{A}_{w_m}$ cannot occur if G_m is connected.) Take a minimal k, corresponding set I and let $S = C_I$.

Claim 1 *G_S has no bridges.*

Proof If there is a bridge, remove it and all edges of the same colour. Clearly $\mathcal{A}_{k-1}$ occurs, contradicting the minimality of k. □

With the notation of Claim 1 suppose then that G_S has i isolated vertices and $n-k+x-i$ non-trivial components, $x \geq 1$. Since non-trivial components without bridges have at least three vertices,

$$i + 3(n - k + x - i) \leq n \tag{3}$$

or

$$\begin{aligned} i &\geq n - \frac{3}{2}k + \frac{3}{2}x \\ &\geq n - \frac{3}{2}k + \frac{3}{2}. \end{aligned}$$

So now let $\mathcal{B}_k$ denote the event

$$\begin{aligned} \{\exists I \subseteq W_m, |I| = k, T \subseteq [n] : \ & t = |T| \leq 3(k-1)/2, \\ & \text{all edges coloured with } I \text{ are contained in } T, \\ & \text{there are } u \geq \max\{k, t\} \ I\text{-coloured edges}\}. \end{aligned}$$

Here T is the set of vertices in the non-trivial components of G_{C_I}. Thus,

$$\mathcal{N}_m \cap \mathcal{A}_k \subseteq \bigcup_{i=3}^{k} \mathcal{B}_i \qquad \text{for } k \geq 3. \tag{4}$$

For large k ($\geq 9n/10$) we consider a slightly different event.

We first rephrase (2) as

$$\kappa(C_{W/J}) \leq |J| + 1 \qquad \text{for all } J \subseteq W. \tag{5}$$

So if $m \geq \max\{m_{\mathcal{C}}, m_{\mathcal{N}}\}$ and there is no MST then there exist $\ell \geq 1$ colours whose deletion produces $\lambda \geq \ell + 2$ components of sizes $n_1, \ldots, n_\lambda$ ($\ell = 0$ is ruled out by the connectivity of G_m).

Claim 2 *Some subsequence of the n_i's sums to between $\ell + 1$ and $n/2$.*

Proof Assume $n_1 \leq n_2 \leq \cdots \leq n_\lambda$.

If $n_\lambda \geq \ell + 1$, one of $n_1, ..., n_{\lambda-1}$ and n_λ suffices.

Suppose then that $n_i \leq \ell$, $1 \leq i \leq \lambda$.

Choose r such that

$$n_1 + \cdots + n_r \leq n/2, \quad n_1 + \cdots + n_{r+1} > n/2$$

and then

$$\begin{aligned} n_1 + \cdots + n_r &> n/2 - n_{r+1} \\ &\geq n/2 - \ell \\ &\geq \ell. \end{aligned}$$

and we can take $n_1, ..., n_r$. □

Note next that if J is minimal in (5) then each colour in J appears at least twice as an edge joining components of $G_{C_{W \setminus J}}$.

So if $m \geq \max\{m_{\mathcal{C}}, m_{\mathcal{N}}\}$ and there is no MST and $\mathcal{A}_k$ does not occur for $k \leq 9n/10$ then there is a set L of $1 \leq \ell < w_m - 9n/10$ colours and a set S of size s, $\ell + 1 \leq s \leq n/2$ such that (i) all $t = \eta(S) = |(S : \bar{S})| \geq 1$ edges are L-coloured, ($(S : \bar{S})$ is the set of edges joining S and $\bar{S} = V \setminus S$), (ii) the lexicographically first $\max\{2\ell - t, 0\}$ non-$(S : \bar{S})$ edges joining up components (of the $W \setminus L$ coloured edges) are also L-coloured. Let $\mathcal{D}_\ell$ denote this event. Then

$$\mathcal{C}_m \cap \left(\bigcup_{k=9n/10}^{w_m-1} \mathcal{A}_k \right) \subseteq \bigcup_{\ell=1}^{w_m-9n/10} \Pr_m(\mathcal{D}_\ell). \tag{6}$$

It follows from (4) and (6) that

$$\Pr(m_{\mathcal{MT}} > \max\{m_{\mathcal{N}}, m_{\mathcal{C}}\}) \leq$$

$$o(1) + \sum_{m=m_0}^{m_1} \left[\sum_{k=3}^{9n/10} \Pr_m(\mathcal{B}_k) + \sum_{\ell=2}^{w_m-9n/10} \Pr_m(\mathcal{D}_\ell) \right] + \Pr\left(\bigcup_{m=m_0}^{m_1} (\mathcal{C}_m \cap \mathcal{A}_{w_m-1}) \right). \tag{7}$$

Here $\Pr_m$ denotes probability w.r.t. G_m and the o(1) term is the probability that G_{m_0} is connected or that $m_{\mathcal{N}} > m_1$. (Our calculations force us to separate out $\mathcal{A}_{w_m-1}$.)

We must now estimate the individual probabilities in (7). It is easier to work with the independent model G_p, $p = m/N$, where each edge occurs independently with probability p and is then randomly coloured. For any event $\mathcal{E}$ we have (see Bollobás [1] Chapter II) the simple bound

$$\Pr_m(\mathcal{E}) \leq 3\sqrt{n \ln n} \Pr_p(\mathcal{E}). \tag{8}$$

where $\Pr_p$ denotes probability w.r.t. the model G_p.

2.1 Few colours

We thus consider $p = \alpha \ln n/n$, $1 - o(1) \leq \alpha \leq 2 + o(1)$. We will initially assume that $|W| = n + c$, $-1 \leq c \leq \epsilon n$ where ϵ is some small fixed positive number ($\epsilon = .01$ is suitable). Then

$$\begin{aligned}
\Pr_p(\mathcal{B}_k) &\leq \sum_{t=1}^{3(k-1)/2} \sum_{u=\max\{t,k\}}^{\binom{t}{2}} \binom{n}{t}\binom{n+c}{k}\binom{\binom{t}{2}}{u}\left(1-\frac{kp}{n+c}\right)^{\binom{n}{2}-u}\left(\frac{kp}{n+c}\right)^{u} \\
&\leq \sum_{t=1}^{3(k-1)/2} \sum_{u=\max\{t,k\}}^{\binom{t}{2}} \frac{n^t e^t}{t^t}\frac{n^k e^{(1+\epsilon)k}}{k^k}\left(\frac{t^2 e}{2u}\right)^{u} n^{-k\alpha(1-\epsilon)/2}\left(\frac{\alpha k \ln n}{n^2}\right)^{u}. \quad (9)
\end{aligned}$$

Case 1: $3 \leq k \leq k_0 = n/(3\ln n)$.

$$\begin{aligned}
\Pr_p(\mathcal{B}_k) &\leq \sum_{t=1}^{3(k-1)/2} \sum_{u=\max\{t,k\}}^{\binom{t}{2}} \left(\frac{e^3 n^{1-\alpha(1-\epsilon)/2}}{k}\right)^{k}\left(\frac{t}{n}\right)^{2u-t}\left(\frac{\alpha e k \ln n}{2u}\right)^{u} \\
&= \sum_{t=1}^{3(k-1)/2} \sum_{u=\max\{t,k\}}^{\binom{t}{2}} \left(\frac{e^3 n^{1-\alpha(1-\epsilon)/2}}{k}\right)^{k}\left(\frac{t}{n}\right)^{u-t}\left(\frac{\alpha e k t \ln n}{2un}\right)^{u} \\
&\leq \sum_{t=1}^{3(k-1)/2} \sum_{u=\max\{t,k\}}^{\binom{t}{2}} \left(\frac{e^3 n^{1-\alpha(1-\epsilon)/2}\alpha e k \ln n}{2kn}\right)^{k}\left(\frac{t}{n}\right)^{u-t}\left(\frac{\alpha e k \ln n}{2n}\right)^{u-k} \\
&= O\left(\left(\frac{e^5 \ln n}{n^{\alpha(1-\epsilon)/2}}\right)^{k}\right).
\end{aligned}$$

It follows from this and (8) that

$$\begin{aligned}
\sum_{m=m_0}^{m_1}\sum_{k=4}^{k_0}\Pr_m(\mathcal{B}_k) &= O((n\ln n)(\sqrt{n\ln n})((\ln n)^4/n^{2\alpha(1-\epsilon)})) \\
&= o(1). \quad (10)
\end{aligned}$$

For $k=3$ we compute $\Pr_m(\mathcal{B}_3)$ directly, but since now $u=t=k=3$ is forced,

$$\begin{aligned}\Pr_m(\mathcal{B}_3) &\leq \binom{n}{3}^2 \left(1-\frac{3}{n+c}\right)^{m-3} \left(\frac{3}{n+c}\right)^3 \frac{\binom{N-3}{m-3}}{\binom{N}{m}} \\ &= O(e^{3\omega}(\ln n)^3 n^{-3/2})\end{aligned}$$

and so

$$\sum_{m=m_0}^{m_1} \Pr_m(\mathcal{B}_3) = o(1). \tag{11}$$

Case 2: $k_0 < k \leq n/2$.

We now write (9) as

$$\begin{aligned}\Pr_p(\mathcal{B}_k) &\leq \sum_{t=1}^{3(k-1)/2} \sum_{u=\max\{t,k\}}^{\binom{t}{2}} \left(\frac{e^3 n^{1-\alpha(1-\epsilon)/2}}{k}\right)^k \left(\frac{t}{n}\right)^{u-t} \left(\frac{\alpha e k t \ln n}{2un}\right)^u \\ &\leq \sum_{t=1}^{3(k-1)/2} \sum_{u=\max\{t,k\}}^{\binom{t}{2}} \left(\frac{e^3 n^{1-\alpha(1-\epsilon)/2}}{k}\right)^k \left(\frac{t}{n}\right)^{u-t} n^{\frac{\alpha t k}{2n}} \\ &\qquad \text{(after maximising the last term over } u\text{)} \\ &= \sum_{t=1}^{3(k-1)/2} \sum_{u=\max\{t,k\}}^{\binom{t}{2}} \left(\frac{e^3 n^{1-\frac{\alpha}{2}(1-\frac{t}{n}-\epsilon)}}{k}\right)^k \left(\frac{t}{n}\right)^{u-t} &(12) \\ &\leq \sum_{t=1}^{3(k-1)/2} \sum_{u=\max\{t,k\}}^{\binom{t}{2}} \left(\frac{e^3 n^{1-\alpha(\frac{1}{8}-\epsilon)}}{k}\right)^k &(13)\end{aligned}$$

since $t \leq 3(k-1)/2 \leq 3n/4$.

(13) and (8) clearly imply

$$\sum_{m=m_0}^{m_1} \sum_{k=k_0}^{n/2} \Pr_m(\mathcal{B}_k) = o(1). \tag{14}$$

Case 3: $n/2 < k \leq 9n/10$

Claim 3 *Choose any constant $A > 0$. Then, in a.e. process, simultaneously for each $m \in [m_0, m_1]$, the sets of $s \leq A$ vertices of G_m which span at least s edges together contain at most $(\ln n)^{A+1}$ vertices.*

Proof We need only prove this for G_{m_1} and since the property is monotone decreasing we need only prove it for G_{p_1}, $p_1 = m_1/N$ ([1], Chapter II.) But

$$\begin{aligned} E_{p_1}(\text{number of vertices}) &\leq \sum_{k=3}^{A} \binom{n}{k} \binom{\binom{k}{2}}{k} p_1^k k \\ &= O(e^{2A}(\ln n)^A). \end{aligned}$$

Now use the Markov inequality. □

It follows that we may rewrite (3) as

$$i + 3(\ln n)^{A+1} + (A+1)(n - k + x - i) \leq n$$

and so

$$\begin{aligned} i &\geq n - \frac{A+1}{A} k - O((\ln n)^{A+1}) \\ &\geq n - \frac{A}{A-1} k. \end{aligned}$$

By making A sufficiently large we see that if $k \leq 9n/10$ then $t \leq 19n/20$ in (12) and consequently

$$\sum_{m=m_0}^{m_1} \sum_{k=n/2}^{9n/10} \Pr_m(\mathcal{B}_k) = o(1). \tag{15}$$

Case 4: $k \geq 9n/10$

$\Pr_p(\mathcal{D}_\ell) \leq$

$$\sum_{s=\ell+1}^{n/2} \binom{n}{s} \binom{n+c}{\ell} \sum_{t=1}^{s(n-s)} \binom{s(n-s)}{t} \left(\frac{\ell p}{n+c}\right)^t (1-p)^{s(n-s)-t} \left(\frac{\ell}{n+c}\right)^{\max\{2\ell-t,0\}}.$$

Let $u(s,\ell,t)$ denote the summand in the above and let $p = \alpha \ln n/n$ and note that $\alpha \in [1-\omega/\ln n, 2+\omega/\ln n]$.

Case 4.1: $t \leq 2\ell$

It will generally be convenient to split s into two ranges:

Case 4.1.1: $s \leq n^{1/10}$

$$\begin{aligned}
u(s,\ell,t) &= \binom{n}{s}\binom{n+c}{\ell}\binom{s(n-s)}{t} p^t (1-p)^{s(n-s)-t} \left(\frac{\ell}{n+c}\right)^{2\ell} \\
&\leq \left(\frac{ne}{s}\right)^s \left(\frac{(n+c)e}{\ell}\right)^\ell \left(\frac{s(n-s)e^{1+p}\alpha \ln n}{tn}\right)^t n^{-\alpha s(n-s)/n} \left(\frac{\ell}{n+c}\right)^{2\ell} \\
&\leq \left(\frac{n^{1-\alpha+\alpha s/n} e}{s}\right)^s \left(\frac{\ell e}{n+c}\right)^\ell \left(\frac{e^2 s(n-s)\ln n}{tn}\right)^t \\
&\leq \left(\frac{n^{1-\alpha+\alpha s/n} e}{s}\right)^s \left(\frac{e^4 s^2 (n-s)^2 (\ln n)^2}{n^3 \ell}\right)^\ell. \qquad (16)
\end{aligned}$$

Now

$$n^{1-\alpha+\alpha s/n} \leq (1+o(1))e^\omega \qquad (17)$$

where $\alpha \geq 1-\omega/\ln n$ and $\omega \to \infty$ slowly.

So if $s \leq 3e^\omega$ then (16) implies that

$$u(s,\ell,t) \leq n^{-(1-o(1))\ell},$$

and if $s > 3e^\omega$

$$\begin{aligned} u(s,\ell,t) &\leq \left(\frac{e^{\omega+5}s(n-s)^2(\ln n)^2}{n^3\ell}\right)^\ell \\ &= O\left(\left(\frac{s}{n^{1-o(1)}}\right)^\ell\right). \end{aligned}$$

Case 4.1.2: $s > n^{1/10}$.

Claim 4 *In a.e. process, every $G_m, m \in [m_0, m_1]$ is such that $\eta(S) \geq \gamma|S|\ln n$ for all $n^{1/10} \leq |S| \leq n/2$, where $\gamma > 0$ is some absolute constant.*

Proof (outline) For $|S| \geq n^{2/3}$ one can use the Chernoff bounds on the tails of the binomial $\eta(S)$. If $|S| \leq n^{2/3}$ we use the fact that with high probability (i) G_{m_0} has $n^{\epsilon'}$ vertices of degree $\leq \epsilon \ln n$ where $\epsilon' = \epsilon'(\epsilon) \to 0$ with ϵ, and (ii) in G_{m_1} no set S of size $\leq n/(\ln n)^2$ contains $3|S|$ edges. □

So if $s \geq n^{1/10}$ then we can take $t \geq \gamma s \ln n > 2\ell$ for some constant $\gamma > 0$ and this case is vacuous.

Case 4.2 : $t > 2\ell$.

$$\begin{aligned} u(s,\ell,t) &\leq \left(\frac{ne}{s}\right)^s \left(\frac{(n+c)e}{\ell}\right)^\ell \left(\frac{s(n-s)e^{1+p}\alpha\ell\ln n}{tn(n+c)}\right)^t n^{-\alpha s(n-s)/n} \\ &= \left(\frac{n^{1-\alpha+\alpha s/n}e}{s}\right)^s \left(\frac{(n+c)e}{\ell}\right)^\ell \left(\frac{s(n-s)e^{1+p}\alpha\ell\ln n}{tn(n+c)}\right)^t \qquad (18) \end{aligned}$$

Case 4.2.1: $t \leq 2n$ and so $((n+c)e/\ell)^\ell \leq (3ne/t)^{t/2}$.

$$u(s,\ell,t) \leq \left(\frac{n^{1-\alpha+\alpha s/n}e}{s}\right)^s \left(\frac{30s\ell\ln n}{t^{3/2}n^{1/2}}\right)^t. \qquad (19)$$

Case 4.2.1.1: $s < n^{1/10}$. Now (17) gives

$$\begin{aligned}\left(\frac{n^{1-\alpha+\alpha s/n}e}{s}\right)^s &\leq \left(\frac{(1+o(1))e^{\omega+1}}{s}\right)^s \\ &\leq e^{(1+o(1))e^\omega} \\ &= e^{\hat{\omega}}, \text{ say,}\end{aligned}$$

and so (19) implies

$$u(s,\ell,t) \leq \left(\frac{s}{n^{\frac{1}{2}-o(1)}}\right)^t. \tag{20}$$

Case 4.2.1.2: $s \geq n^{1/10}$.

Using Claim 4 and (19),

$$u(s,\ell,t) \leq n^{-s/11}\left(\frac{\ell}{\sqrt{tn}}\right)^t.$$

Case 4.2.2: $t \geq 2n$ and so $((n+c)e/\ell)^\ell \leq e^{n+c} \leq e^{(1+\epsilon)t/2}$.

From (19),

$$u(s,\ell,t) \leq \left(\frac{(1+o(1))e^{\omega+1}}{s}\right)^s \left(\frac{30s\ell\ln n}{tn}\right)^t.$$

Case 4.2.2.1: $s < n^{1/10}$.

Arguing as in (20),

$$u(s,\ell,t) \leq \left(\frac{s}{n^{1-o(1)}}\right)^t.$$

Case 4.2.2.2: $s \geq n^{1/10}$.

From Claim 4

$$u(s,\ell,t) \leq \left(\frac{(1+o(1))e^{\omega+1}}{s}\right)^s \left(\frac{A\ell}{n}\right)^t.$$

for some constant $A > 0$. Now this clearly implies

$$u(s, \ell, t) = O(2^{-n}) \tag{21}$$

for $\ell \leq n/(3A)$. For $\ell > n/(3A)$ we have $s \geq \ell$ and

$$u(s, \ell, t) \leq n^{-s/2} A^n$$

and so (21) holds here also.

Summarising,

$$\begin{aligned}
\Pr(\mathcal{D}_\ell) &= O\left(\sum_{t=1}^{2\ell} \sum_{s=\ell+1}^{n^{1/10}} \left(\frac{s}{n^{1-o(1)}}\right)^\ell + \sum_{t=2\ell+1}^{2n} \sum_{s=\ell+1}^{n^{1/10}} \left(\frac{s}{n^{\frac{1}{2}-o(1)}}\right)^t\right. \\
&\quad + \sum_{t=2\ell+1}^{2n} \sum_{s=n^{1/10}}^{n/2} \left(\frac{s}{\sqrt{tn}}\right)^t + \sum_{s=1}^{n^{1/10}} \sum_{t=2n+1}^{s(n-s)} \left(\frac{s}{n^{\frac{1}{2}-o(1)}}\right)^t \\
&\quad \left. + \sum_{s=n^{1/10}}^{n/2} \sum_{t=2n+1}^{s(n-s)} 2^{-n}\right) \\
&= O(\ell n^{-(.9-o(1))\ell}).
\end{aligned}$$

where the double summations correspond to the five cases enumerated above.

Thus, we see that

$$\begin{aligned}
\sum_{m=m_0}^{m_1} \sum_{\ell=2}^{n/10} \Pr_m(\mathcal{D}_\ell) &= O((n \ln n)(\sqrt{n \ln n}) n^{-1.7}) \\
&= o(1). \qquad (22)
\end{aligned}$$

We are thus left with $\Pr\left(\bigcup_{m=m_0}^{m_1} (\mathcal{C}_m \cap \mathcal{A}_{w_m - 1})\right)$.

We consider G_{m_0}. We know that a.e. G_{m_0} consists of a giant connected component C plus $O(e^\omega)$ isolated vertices T. If $\bigcup_{m=m_0}^{m_1} (\mathcal{C}_m \cap \mathcal{A}_{w_m - 1})$ occurs at some time during the process then either

(i) there exist $u, v \in T$ such that the first edges of the process that are incident with each of u and v are the same colour,

OR

(ii) there exists a colour r and a set S, $2 \leq |S| \leq n/2$ such that in G_{m_0} the $t \geq 2$ $(S : \bar{S})$ edges are all of colour r.

(Suppose that deleting the edges of colour r from G_m produces at least three components. If colour k has not occurred by time m_0 then two of these components must be vertices from T, contradicting (i). If G_{m_0} has edges of colour r then deleting these edges must beak C into at least three pieces.)

Clearly

$$\Pr((i)) = o(1) + O(e^{2\omega}/n) = o(1).$$

Furthermore

$$\begin{aligned} \Pr_p((ii)) &\leq \sum_{s=2}^{n/2} \binom{n}{s} n \sum_{t=2}^{s(n-s)} \binom{s(n-s)}{t} \left(\frac{p}{n+c}\right)^t (1-p)^{s(n-s)-t} \\ &\leq 2 \sum_{s=2}^{n/2} \binom{n}{s} n \sum_{t=2}^{10 \ln n} \frac{(s(n-s))^t}{t!} \left(\frac{\alpha \ln n}{n^2}\right)^t n^{-\alpha s} \\ &\leq n \sum_{s=2}^{n/2} \left(\frac{n^{1-\alpha}}{s}\right)^s \sum_{t=2}^{10 \ln n} \left(\frac{s\alpha \ln n}{n}\right)^t \\ &= O(n^{-(1-o(1))}). \end{aligned}$$

The upper bound is good enough to apply (8) and so $\Pr_{m_0}((ii)) = o(1)$. Thus

$$\Pr\left(\bigcup_{m=m_0}^{m_1} (\mathcal{C}_m \cap \mathcal{A}_{w_m-1})\right) = o(1). \tag{23}$$

The result for $|W| \leq (1+\epsilon)n$ follows from (7),(10),(11),(14),(15),(22) and (23).

2.2 Many colours

We now deal with the case where $|W| > (1+\epsilon)n$. Our main tool is a monotonicity result that in essence says "the more colours, the more likely an MST exists". We frame it in a general context. Assume that we are given a fixed collection $X_1, X_2, \ldots, X_M$ of subsets of a finite set X. The elements of X are randomly coloured with s colours. We identify the event

$$\mathcal{E} = \{\exists i, 1 \leq i \leq M : X_i \text{ is multicoloured}\},$$

and let

$$\pi(s) = \mathbf{Pr}(\mathcal{E}) \qquad \text{for } s \geq 1.$$

Theorem 3

$$\pi(s+1) \geq \pi(s).$$

□

We defer the proof of this theorem and show how it can be used to finish the proof of Theorem 1.

When we apply Theorem 3 we have a connected graph G and $X_1, X_2, \ldots, X_M$ is the collection of edge sets of spanning trees of G. The theorem then implies that when we randomly colour such a graph, the more colours we choose from, the more likely we are to produce an MST.

Suppose now that $|W| = s > s_0 = \lceil (1+\epsilon)n \rceil$. Let $\mathbf{Pr}_s$ denote event probabilities when s colours are used. Observe first that

$$\mathbf{Pr}_{s_0}(m_{\mathcal{N}} > m_0) = o(1).$$

Let $\mathcal{G}_m$ denote the set of connected graphs with vertex set $[n]$ and m edges. Then

$$\begin{aligned}
\mathbf{Pr}_s(m_{\mathcal{MT}} > \max\{m_{\mathcal{C}}, m_{\mathcal{N}}\}) &= o(1) + \mathbf{Pr}_s(m_{\mathcal{MT}} > m_{\mathcal{C}} > m_0 > m_{\mathcal{N}}), \\
&\leq o(1) + \sum_{m=m_0+1}^{m_1} \sum_{G \in \mathcal{G}_m} \mathbf{Pr}_s(G = G_{m_{\mathcal{C}}},\ \text{no MST }, m_0 > m_{\mathcal{N}}), \\
&\leq o(1) + \sum_{m=m_0+1}^{m_1} \sum_{G \in \mathcal{G}_m} \mathbf{Pr}(G = G_{m_{\mathcal{C}}}) \mathbf{Pr}_s(G \text{ has no } MST), \\
&\leq o(1) + \sum_{m=m_0+1}^{m_1} \sum_{G \in \mathcal{G}_m} \mathbf{Pr}(G = G_{m_{\mathcal{C}}}) \mathbf{Pr}_{s_0}(G \text{ has no } MST), \\
&\leq o(1) + \mathbf{Pr}_{s_0}(m_0 \leq m_{\mathcal{N}}) \\
&\quad + \sum_{m=m_0+1}^{m_1} \sum_{G \in \mathcal{G}_m} \mathbf{Pr}_{s_0}(G = G_{m_{\mathcal{C}}},\ \text{no MST }, m_0 > m_{\mathcal{N}}), \\
&= o(1) + \mathbf{Pr}_{s_0}(m_{\mathcal{MT}} > \max\{m_{\mathcal{C}}, m_{\mathcal{N}}\}), \\
&= o(1),
\end{aligned}$$

and this completes the proof of Theorem 1.

We now prove Theorem 3. We first generalise the colouring of X to non-uniform colourings i.e. given $p_1 + p_2 + \cdots + p_{s+1} = 1, p_i \geq 0, 1 \leq i \leq s+1$, let

$$\begin{aligned}
\rho(p_1, p_2, \ldots p_{s+1}) &= \mathbf{Pr}(\mathcal{E} \text{ when the elements of } X \text{ are independently} \\
&\qquad \text{coloured } j \text{ with probability } p_j, 1 \leq j \leq s+1).
\end{aligned}$$

Then

$$\pi(X, s) = \rho\left(\frac{1}{s}, \frac{1}{s}, \ldots, \frac{1}{s}, 0\right),$$

and

$$\pi(X, s+1) = \rho\left(\frac{1}{s+1}, \frac{1}{s+1}, \ldots, \frac{1}{s+1}, \frac{1}{s+1}\right).$$

The theorem follows fairly easily from symmetry and

$$\rho(p_1, p_2, \ldots, p_{s+1}) \leq \rho\left(p_1, p_2, \ldots, p_{s-1}, \frac{p_s + p_{s+1}}{2}, \frac{p_s + p_{s+1}}{2}\right). \tag{24}$$

We prove (24) by conditioning on the set of elements $Y \subseteq X$ which are coloured with the first $s-1$ colours and how Y is coloured. Let $Z = X \setminus Y$ and $Z_i = X_i \setminus Y, 1 \leq i \leq M$.

We first eliminate from further consideration those i for which $X_i \cap Y$ is not multicoloured. As for the rest, unless $|Z_i| = 2$,

$$\mathbf{Pr}(X_i \text{ becomes multicoloured } |Y) = 0 \text{ or } 1.$$

We have thus reduced the problem to the case where $|Z_i| = 2$ for all i, and each element is independently coloured s with probability $p = p_s/(p_s + p_{s+1})$ or $s+1$ with probability $1-p$. The Z_i can be thought of as the edges of a graph H, the vertices of which are randomly coloured. There is now a multicoloured X_i if and only if one of the components of H contains two vertices of a different colour, for then, trivially, there is an edge with endpoints of a different colour.

But for a component C with r vertices,

$$\begin{aligned} \mathbf{Pr}(C \text{ is mono-coloured}) &= p^r + (1-p)^r \\ &\geq \left(\tfrac{1}{2}\right)^r + \left(\tfrac{1}{2}\right)^r \end{aligned}$$

and (24) and the theorem follows.

Acknowledgement: We thank Andrzej Ruciński for pointing out an error in an earlier version of the paper, and Noga Alon and Rodney Canfield for their help in proving Theorem 3.

References

[1] B.Bollobás, *Random graphs*, Academic press, 1985.

[2] J.Edmonds, *Submodular functions, matroids and certain polyhedra*, in Combinatorial Structures and their Applications, R.Guy et al, eds., Gordon and Breach, 1970, pp69-87.