Random minimum length spanning trees in regular graphs

Andrew Beveridge
Department of Mathematical Sciences
Carnegie Mellon University

Alan Frieze*
Department of Mathematical Sciences
Carnegie Mellon University

Colin McDiarmid
Department of Statistics
University of Oxford

February 9, 1998

Abstract

Consider a connected r-regular n-vertex graph G with random independent edge lengths, each uniformly distributed on $(0,1)$. Let $mst(G)$ be the expected length of a minimum spanning tree. We show that $mst(G)$ can be estimated quite accurately under two distinct circumstances. Firstly, if r is large and G has a modest edge expansion property then $mst(G) \sim \frac{n}{r}\zeta(3)$, where $\zeta(3) = \sum_{j=1}^{\infty} j^{-3} \sim 1.202$. Secondly, if G has large girth then there exists an explicitly defined constant c_r such that $mst(G) \sim c_r n$. We find in particular that $c_3 = 9/2 - 6\log 2 \sim 0.341$.

1 Introduction

Given a graph $G = (V, E)$ with edge lengths $\mathbf{x} = (x_e : e \in E)$, let $msf(G, \mathbf{x})$ denote the minimum length of a spanning forest. When $\mathbf{X} = (X_e : e \in E)$ is a family of independent random variables, each uniformly distributed on the interval $(0,1)$, denote the expected value $\mathbf{E}(msf(G, \mathbf{X}))$ by $msf(G)$. This quantity gives a measure of the connectivity of G. In the most important case when G is connected, we use mst in place of msf in order to indicate *minimum spanning tree.*

Consider the complete graph K_n and the complete bipartite graph $K_{n,n}$. It is known (see [4, 5]) that, as $n \to \infty$, $mst(K_n) \to \zeta(3)$ and $mst(K_{n,n}) \to 2\zeta(3)$. Here $\zeta(3) = \sum_{j=1}^{\infty} j^{-3} \sim 1.202$. Also, it has recently been shown [12] that, for the d-cube Q_d, which has 2^d nodes and is regular of degree d, we have $(d/2^d) mst(Q_d) \to \zeta(3)$ as $d \to \infty$.

*Supported in part by NSF Grant CCR9530974

The results about mst quoted above (and others from [5]) are for particular regular graphs with growing degrees, and show that mst is about $\zeta(3)$ times the number of nodes divided by the degree. The results below provide a generalisation of all these results about mst. The first result gives a rather general lower bound. Let $\delta(G)$ and $\Delta(G)$ denote the minimum and maximum degree respectively of the graph G.

Theorem 1 *For any n-vertex graph G with no isolated vertices,*

$$msf(G) \geq (1+o(1))(n/\Delta)\zeta(3)$$

where $\Delta = \Delta(G)$ and the $o(1)$ term is with respect to $\Delta \to \infty$. In other words, for any $\epsilon > 0$ there exist Δ_0 such that, for any graph G with no isolated vertices and with $\Delta = \Delta(G) \geq \Delta_0$, we have

$$msf(G) \geq (1-\epsilon)(n/\Delta)\zeta(3).$$

The above result in fact gives the right value for graphs $G = (V, E)$ that are regular or nearly regular and have a modest edge expansion property. For $S \subseteq V$, let $(S : \bar{S})$ be the set of edges with one end in S and the other in $\bar{S} = V \setminus S$.

Theorem 2 *Let $\alpha = \alpha(r) = O(r^{-\frac{1}{3}})$ and let $\rho = \rho(r)$ and $\omega = \omega(r)$ tend to infinity with r. Suppose that the graph $G = (V, E)$ satisfies*

$$r \leq \delta(G) \leq \Delta(G) \leq (1+\alpha)r, \tag{1}$$

and

$$|(S : \bar{S})|/|S| \geq \omega r^{2/3} \log r \quad \text{for all } S \subseteq V \text{ with } r/2 < |S| \leq \min\{\rho r, |V|/2\}. \tag{2}$$

Then

$$msf(G) = (1+o(1))\frac{|V|}{r}\zeta(3)$$

where the $o(1)$ term is with respect to $r \to \infty$.

Note that for $|S| = k$ we have

$$|S : \bar{S}|/|S| \geq \delta - k + 1 \tag{3}$$

and so we are really getting some expansion here for $|S| \leq \min\{\rho r, |V|/2\}$.

For regular graphs we of course take $\alpha = 0$. For K_n, $K_{n,n}$ and Q_d we can define ω, ρ such that the condition (2) holds: when $G = Q_d$ we use the result that

$$|(S : \bar{S})|/|S| \geq d - \log_2 |S|, \tag{4}$$

see for example Bollobás and Leader [3].

There are further similar results. Let $[d]$ denote the set $\{1, \dots, d\}$. Consider the d-dimensional mesh $M_{d,n}^{(1)} = (V_{d,n}, E_{d,n}^{(1)})$, where the vertex set $V_{d,n} = \{0, 1, \dots, n-1\}^d$ and if

$x, y \in V_{d,n}$ then $\{x, y\}$ is in the edge set $E^{(1)}_{d,n}$ if and only if there exists $j \in [d]$ such that $x_i = y_i$ if $i \neq j$ and $x_j - y_j = \pm 1$. Thus $M^{(1)}_{d,n}$ has n^d vertices and has maximum degree $2d$ for $n \geq 3$. We also consider the 'wrap-around' version $M^{(2)}_{d,n} = (V_{d,n}, E^{(2)}_{d,n})$, where if $x, y \in V_{d,n}$ then $\{x, y\} \in E^{(2)}_{d,n}$ if and only if there exists $j \in [n]$ such that $x_i = y_i$ if $i \neq j$ and $x_j - y_j = \pm 1 \bmod n$. Thus $M^{(2)}_{d,n}$ is $2d$-regular for $n \geq 3$. Both $M^{(1)}_{d,2}$ and $M^{(2)}_{d,2}$ are the d-cube Q_d, which is d-regular. The first part of the theorem below is Penrose's result on the d-cube mentioned above.

Theorem 3 *If $d \to \infty$ then*

$$mst(Q_d) \sim \frac{2^d}{d}\zeta(3),$$

$$mst(M^{(2)}_{d,n}) \sim \frac{n^d}{2d}\zeta(3)$$

uniformly over $n \geq 3$, and if also $n \to \infty$ in such a way that $d = o(n)$ then

$$mst(M^{(1)}_{d,n}) \sim \frac{n^d}{2d}\zeta(3).$$

We now move on to discuss the second circumstance under which we can estimate $msf(G)$ quite accurately. Instead of considering graphs with large degrees, we consider r-regular graphs with large girth, or at least with few edges on short cycles. Recall that the girth of a graph G is the length of a shortest cycle in G.

Theorem 4 *For $r \geq 2$ let*

$$c_r = \frac{r}{(r-1)^2} \sum_{k=1}^{\infty} \frac{1}{k(k+\rho)(k+2\rho)},$$

where $\rho = 1/(r-1)$. Then, for any $r \geq 2$ and any r-regular graph G

$$|msf(G) - c_r n| \leq \frac{3n}{2g},$$

where n denotes the number of vertices and g denotes the girth of G. The constants c_r satisfy $c_2 = \frac{1}{2}$, $c_3 = 9/2 - 6\log 2 \sim 0.341$, $c_4 = 9 - 3\log 3 - \pi\sqrt{3} \sim 0.264$, and $c_5 = 15 - 10\log 2 - 5\pi/2 \sim 0.215$; and $c_r \sim \zeta(3)/r$ as $r \to \infty$.

Corollary 5 *For each $r \geq 2$ and $g \geq 3$, there exists $\delta = \delta(r, g) > 0$ with the following property. For every r-regular graph G with n vertices such that there is a set of at most δn edges which hit all cycles of length less than g, we have*

$$|msf(G) - c_r n| \leq \frac{2n}{g}.$$

From this corollary, we obtain easily a result about random regular graphs. Let $G_{n,r}$ denote a random r-regular graph with vertex set $\{1,\ldots,n\}$. Let the random variable $L_{n,r}$ be the minimum length of a spanning forest of the random regular graph $G_{n,r}$ when it has independent edge lengths each uniformly distributed on $(0,1)$. Thus in the notation above we may write $L_{n,r} = msf(G_{n,r}, \mathbf{X})$ and $\mathbf{E}(L_{n,r}) = \mathbf{E}(msf(G_{n,r}))$.

Using the configuration model of random regular graphs see e.g. [2], it can easily be proved that

$$\mathbf{Pr}(G_{n,r} \text{ contains} \geq n^{1/2} \text{ edges on cycles of length} \leq \sqrt{\log n}) \leq n^{-(1/2-o(1))}.$$

We therefore have

Corollary 6 *For each integer $r \geq 3$,*

$$(1/n)\mathbf{E}(L_{n,r}) \to c_r.$$

Remark: Since for $r \geq 3$, $G_{n,r}$ is connected with probability $1 - O(n^{-2})$, this result is not changed if we condition on $G_{n,r}$ being connected.

Further information on the constants c_r is given in Propositions 10 and 11 below. It is straightforward to extend these results to more general distributions on the edge lengths – see [5].

We also prove some results about how concentrated $mst(G, \mathbf{X})$ is about its mean.

Theorem 7 **(a)** *For any r-regular graph $G = (V, E)$ with n vertices and $r = o((n/\log n)^{1/2})$,*

$$\mathbf{Pr}(|mst(G, \mathbf{X}) - mst(G)| \geq \epsilon n/r) \leq e^{-\epsilon^2 n/(5r^2)}$$

if n is sufficiently large.

(b) *There is a constant $K > 0$ such that the following holds. Suppose that*

$$|(S : \bar{S})| \geq \gamma r|S| \quad \text{for all } S \subset V \text{ with } |S| \leq n/2.$$

Then for any $0 < \epsilon \leq 1$,

$$\mathbf{Pr}(|mst(G, \mathbf{X}) - mst(G)| \geq \epsilon n/r) \leq n^2 e^{-K\epsilon^2\gamma^2 n/(\log n)^2},$$

for n sufficiently large.

The following two propositions are easier than Corollary 6, and have short proofs. The first concerns random 2-regular graphs, where we can give a more precise result than for general r.

Proposition 8

$$\mathbf{E}(L_{n,2}) = n/2 - \log n + O(\sqrt{\log n}).$$

Finally, let us consider random graphs $G_{n,p}$ which are not too sparse. Consider any edge-probability $p = p(n)$ which is above the connectivity threshold, that is $P(G_{n,p} \text{ connected}) \to 1$ as $n \to \infty$. (Thus we are assuming that $p(n) = \frac{1}{2}n(\log n + \omega(n))$ where $\omega(n) \to \infty$ as $n \to \infty$.)

Proposition 9 *If $p = p(n)$ is above the threshold for connectivity, then $p\, msf(G_{n,p}) \to \zeta(3)$ as $n \to \infty$, in probability and in any mean.*

2 Proofs

Given a graph $G = (V, E)$ with $|V| = n$ and $0 \le p \le 1$, let G_p be the random subgraph of G with the same vertex set which contains those edges e with $X_e \le p$. [Here we are assuming that as before we have a family $\mathbf{X} = (X_e : e \in E)$ of independent random variables each uniformly distributed on $(0,1)$.] Note that the edges of G are included independently with probability p. In this notation, the usual random graph $G_{n,p}$ could be written as $(K_n)_p$. Let $\kappa(G)$ denote the number of components of G. We shall first give a rather precise description of $msf(G)$.

Lemma 1 *For any graph G,*

$$msf(G) = \int_{p=0}^{1} \mathbf{E}(\kappa(G_p))dp - \kappa(G). \tag{5}$$

Proof We shall follow the proof method in [1] and [7]. Let F denote the random set of edges in the minimal spanning forest. For any $0 \le p \le 1$, $\sum_{e \in F} 1_{(X_e > p)}$ is the number of edges of F which are not in G_p, which equals $\kappa(G_p) - \kappa(G)$. But

$$msf(G, \mathbf{X}) = \sum_{e \in F} X_e = \sum_{e \in F} \int_{p=0}^{1} 1_{(X_e > p)} dp = \int_{p=0}^{1} \sum_{e \in F} 1_{(X_e > p)} dp.$$

Hence

$$msf(G, \mathbf{X}) = \int_{p=0}^{1} \kappa(G_p)dp - \kappa(G),$$

and the result follows on taking expectations. □

2.1 Large Degrees

We substitute $p = x/r$ in (5) to obtain

$$msf(G) = \frac{1}{r} \int_{x=0}^{r} \mathbf{E}(\kappa(G_{x/r}))dx - \kappa(G).$$

Now let $C_{k,x}$ denote the total number of components in $G_{x/r}$ with k vertices. Thus

$$msf(G) = \frac{1}{r} \int_{x=0}^{r} \sum_{k=1}^{n} \mathbf{E}(C_{k,x})dx - \kappa(G). \tag{6}$$

We decompose

$$C_{k,x} = \tau_{k,x} + \sigma_{k,x}$$

where

$\tau_{k,x}$ denotes the number of tree components of $G_{x/r}$ with k vertices

and

$\sigma_{k,x}$ denotes the number of non-tree components in $G_{x/r}$ with k vertices.

We will find, perhaps not unexpectedly, that the number of components of $G_{x/r}$ is usually dominated by the number of components which are small trees. Imagine taking all trees T in G which have k vertices and giving them a root. Fix a vertex $v \in V$ and let $\mathcal{T}(v,k)$ be the set of trees obtained in this way which have root v. Let $t(v,k) = |\mathcal{T}(v,k)|$.

Lemma 2

$$\frac{k^{k-2}(\delta-k)^{k-1}}{(k-1)!} \leq t(v,k) \leq \frac{k^{k-2}\Delta^{k-1}}{(k-1)!}.$$

Proof Given a tree $T \in \mathcal{T}(v,k)$ we label v with k and then define a labelling $f : V(T) \setminus \{v\} \to \{1, \ldots, k-1\}$ of the remaining vertices. Now consider pairs (T, f) where $T \in \mathcal{T}(v,k)$ and f is such a labelling. Clearly each rooted $T \in \mathcal{T}(v,k)$ is in $(k-1)!$ such pairs. Furthermore each such pair defines a unique spanning tree T' of K_k, where (i,j) is an edge of T' if and only if there is an edge $\{x,y\}$ of T such that $f(x) = i$ and $f(y) = j$. Each spanning tree T' of K_k nodes lies in between $(\delta-k)^{k-1}$ and Δ^{k-1} such pairs. Take a fixed breadth first search of T' starting at k and on reaching vertex ℓ for the first time, define $f^{-1}(\ell)$. There will always be between $\delta - k$ and Δ choices. Thus

$$(\delta-k)^{k-1}k^{k-2} \leq \#\text{pairs}\,(T,f) = t(v,k)(k-1)! \leq \Delta^{k-1}k^{k-2}$$

and the lemma follows. □

Now consider a fixed sub-tree T of G containing k vertices. Suppose that the vertices of T induce $a(T)$ edges in G, and the sum of their degrees in G is $b(T)$. Then the probability $\pi(x,T)$ that it forms a component of $G_{x/r}$ satisfies

$$\pi(x,T) = \left(\frac{x}{r}\right)^{k-1}\left(1-\frac{x}{r}\right)^{b(T)-a(T)-k+1}. \tag{7}$$

Also

$$k-1 \leq a(T) \leq \binom{k}{2} \text{ and } k\delta \leq b(T) \leq k\Delta. \tag{8}$$

It follows from Lemma 2, (7) and (8) that

$$\begin{aligned} \mathbf{E}(\tau_{k,x}) &\leq \frac{1}{k}\sum_v t(v,k)\left(\frac{x}{r}\right)^{k-1}\left(1-\frac{x}{r}\right)^{k\delta-(k+2)(k-1)/2} && (9)\\ &\leq \frac{nk^{k-2}}{k!}\left(\frac{\Delta}{r}\right)^{k-1}x^{k-1}\left(1-\frac{x}{r}\right)^{k\delta-k^2}. && (10)\end{aligned}$$

Similarly,

$$\begin{aligned} \mathbf{E}(\tau_{k,x}) &\geq \frac{1}{k}\sum_v t(v,k)\left(\frac{x}{r}\right)^{k-1}\left(1-\frac{x}{r}\right)^{k\Delta-2k+2} && (11)\\ &\geq \frac{nk^{k-2}}{k!}\left(\frac{\delta-k}{r}\right)^{k-1}x^{k-1}\left(1-\frac{x}{r}\right)^{k\Delta}. && (12)\end{aligned}$$

The $1/k$ factor in front of the sums in (9) and (11) comes from the fact that each k-vertex tree appears k times in the sum $\sum_v t(v,k)$. The following will be needed below:

$$\int_{x=0}^{\infty} x^{k-1}e^{-kx}dx = \frac{(k-1)!}{k^k} \geq \frac{1}{ke^k},$$

and for $a \geq 1$

$$\int_{x=a}^{\infty} x^{k-1}e^{-kx}dx \leq \int_{x=a}^{\infty} (xe^{-x})^k dx \leq \int_{x=a}^{\infty} e^{-kx/2}dx = \frac{2}{k}e^{-ka/2}.$$

Now, if $a, b \to \infty$, then

$$\begin{aligned} \int_{x=0}^{a} \sum_{k=1}^{b} \frac{k^{k-3}}{(k-1)!} x^{k-1}e^{-kx}dx &= (1+o(1)) \sum_{k=1}^{b} \frac{1}{k^3} \\ &= (1+o(1))\zeta(3). \end{aligned} \tag{13}$$

We may now prove Theorem 1: after that we shall continue the development here to prove Theorem 2.

Proof of Theorem 1 We use four stages.

(a) Let $\epsilon > 0$. Let a and b be sufficiently large that

$$\int_{x=0}^{a} \sum_{k=1}^{b} \frac{k^{k-3}}{(k-1)!} x^{k-1}e^{-kx}dx \geq (1-\epsilon)\zeta(3).$$

Now, if $0 \leq x \leq r/2$ and $0 \leq \alpha \leq 1/2$, then

$$(1-x/r)^{kr(1+\alpha)} \geq \exp\left(-k(1+\alpha)(x+2x^2/r)\right) \geq e^{-kx}\exp(-xk\alpha - 3x^2k/r).$$

Let r_0 be sufficiently large that for $r \geq r_0$ we have $(1-b/r)^{b-1} \geq (1-\epsilon)$ and $\exp(-3a^2b/r) \geq (1-\epsilon)$. Let $0 < \eta < 1/2$ be sufficiently small that $\exp(-ab\eta) \geq (1-\epsilon)$.

Now suppose that $r \geq r_0$, that the graph G has $\delta = \delta(G) = r$, and that $\Delta = \Delta(G) \leq (1+\eta)r$. Then by (12) and the above, for $0 \leq x \leq a$ and $1 \leq k \leq b$,

$$\mathbf{E}(\tau_{k,x}) \geq \frac{n}{k}\frac{k^{k-2}}{(k-1)!}\left(1-\frac{k}{r}\right)^{k-1} x^{k-1}\left(1-\frac{x}{r}\right)^{k\Delta} \geq \frac{n}{k}\frac{k^{k-2}}{(k-1)!}x^{k-1}e^{-kx}(1-\epsilon)^3.$$

Hence $msf(G) \geq (1-\epsilon)^4 \frac{n}{r}\zeta(3)$.

(b) Next we drop the assumption on $\delta(G)$. Let $\epsilon > 0$. We shall show that there exist r_1 and $\beta > 0$ such that, for any connected n-vertex graph G with $r_1 \leq \Delta = \Delta(G) \leq \beta n$, we have

$$mst(G) \geq (1-\epsilon)\frac{n}{\Delta}\zeta(3).$$

To do this, let r_0 and $\eta > 0$ be such that for any $r \geq r_0$ and any graph G with $\delta = \delta(G) = r$ and $\Delta = \Delta(G) \leq (1+\eta)r$, we have $msf(G) \geq (1-\epsilon)\zeta(3)$. We have just seen that this is possible. Let $r_1 = \max\{r_0, 2/\eta\}$ and $\beta > 0$ be such that if $r_1 \le r \le \beta n$ then $r + \frac{r^2}{n-r} + 1 \le (1+\eta)r$.

Now let G be a connected n-vertex graph with $r_1 \le r = \Delta(G) \le \beta n$. We shall add edges to G to produce a graph G' which has minimum degree r and maximum degree $\Delta' \le (1+\eta)r$: then

$$mst(G) \ge mst(G') \ge (1-\epsilon)\frac{n}{\Delta'}\zeta(3),$$

and the desired result follows. To get G' we add edges between vertices of degree less than r until the vertices S of degree less than r form a clique. We then add new edges from S to $\bar{S} = V \setminus S$ until the vertices in S have degree r. When adding an $(S : \bar{S})$ edge we choose a vertex of current smallest degree in $\bar{S}$. In this way we end up with $\delta(G') = r$ and

$$\Delta' \le r + \frac{r^2}{n-r} + 1 \le (1+\eta)r,$$

as required.

(c) Next we shall deduce the corresponding result for connected graphs but without the condition that $\Delta \le \beta n$.

Let $\epsilon > 0$. Choose r_1 and $\beta > 0$ as above for $\epsilon/3$. Let r_2 be the maximum of r_1 and $\lceil 6/\epsilon \rceil$. Consider a connected n-vertex graph G with $\Delta = \Delta(G) \ge r_2$. Let $k = \lceil (2/\beta) \rceil$, and form k disjoint copies $G_1, \ldots, G_k$ of G. For each $i = 1, \ldots, k-1$ add a perfect matching between G_i and G_{i+1}. The new graph H is connected, and has kn vertices and maximum degree $\Delta + 2$, and thus satisfies $\Delta(H) \le 2n \le \beta|V(H)|$. Hence

$$mst(H) \ge (1-\epsilon/3)(kn/(\Delta+2))\zeta(3) \ge (1-2\epsilon/3)(kn/\Delta)\zeta(3),$$

since $2/(\Delta+2) < 2/r_1 \le \epsilon/3$. But $mst(H) \le k\ mst(G) + (k-1)/(n+1)$, and so

$$mst(G) \ge (1/k)mst(H) - 1/n \ge (1-2\epsilon/3)(n/\Delta)\zeta(3) - 1/n \ge (1-\epsilon)(n/\Delta)\zeta(3),$$

for $n \ge 3/\epsilon$.

(d) Finally we remove the assumption of connectedness. Let c be the infimum of $mst(K_n)$ over all positive integers n. Then $c > 0$ - indeed it is easy to see that $c \ge 1/2$. Let $\epsilon > 0$. Let r_2 be as above, and let r_3 be the maximum of r_2 and $\lceil \zeta(3)r_2/c \rceil$. Consider a graph G with $\Delta = \Delta(G) \ge r_3$. List the components of G as $G_1, \ldots, G_k$ where $G_i = (V_i, E_i)$. If $|V_i| < r_2$ then

$$mst(G_i) \ge c \ge r_2\zeta(3)/r_3 \ge |V_i|\zeta(3)/\Delta(G),$$

and if $|V_i| \ge r_2$ then

$$mst(G_i) \ge (1-\epsilon)|V_i|\zeta(3)/\Delta(G).$$

Hence

$$mst(G) = \sum_{i=1}^{k} mst(G_i) \ge (1-\epsilon)(\sum_{i=1}^{k}|V_i|)\zeta(3)/\Delta(G) = (1-\epsilon)|V(G)|\zeta(3)/\Delta(G),$$

as required. This completes the proof of Theorem 1. $\square$

Proof of Theorem 2

In order to use (6) we need to consider a number of separate ranges for x and k. Let $A = 2r^{1/3}/\omega$, $B = \lfloor (Ar)^{1/4} \rfloor$ so that each of $B\alpha$, AB^2/r and $A/B \to 0$ as $r \to \infty$.

Range 1: $0 \le x \le A$ and $1 \le k \le B$. By (10) we have

$$\mathbf{E}(\tau_{k,x}) \le \frac{nk^{k-2}}{k!} x^{k-1} e^{-kx} \exp(k\alpha + xk^2/r),$$

since $(\Delta/r)^{k-1} \le (1+\alpha)^k \le \exp(k\alpha)$, and $(1-x/r)^{k\delta - k^2} \le \exp(-xk + xk^2/r)$. Also,

$$\exp(k\alpha + xk^2/r) \le \exp(B\alpha + AB^2/r) = 1 + o(1).$$

Hence

$$\begin{aligned} \frac{1}{r}\int_{x=0}^{A} \sum_{k=1}^{B} \mathbf{E}(\tau_{k,x})dx &\le (1+o(1))\frac{n}{r}\int_{x=0}^{A}\sum_{k=1}^{B}\frac{k^{k-2}}{k!}x^{k-1}e^{-kx}dx \\ &\le (1+o(1))\frac{n}{r}\zeta(3). \end{aligned} \tag{14}$$

Let $\sigma_{k,u,x}$ be the number of non-tree components of $G_{x/r}$ which have k vertices and $k-1+u$ edges. Then

$$\mathbf{E}(\sigma_{k,u,x}) \le \frac{1}{k}\sum_{v\in V} t(v,k)\binom{k}{2}^u \left(\frac{x}{r}\right)^{k-1+u}\left(1-\frac{x}{r}\right)^{kr-k^2}.$$

So

$$\begin{aligned} \mathbf{E}(\sigma_{k,x}) &\le \frac{nk^{k-2}}{k!}\Delta^{k-1}\sum_{u=1}^{\infty}\left(\frac{k^2}{2}\right)^u\left(\frac{x}{r}\right)^{k-1+u}\left(1-\frac{x}{r}\right)^{kr-k^2} \\ &\le \frac{nk^{k-2}}{k!}\left(\frac{\Delta}{r}\right)^{k-1} x^{k-1}e^{-xk}e^{xk^2/r}\sum_{u=1}^{\infty}\left(\frac{k^2x}{2r}\right)^u \\ &\le \left(\frac{e^{k\alpha + xk^2/r}}{2 - xk^2/r}\right)\frac{n}{r}\frac{k^k}{k!}x^k e^{-kx} \\ &\le \frac{n}{r}\frac{k^k}{k!}x^k e^{-kx} \end{aligned}$$

if r is sufficiently large. Thus,

$$\begin{aligned} \frac{1}{r}\int_{x=0}^{A}\sum_{k=1}^{B}\mathbf{E}(\sigma_{k,x}) &\le \frac{n}{r^2}\sum_{k=1}^{B}\frac{k^k}{k!}\int_{x=0}^{\infty}x^k e^{-kx}dx \\ &= \frac{n}{r^2}\sum_{k=1}^{B}\frac{1}{k^2} \\ &\le 2\frac{n}{r^2} = o(n/r). \end{aligned} \tag{15}$$

Range 2: $x \le A$ and $k \ge B$. Using the bound

$$\sum_{k=\ell}^{n} C_{k,x} \le \frac{n}{\ell} \tag{16}$$

for all ℓ, x we get

$$\frac{1}{r}\int_{x=0}^{A} \sum_{k=B}^{n} \mathbf{E}(C_{k,x})dx \le \frac{1}{r}\int_{x=0}^{A} \frac{n}{B}dx = \frac{A}{B}\cdot\frac{n}{r} = o(n/r). \tag{17}$$

We next have to consider larger values of x in our integral. Now G contains at most $n(e\Delta)^k$ connected subgraphs with k vertices. To see this, choose $v \in V$ and note that G contains fewer than $(e\Delta)^k$ k-vertex trees rooted at v. This follows from the formula (29) below for the number of subtrees of an infinite rooted r-ary tree which contain the root.

Also, from (3) we get $S \subseteq V$, $|S| = k$ implies $|S : \bar{S}| \ge k\delta - k(k-1) \ge k(r-k)$. Thus

$$\begin{aligned} \mathbf{E}(C_{k,x}) &\le n(e\Delta)^k \left(1 - \frac{x}{r}\right)^{k(r-k)} \\ &\le n(re^{1+\alpha - x(1-k/r)})^k. \end{aligned} \tag{18}$$

Range 3: $x \ge A$ and $k \le r/2$. Equation (18) implies that for large r,

$$\mathbf{E}(C_{k,x}) \le ne^{-kA/3}. \tag{19}$$

Thus

$$\frac{1}{r}\int_{x=A}^{r} \sum_{k=1}^{r/2} \mathbf{E}(C_{k,x})dx \le nre^{-A/3} = o(n/r). \tag{20}$$

Range 4: $x \ge A$ and $r/2 < k \le k_0 = \min\{\rho r, n/2\}$. It is only here that we use the expansion condition (2). We find

$$\mathbf{E}(C_{k,x}) \le n(er)^k \left(1 - \frac{x}{r}\right)^{k\omega r^{2/3}\log r} \le n\left(\frac{e}{r}\right)^k. \tag{21}$$

So,

$$\frac{1}{r}\int_{x=A}^{r} \sum_{k=r/2+1}^{k_0} \mathbf{E}(C_{k,x})dx \le n\left(\frac{e}{r}\right)^{r/2} = o(n/r). \tag{22}$$

We split the remaining range into two cases.

Range 5: $x \ge A$ and $k > k_0$.

Case 1: $n \geq 2\rho r$, so that $k_0 = \rho r$.
If $k \geq k_0$ we use (16) to deduce that

$$\frac{1}{r}\int_{x=A}^{r}\sum_{k=\rho r}^{n}\mathbf{E}(C_{k,x})dx \leq \frac{n}{\rho r} = o(n/r). \tag{23}$$

Part (b) now follows from (6), (14), (15), (17), (20), (22) and (23).
Case 2: $n < 2\rho r$, so that $k_0 = n/2$.
For larger r, we have to use the $-\kappa(G)$ term in (6), ignored in the previous case. Here (2) implies $\kappa(G) = 1$. We deduce from (19) and (21) that

$$\mathbf{Pr}(G_{A/r} \text{ is not connected }) \leq 2ne^{-A/3} + 2n\left(\frac{e}{r}\right)^{r/2}. \tag{24}$$

Then,

$$\frac{1}{r}\int_{x=A}^{r}\sum_{k=0}^{n}\mathbf{E}(C_{k,x})dx = 1 - O(n^{-K})$$

for any constant $K > 0$, and the proof is completed by (6), (14), (15), (17). □

Remark: It is worth pointing out that it is not enough to have $r \to \infty$ in order to have Theorem 2, that is, we need some extra condition such as the expansion condition (2). For consider the graph G_0 obtained from n/r r-cliques $C_1, C_2, \ldots, C_{n/r}$ by deleting an edge (x_i, y_i) from C_i, $1 \leq i \leq n/r$ then joining the cliques into a cycle of cliques by adding edges (y_i, x_{i+1}) for $1 \leq i \leq n/r$. It is not hard to see that

$$mst(G_0) \sim \frac{n}{r}\left(\zeta(3) + \frac{1}{2}\right)$$

if $r \to \infty$ with $r = o(n)$. We conjecture that this is the worst-case, that is

Conjecture: Assuming only the conditions of Theorem 1,

$$mst(G) \leq (1 + o(1))\frac{n}{r}\left(\zeta(3) + \frac{1}{2}\right).$$

2.1.1 Proof of Theorem 3

We consider $M_{d,n}^{(2)}$ first. We prove the equivalent of (4). For this we need a technical lemma.

Lemma 3 *Assume $s_1, s_2, \ldots, s_n \geq 0$ and $s = s_1 + s_2 + \cdots + s_n$ then*

$$\frac{1}{2}s\log_2 s \geq \frac{1}{2}\sum_{i=1}^{n} s_i \log_2 s_i + \sum_{i=1}^{n}\min\{s_i, s_{i+1}\}. \tag{25}$$

(Here $s_{n+1} = s_1$ and $s_i \log_2 s_i = 0$ when $s_i = 0$.)

Proof We prove (25) by induction on n. The case $n=2$ is proved in [3]. Assume (25) is true for some $n \geq 2$ and consider $n+1$.

$$\begin{aligned}
\Lambda &= \frac{1}{2}\sum_{i=1}^{n+1} s_i \log_2 s_i + \sum_{i=1}^{n+1} \min\{s_i, s_{i+1}\} \\
&\leq \frac{1}{2}(s - s_{n+1})\log_2(s - s_{n+1}) + \frac{1}{2}s_{n+1}\log_2 s_{n+1} + \min\{s_n, s_{n+1}\} + \min\{s_{n+1}, s_1\} - \min\{s_n, s_1\}
\end{aligned}$$

by induction.

Case 1 $\min\{s_1, s_n, s_{n+1}\} = s_1$:

$$\begin{aligned}
\Lambda &\leq \frac{1}{2}(s - s_{n+1})\log_2(s - s_{n+1}) + \frac{1}{2}s_{n+1}\log_2 s_{n+1} + \min\{s_n, s_{n+1}\} \\
&\leq \frac{1}{2}(s - s_{n+1})\log_2(s - s_{n+1}) + \frac{1}{2}s_{n+1}\log_2 s_{n+1} + \min\{s - s_{n+1}, s_{n+1}\} \\
&\leq \frac{1}{2}s\log_2 s.
\end{aligned}$$

Case 2 $\min\{s_1, s_n, s_{n+1}\} = s_n$: similar.

Case 3 $\min\{s_1, s_n, s_{n+1}\} = s_{n+1}$:

$$\begin{aligned}
\Lambda &\leq \frac{1}{2}(s - s_{n+1})\log_2(s - s_{n+1}) + \frac{1}{2}s_{n+1}\log_2 s_{n+1} + 2s_{n+1} - \min\{s_n, s_1\} \\
&\leq \frac{1}{2}(s - s_{n+1})\log_2(s - s_{n+1}) + \frac{1}{2}s_{n+1}\log_2 s_{n+1} + s_{n+1} \\
&= \frac{1}{2}(s - s_{n+1})\log_2(s - s_{n+1}) + \frac{1}{2}s_{n+1}\log_2 s_{n+1} + \min\{s - s_{n+1}, s_{n+1}\} \\
&\leq \frac{1}{2}s\log_2 s.
\end{aligned}$$

□

Now consider $S \subseteq V_{d,n}$ with $|S| = s$. We now prove by induction on s that

$$S \text{ contains at most } \tfrac{1}{2}s\log_2 s \text{ edges.} \tag{26}$$

Let S_i be the set of vertices $x \in S$ with $x_n = i$. Let $s_i = |S_i|$, $i = 1, 2, \ldots, n$. Each S_i can be considered a subset of $V_{d,n-1}$ and we can assume inductively that each S_i contains at most $\frac{1}{2}s_i\log_2 s_i$ edges. Therefore S contains at most Λ edges and (26) follows from Lemma 3. It follows that $|S : \bar{S}| \geq 2ds - s\log_2 s$ and so $M_{d,n}^{(2)}$ has adequate expansion to apply Theorem 2.

Now consider the spanning subgraph $M_{d,n}^{(1)}$ of $M_{d,n}^{(2)}$. Since each edge of $M_{d,n}^{(2)}$ is equally likely to be in a minimum spanning tree T, the expected number of 'wrap-around' edges in T equals $(n^d - 1)/n < n^{d-1}$. Hence

$$mst(M_{d,n}^{(2)}) \leq mst(M_{d,n}^{(1)}) \leq mst(M_{d,n}^{(2)}) + n^{d-1},$$

which completes the proof. □

2.2 Large Girth

We note first that all components of G_p with fewer than g vertices are trees. Here g denotes the girth of G. Hence

$$\left| mst(G) - \int_{p=0}^{1} \sum_{k=1}^{g-1} \mathbf{E}(\tau_{k,p}) dp \right| \leq \frac{n}{g}. \tag{27}$$

Here $\tau_{k,p}$ is the number of (tree) components with k vertices in G_p and n/g is an upper bound for the number of components of G_p with g or more vertices.

Let $t(v,k)$ be as in Lemma 2. This time we have an exact formula for $t(v,k)$ when k is less than the girth g of G.

Lemma 4 *For* $k < g$,

$$t(v,k) = \frac{r((r-1)k)!}{(k-1)!((r-2)k+2)!}.$$

Proof We use the formula

$$t(v,k) = \sum_{i=1}^{k} \frac{1}{(r-2)i+1} \binom{(r-1)i}{i} \frac{1}{(r-2)(k-i)+1} \binom{(r-1)(k-i)}{k-i}. \tag{28}$$

This follows from the formula

$$\frac{1}{(r-1)m+1} \binom{rm}{m} \tag{29}$$

for the number of m-vertex subtrees of an infinite rooted r-ary tree which contain the root – see Knuth [8], Problem 2.3.4.4.11. To obtain (28) we take each tree with k vertices rooted at v and view it as an $(r-1)$-ary tree with i vertices rooted at v plus an $(r-1)$-ary tree with $k-i$ vertices rooted at the largest (numbered) neighbour of v. Let

$$a_k = \sum_{i=0}^{k} \frac{1}{(r-2)i+1} \binom{(r-1)i}{i} \frac{1}{(r-2)(k-i)+1} \binom{(r-1)(k-i)}{k-i}.$$

[Sum from $i=0$ as opposed to $i=1$ in (28).] Then

$$\begin{aligned}
\sum_{k=0}^{\infty} a_k x^k &= \sum_{k=0}^{\infty} \sum_{i=0}^{k} \frac{1}{(r-2)i+1} \binom{(r-1)i}{i} \frac{1}{(r-2)(k-i)+1} \binom{(r-1)(k-i)}{k-i} x^k \\
&= \sum_{i=0}^{\infty} \frac{1}{(r-2)i+1} \binom{(r-1)i}{i} x^i \sum_{k=i}^{\infty} \frac{1}{(r-2)(k-i)+1} \binom{(r-1)(k-i)}{k-i} x^{k-i} \\
&= \left(\sum_{i=0}^{\infty} \frac{1}{(r-2)i+1} \binom{(r-1)i}{i} x^i \right)^2 \\
&= \left(\sum_{i=0}^{\infty} \frac{1}{(r-1)i+1} \binom{(r-1)i+1}{i} x^i \right)^2 \\
&= B_{r-1}(x)^2,
\end{aligned}$$

where

$$B_t(x) = \sum_{i=0}^{\infty} \frac{1}{ti+1} \binom{ti+1}{i} x^i$$

is the *Generalised Binomial Series.* The identity

$$B_t(x)^s = \sum_{i=0}^{\infty} \frac{s}{ti+s} \binom{ti+s}{i} x^i$$

is given for example in Graham, Knuth and Patashnik [6]. Thus,

$$a_k = \frac{2}{(r-1)k+2} \binom{(r-1)k+2}{k}.$$

The lemma follows from

$$t(v,k) = a_k - \frac{1}{(r-2)k+1} \binom{(r-1)k}{k}.$$

□

We may now prove the first part of Theorem 4. We have

$$\begin{aligned}
& \int_{p=0}^{1} \sum_{k=1}^{g-1} \mathbf{E}(\tau_{k,p}) dp \qquad (30) \\
= \; & \frac{1}{k} \int_{p=0}^{1} \sum_{k=1}^{g-1} \sum_{v \in V} t(v,k) p^{k-1} (1-p)^{rk-2k+2} dp \\
= \; & \sum_{k=1}^{g-1} \frac{n}{k} \frac{r((r-1)k)!}{(k-1)!((r-2)k+2)!} \frac{(k-1)!((r-2)k+2)!}{((r-1)k+2)!} \\
= \; & \sum_{k=1}^{g-1} \frac{nr}{k((r-1)k+1)((r-1)k+2)} \\
= \; & \frac{nr}{(r-1)^2} \sum_{k=1}^{g-1} \frac{1}{k(k+\rho)(k+2\rho)}
\end{aligned}$$

where $\rho = 1/(r-1)$. Theorem 4 now follows from (27) and

$$\begin{aligned}
\frac{r}{(r-1)^2} \sum_{k=g}^{\infty} \frac{1}{k(k+\rho)(k+2\rho)} & \leq \frac{r}{(r-1)^2} \sum_{k=g}^{\infty} k^{-3} \\
& \leq \frac{r}{(r-1)^2} \int_{g-1}^{\infty} x^{-3} dx \\
& = \frac{r}{(r-1)^2} \frac{1}{2(g-1)^2} \\
& \leq \frac{1}{2g}.
\end{aligned}$$

□

Proof of Corollary 5 Start with a 2-edge-connected r-regular graph with girth at least $g-2$, and form a new graph H by 'splitting' an edge so that two vertices have degree 1 and all the others have degree r.

Let F be a set of edges in G which meet each cycle of length less than g. From the graph G, form a new graph $\hat{G}$ as follows. For each edge $f=\{u,v\} \in F$, take a new copy H_f of H and identify the vertices u and v with the vertices of degree 1 in H_f. Then $\hat{G}$ has girth at least g, $|V(\hat{G})| = n + |F|(|V(H)|-2) = (1+o(1))n$, and $|msf(\hat{G}) - msf(G)| \leq |F||E(H)| = o(n)$. □

2.2.1 Proof of Theorem 7

Our main tool here is a concentration inequality of Talagrand [14], see Steele [13] for a good exposition. Let A be a (measurable) non-empty subset of $\boldsymbol{R}^E$. For $\mathbf{x}, \beta \in \boldsymbol{R}^E$ with $||\beta||_2 = 1$ let

$$d_A(\mathbf{x},\beta) = \inf_{\mathbf{y}\in A} \sum_{e\in E} \beta_e 1_{\{\mathbf{x}_e \neq \mathbf{y}_e\}}. \tag{31}$$

and let

$$d_A(\mathbf{x}) = \sup_{\beta} d_A(\mathbf{x},\beta).$$

Talagrand shows that for all $t>0$,

$$\mathbf{Pr}(\mathbf{X}\in A)\mathbf{Pr}(d_A(\mathbf{X}) \geq t) \leq e^{-t^2/4}. \tag{32}$$

(a) For $a \in \boldsymbol{R}$ let

$$S(a) = \{\mathbf{y} \in \boldsymbol{R}^E :\ mst(G,\mathbf{y}) \leq a\}.$$

Given $\mathbf{x}$ we let $T=T(\mathbf{x})$ be a minimum spanning tree of G using these weights ($T(\mathbf{X})$ is unique with probability 1). Let $L = L(\mathbf{x}) = (\sum_{e\in T} \mathbf{x}_e^2)^{1/2}$. Note that $L(\mathbf{x}) \leq n^{1/2}$. Define, $\beta = \beta(\mathbf{x})$ by

$$\beta_e = \begin{cases} \mathbf{x}_e/L: & e\in T \\ 0: & \text{otherwise} \end{cases}$$

Then for $\mathbf{y}\in S(a)$ we have

$$\begin{aligned} mst(G,\mathbf{x}) &\leq mst(G,\mathbf{y}) + \sum_{e\in T(\mathbf{x})} (\mathbf{x}_e - \mathbf{y}_e)^+ \\ &\leq mst(G,\mathbf{y}) + L(\mathbf{x}) \sum_{e\in E} \beta_e 1_{\{\mathbf{x}_e \neq \mathbf{y}_e\}}. \end{aligned}$$

By choosing $\mathbf{y}$ achieving the minimum in (31) (the infimum is achieved) we see that

$$mst(G,\mathbf{x}) \leq a + L(\mathbf{x}) d_a(\mathbf{x},\beta) \leq a + n^{1/2} d_a(\mathbf{x},\beta).$$

Applying (32) with $A = S(a)$ we get

$$\mathbf{Pr}(mst(G,\mathbf{X}) \leq a)\mathbf{Pr}(mst(G,\mathbf{X}) \geq a + n^{1/2}t) \leq e^{-t^2/4}. \tag{33}$$

Let M denote the median of $mst(G,\mathbf{X})$. Then with $a = M$ and $t = \epsilon n^{1/2}/r$,

$$\mathbf{Pr}(mst(G,\mathbf{X}) \geq M + \epsilon n/r) \leq 2e^{-\epsilon^2 n/(4r^2)}. \tag{34}$$

With $a = M - \epsilon n/r$,

$$\mathbf{Pr}(mst(G,\mathbf{X}) \leq M - \epsilon n/r) \leq 2e^{-\epsilon^2 n/(4r^2)}. \tag{35}$$

Equations (34) and (35) plus $r = o((n/\log n)^{1/2})$ imply that

$$|M - mst(G)| = o(n/r)$$

and so it is a simple matter to replace M by $mst(G)$ in (34), (35) to obtain (a).

(b) We change the definition of β slightly. For minimum spanning tree $T(\mathbf{x})$ we let $T_1(\mathbf{x}) = \{e \in T: \ \mathbf{x}_e \leq 12\log n/(\gamma r)\}$. Then let

$$L_1(\mathbf{x}) = \left(\sum_{e\in T_1} \mathbf{x}_e^2\right)^{1/2} \leq \frac{12n^{1/2}\log n}{\gamma r}.$$

Then define

$$\beta_e = \begin{cases} \mathbf{x}_e/L_1: & e \in T_1 \\ 0: & \text{otherwise} \end{cases}$$

Also let

$$\phi(\mathbf{x}) = \sum_{e\in T\setminus T_1} \mathbf{x}_e.$$

Then for $\mathbf{y} \in S(a)$ we have

$$\begin{aligned} mst(G,\mathbf{x}) &\leq mst(G,\mathbf{y}) + \sum_{e\in T_1} (\mathbf{x}_e - \mathbf{y}_e)^+ + \phi(\mathbf{x}) \\ &\leq mst(G,\mathbf{y}) + L_1(\mathbf{x}) \sum_{e\in E} \beta_e 1_{\{\mathbf{x}_e \neq \mathbf{y}_e\}} + \phi(\mathbf{x}). \end{aligned}$$

By choosing $\mathbf{y}$ achieving the minimum in (31) we see that

$$mst(G,\mathbf{x}) \leq a + L_1(\mathbf{x}) d_a(\mathbf{x},\beta) + \phi(\mathbf{x}).$$

Applying (32) we get

$$\mathbf{Pr}(mst(G,\mathbf{X}) \leq a)\mathbf{Pr}(mst(G,\mathbf{X}) \geq a + t\frac{12n^{1/2}\log n}{\gamma r} + \phi(\mathbf{X})) \leq e^{-t^2/4}. \tag{36}$$

We will show below that

$$\mathbf{Pr}(\phi(\mathbf{X}) \geq \epsilon n/(3r)) \leq e^{-\gamma n/(20(\log n)^2)}. \tag{37}$$

So putting $a = M$ and $t = \epsilon\gamma n^{1/2}/(36\log n)$ into (36) we get

$$\mathbf{Pr}(mst(G,\mathbf{X}) \geq M + 2\epsilon n/(3r)) \leq 2e^{-\epsilon^2\gamma^2 n/(5184(\log n)^2)} + \mathbf{Pr}(\phi(\mathbf{X}) \geq \epsilon n/(3r)).$$

On the other hand, putting $a = M - 2\epsilon n/(3r)$ and $t = \epsilon\gamma n^{1/2}/(36\log n)$ we get

$$\mathbf{Pr}(mst(G,\mathbf{X}) \leq M - 2\epsilon n/(3r))\mathbf{Pr}(mst(G,\mathbf{X}) \geq M - \epsilon n/(3r) + \phi(\mathbf{X})) \leq e^{-t^2/4}.$$

But

$$\mathbf{Pr}(mst(G,\mathbf{X}) \geq M - \epsilon n/(3r) + \phi(\mathbf{X})) \geq \frac{1}{2} - \mathbf{Pr}(\phi(\mathbf{X}) \geq \epsilon n/(3r))$$

and we can finish as in (a).

Proof of (37) Let

$$\begin{aligned}
\pi(m,k,p) &= \mathbf{Pr}(G_p \text{ contains } \geq m \text{ components of size } k) \\
&\leq \binom{n}{k}^m (1-p)^{\gamma krm/2} \\
&\leq \left(\frac{ne}{k}e^{-p\gamma r/2}\right)^{mk} \\
&\leq e^{-mkp\gamma r/3}
\end{aligned}$$

if $p \geq p_0 = \min\{1, 12\log n/(\gamma r)\}$. Next let

$$p_i = \min\{1, 2^i p_0\} \text{ for } 0 \leq i \leq i_0 = \lceil \log_2 p_0^{-1} \rceil$$

and

$$m_{k,p} = \frac{\epsilon n}{6kpr(\log n)^2}.$$

Now

$$\phi(\mathbf{X}) \leq \sum_{i=0}^{i_0-1}\sum_{k=1}^{n} C_{k,p_i}p_{i+1}$$

and so if

$$G_{p_i} \text{ contains } < m_{k,p_i} \text{ components of size } k \text{ for } 0 \leq i < i_0,\ 1 \leq k \leq n \tag{38}$$

then

$$\phi(\mathbf{X}) \leq \sum_{i=0}^{i_0-1}\sum_{k=1}^{n} \frac{\epsilon n}{3kr(\log n)^2} \leq \frac{\epsilon n}{3r}.$$

Furthermore, the probability that (38) fails to hold is at most

$$\sum_{i=0}^{i_0-1}\sum_{k=1}^{n} \pi(m_{k,p_i}, k, p_i) \leq \sum_{i=0}^{i_0-1}\sum_{k=1}^{n} e^{-\epsilon\gamma n/(18(\log n)^2)}$$

which proves (37). □

We now consider the values of the constants c_r more carefully.

Proposition 10 *The constants* c_r *satisfy* $c_2 = 1/2$, $c_3 = 9/2 - 6\log 2 \sim 0.341$, $c_4 = 9 - 3\log 3 - \pi\sqrt{3} \sim 0.264$ *and* $c_5 = 15 - 10\log 2 - 5\pi/2 \sim 0.215$*; and in general, for* $r \geq 3$, $c_r = r\sum_{j=0}^{r-2} g(\omega^j)$, *where* $g(x) = \frac{(x-1)^2}{2x^2}\log(\frac{1}{1-x}) + \frac{3}{4}$ *and* $\omega = e^{2\pi i/(r-1)}$.

Proof Let Σ_r denote the sum in Theorem 4, so that $c_r = r\Sigma_r$. Note first that

$$\begin{aligned}
&\sum_{k=1}^{\infty} \frac{x^k}{k(k+1)(k+2)} \\
&= \sum_{k=1}^{\infty} x^k \left(\frac{1}{2k} - \frac{1}{k+1} + \frac{1}{2(k+2)}\right) \\
&= \frac{1}{2}\log\left(\frac{1}{1-x}\right) - \frac{1}{x}\left(\log\left(\frac{1}{1-x}\right) - x\right) + \frac{1}{2x^2}\left(\log\left(\frac{1}{1-x}\right) - x - \frac{x^2}{2}\right) \\
&= \frac{(x-1)^2}{2x^2}\log\left(\frac{1}{1-x}\right) + \frac{3}{4} - \frac{1}{2x}.
\end{aligned}$$

Thus $\Sigma_2 = \frac{1}{4}$. Also, for $r \geq 3$, note that $\omega^{r-1} = 1$ and $1 + \omega + \cdots + \omega^{r-2} = 0$. Hence, for $r \geq 3$

$$\Sigma_r = (r-1)\sum_{k:(r-1)|k} \frac{1}{k(k+1)(k+2)} = \sum_{j=0}^{r-2} g(\omega^j).$$

For $r = 3$, $\omega = -1$ so

$$\Sigma_3 = g(1) + g(-1) = \frac{3}{2} - 2\log 2,$$

and thus c_3 is as given. For $r = 4$ we find after some calculation that

$$Re(g(\omega)) = \frac{3}{4} - \frac{3\log 3}{8} - \frac{\pi\sqrt{3}}{8}.$$

But $\Sigma_4 = \frac{3}{4} + 2Re(g(\omega))$ and so c_4 is as given. For $r = 5$, $\omega = i$ and we find that

$$\Sigma_5 = g(1) + g(i) + g(-1) + g(-i) = 3 - 2\log 2 - \pi/2,$$

and so c_5 is as given. □

Proposition 11 *For any* $r \geq 2$,

$$\frac{\zeta(3)}{r+1} < c_r < \frac{r\zeta(3)}{(r-1)^2}.$$

Also

$$\begin{aligned}
c_r &= r\sum_{k=3}^{\infty}\left(-\frac{1}{r-1}\right)^{k-1}(2^{k-2}-1)\zeta(k) \\
&= \frac{r}{(r-1)^2}\zeta(3) - 3\frac{r}{(r-1)^3}\zeta(4) + 7\frac{r}{(r-1)^4}\zeta(5) - \cdots.
\end{aligned}$$

Both of these results show that $c_r \sim \zeta(3)/r$ *as* $r \to \infty$.

Proof We may write

$$c_r = r(r-1)^{-2} \sum_{k=1}^{\infty} (k(k+1/(r-1))(k+2/(r-1)))^{-1}.$$

It follows that $c_r < \frac{r}{(r-1)^2}\zeta(3)$, and

$$c_r > r(r-1)^{-2}\left(1+\frac{1}{r-1}\right)^{-1}\left(1+\frac{2}{r-1}\right)^{-1}\zeta(3) = \frac{1}{r+1}\zeta(3).$$

Also, for any $0 \le x \le 1$

$$\sum_{k=1}^{\infty}(k(k+x))^{-1} = \sum_{k=1}^{\infty} k^{-2}\sum_{j=0}^{\infty}(-x/k)^j = \sum_{k=2}^{\infty}(-x)^{k-2}\zeta(k).$$

Hence, for any $a > 1$

$$\begin{aligned}\sum_{k=1}^{\infty}(k(k+1/a)(k+2/a))^{-1} &= \sum_{k=1}^{\infty}\frac{a}{k}\left(\frac{1}{k+1/a}-\frac{1}{k+2/a}\right)\\ &= \sum_{k=3}^{\infty}(-1/a)^{k-3}(2^{k-2}-1)\zeta(k).\end{aligned}$$

Thus

$$\begin{aligned}c_r &= r\sum_{k=3}^{\infty}\left(-\frac{1}{r-1}\right)^{k-1}(2^{k-2}-1)\zeta(k)\\ &= \frac{r}{(r-1)^2}\zeta(3) - 3\frac{r}{(r-1)^3}\zeta(4) + 7\frac{r}{(r-1)^4}\zeta(5) - \cdots\end{aligned}$$

□

It remains only to prove Propositions 8 and 9.

Proof of Proposition 8 We estimate the maximum total weight of edges that can be deleted without increasing the number of components, which are all cycles. Let C_k be the random number of k-cycles in $G_{n,2}$. Then using the configuration model we can prove that for $k \ge 3$, $\mathbf{E}(C_k) = (1+O(\frac{k}{n}))\frac{1}{k}$. So the expected 'savings' from k-cycles is

$$\left(1+O\left(\frac{k}{n}\right)\right)\frac{1}{k}\left(1-\frac{1}{k+1}\right) = \left(1+O\left(\frac{k}{n}\right)\right)\frac{1}{k+1}.$$

Hence the total expected savings from cycles of length at most k is

$$\left(1+O\left(\frac{k}{n}\right)\right)\sum_{i=3}^{k}\frac{1}{i+1} = \left(1+O\left(\frac{k}{n}\right)\right)(\log k + O(1)).$$

Take $k \sim n/\sqrt{\log n}$. Then the total savings is at least

$$\left(1 + O\left(\frac{k}{n}\right)\right)(\log k + O(1)) = \log n + O(\sqrt{\log n}),$$

and is at most this value plus $n/k \sim \sqrt{\log n}$. □

Proof of Proposition 9 Consider the complete graph K_n, with independent edge lengths X_e on the edges e, each uniformly distributed on $(0,1)$. Call this the random network $(K_n, \mathbf{X})$. Form a random subgraph H on the same set of vertices by including the edge e exactly when $X_e \leq p$, and give e the length X_e/p. We thus obtain a random graph $G_{n,p}$ with independent edge lengths, each uniformly distributed on $(0,1)$. Call this the random network $(H, \mathbf{Y})$. We observe

$$mst(K_n, \mathbf{X})\, \mathbf{1}_{(H \text{ connected})} \leq p\, msf(H, \mathbf{Y}) \leq mst(K_n, \mathbf{X}).$$

The theorem now follows easily from the fact that that $mst(K_n, \mathbf{X}) \to \zeta(3)$ as $n \to \infty$, in probability and in any mean [4, 5]. □

Acknowledgement We would like to thank Noga Alon, Bruce Reed and Gunter Rote for helpful comments.

References

[1] F. Avram and D. Bertsimas, The minimum spanning tree constant in geometrical probability and under the independent model: a unified approach, *Annals of Applied Probability* **2** (1992) 113 - 130.

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

[3] B. Bollobás and I. Leader, *Exact face-isoperimetric inequalities*, European Journal on Combinatorics **11** (1990), 335-340.

[4] A.M. Frieze, On the value of a random minimum spanning tree problem, *Discrete Applied Mathematics* **10** (1985) 47 - 56.

[5] A.M. Frieze and C.J.H. McDiarmid, On random minimum length spanning trees, *Combinatorica* **9** (1989) 363 - 374.

[6] R.Graham, D.E.Knuth and O.Patashnik, *Concrete Mathematics*, Addison Wesley 1989.

[7] S. Janson, The minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph, *Random Structures and Algorithms* **7** (1995) 337 - 355.

[8] D.E.Knuth, The art of computer programming, Volume 1, Fundamental Algorithms, Addison-Wesley, 1968.

[9] L. Lovász, *Combinatorial Problems and Exercises,* North-Holland, 1993.

[10] C.J.H. McDiarmid, On the method of bounded differences, in Surveys in Combinatorics, ed J. Siemons, London Mathematical Society Lecture Note Series 141, Cambridge University Press, 1989.

[11] B.D. McKay and N.C. Wormald, Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{\frac{1}{2}})$, *Combinatorica* **11** (1991) 369 - 382.

[12] M. Penrose, *Random minimum spanning tree and percolation on the n-cube*, Random Structures and Algorithms **12** (1998) 63 - 82.

[13] M. Steele, *Probability Theory and Combinatorial Optimization*, SIAM 1997.

[14] M.Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces*, Publ. Math. IHES **81** (1995) 73 - 205.

[15] D.B. West, *Introduction to Graph Theory,* Prentice Hall, 1996.