On the distribution of the minimum weight clique

Alan Frieze, Wesley Pegden, Gregory Sorkin

June 15, 2016

Abstract

We determine, asymptotically in $n$, the distribution of the weight of a minimum-weight $k$-clique (or any strictly balanced graph $H$) in a randomly edge-weighted complete graph $K_n$.

1 Introduction

In this note we consider the distribution of the minimum weight copy of a fixed subgraph $H$ in a randomly edge weighted complete graph $K_n$. This seemingly natural problem seems to be absent from the literature thus far. It may be of particular interest where $H$ is a $k$-clique; in this case, the distribution we seek would allow one to do statistics rigorously when evaluating whether an observation of a small-weight clique in a randomly weighted network is significant.

We focus on the case where each edge $e$ in $G$ has an independent weight $X_e$ which is uniform in $[0,1]$, and $H$ is a complete graph $K_k$, but our results can be extended to more general settings, as we discuss in Section 6. We only consider finding subgraphs $H$ of the complete graph $K_n$, but the same approach may be applicable to other networks of interest.

Our result when $H$ is a clique is the following:

**Theorem 1** Let $k \geq 3$ be a positive integer. Let the edges of $K_n$ be given independent uniform $[0,1]$ weights, and let $W_k$ denote the minimum weight of a $k$-clique in this graph. Then we have, asymptotically in $n$,

(a) \[ \Pr \left( W_k \geq \frac{\alpha}{n^{2/(k-1)}} \right) = \exp \left\{ -\frac{\alpha \binom{k}{2}}{\binom{k}{2}! k!} \right\} + o(1). \]

(b) \[ \mathbb{E}(W_k) \cdot n^{2/(k-1)} \approx \phi_k := \frac{\binom{k}{2}! k!}{\binom{k}{2} \Gamma \left( \frac{1}{\binom{k}{2}} \right)}. \]
Here $\Gamma$ is the gamma function, and $A_n \approx B_n$ means that $A_n/B_n \to 1$ as $n \to \infty$.

We can easily generalise Theorem 1 to a more general class of subgraphs $H$ than cliques. The density $\text{den}(H)$ of a graph $H$ is defined to be $e(H)/v(H)$ where $e(H)$ and $v(H)$ denote the number of edges and vertices of $H$. A graph $H$ is said to be strictly balanced if

$$\text{den}(H) > \text{den}(H')$$

for all strict subgraphs $H' \subset H$.

**Theorem 2** Let $H$ be a fixed strictly balanced graph, let the edges of $K_n$ be given independent uniform $[0, 1]$ edge weights, and let $W_H$ denote the minimum weight of a subgraph isomorphic to $H$. Then we have

(a) 

$$\mathbb{P} \left( W_H \geq \frac{\alpha}{n^{1/\text{den}(H)}} \right) = \exp \left\{ - \frac{\alpha e(H)}{e(H)! \text{aut}(H)} \right\} + o(1).$$

(b) 

$$\mathbb{E}(W_H) \cdot n^{1/\text{den}(H)} \approx \phi_H := \frac{(e(H)!)^{1/e(H)}}{e(H)} \Gamma \left( \frac{1}{e(H)} \right).$$

Here, $\text{aut}(H)$ is the number of automorphisms of $H$.

In the case when $H$ is a clique we have made some effort to control the $o(1)$ error, to demonstrate that this approach is useful for reasonable problem sizes.

We define

$$\gamma := \frac{4}{(k-1)(3k+4)} \quad \text{and} \quad \omega := n^\gamma \quad \text{and} \quad \xi_k := \frac{k}{3k+4} \quad \text{and} \quad p := \frac{\omega}{n^{2/(k-1)}}. \quad (2)$$

We show that

**Theorem 3** If $n \geq 5k$, $\alpha \leq \omega$ and $p < 1/3$ then we have that

$$\left| \mathbb{P} \left( W_k \geq \frac{\alpha}{n^{2/(k-1)}} \right) - \exp \left\{ - \frac{\alpha (k)}{(k)!} \right\} \right| \leq \frac{1}{n^{\xi_k}} \left( 2 + \frac{3}{(k-2)!} + \frac{(3k!\xi_k \log n)^{1/2}}{e} \right).$$

Note that Theorem 1(a) is implied by Theorem 3. Assuming that $n$ is large, when $\alpha > \omega$ the claim about probabilities are covered by the $o(1)$ term. This is because for such $\alpha$ the claim in Theorem 1(a) is simply that $\mathbb{P} \left( W_k \geq \alpha/n^{2/(k-1)} \right) = o(1)$, which is implied by $\mathbb{P} \left( W_k \geq \omega/n^{2/(k-1)} \right) = o(1)$.

2 Outline of the approach

For the bulk of the paper we focus on the case where $H$ is a clique (Theorem 1). In Section 6.2, we discuss how to extend the analysis to prove Theorem 2.
For purposes of analysis we restrict our attention to edges of weight at most \( p \), and thus to cliques all of whose edges have weight at most \( p \). This set of edges forms a random graph \( G = G_{n,p} \). We choose \( p \) large enough that the lowest clique weight \( W_k \) satisfies \( W_k \leq p \) w.h.p.\(^1\) (so each edge constituting this clique also has cost \( \leq p \), and it is among the set of remaining cliques), but small enough that w.h.p. the \( k \)-cliques of \( G \) are disjoint (thus their edge costs are independent, making it straightforward to find the distribution of the cost of the cheapest one). Fortunately, it is possible to satisfy both conditions simultaneously, not just for cliques but for any strictly balanced graph \( H \): the threshold value \( p \) for which \( G_{n,p} \) is expected to contain a copy of \( H \) is a power of \( n \) lower than that at which it will contain overlapping copies of \( H \). Choosing an intermediate power rules out overlapping copies of \( H \) (by the first-moment method), while giving a large expected number of copies of \( H \). Using the Stein-Chen method we show that the actual number of copies is almost certainly large, and highly predictable. The edge weights within these copies are i.i.d. \( U(0, p) \), by disjointness, and there is constant probability (independent of \( n \)) that a given copy has total weight \( \leq p \), so (since there are many copies) there is almost certainly one of cost at most \( p \).

We use the Stein-Chen method to show that the number of cliques is asymptotically Poisson. This is well known, but we include a proof for completeness and for explicit error terms. This will imply that the number of cliques is concentrated, as well as large in expectation, allowing us to compute asymptotically \( P(W_k \geq x) \) for \( x \leq p \).

3 Stein-Chen Method

We use a version of the Stein-Chen method given in [1, Theorem 1]. This formulation controls the distribution of \( X = \sum_{\alpha \in I} X_\alpha \) where \( I \) is an arbitrary index set and each \( X_\alpha \) is a Bernoulli random variable \( \text{Be}(p_\alpha) \), that is, 1 with this probability and 0 otherwise. In our case, \( X \) is the number of \( k \)-cliques in \( G_{n,p} \),

\[
I = \binom{|I|}{k}
\]

is the set of all \( k \)-cliques of \( K_n \) (potential \( k \)-cliques of \( G \), and \( X_\alpha = 1 \) if clique \( \alpha \) is present and 0 otherwise. Here \( p \) will as defined in (2).

The method of [1, Theorem 1] introduces, for each \( \alpha \in I \), a “neighborhood of dependence” \( B_\alpha \subseteq I \) with the property that \( X_\alpha \) is independent of all the \( X_\beta \) for \( \beta \) outside of \( B_\alpha \). (The method actually allows for “near independence”, but we do not require this.) Define

\[
b_1 := \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} p_\alpha p_\beta,
\]

\[
b_2 := \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B_\alpha} p_\alpha p_\beta, \text{ where } p_\alpha p_\beta = \mathbb{E}(X_\alpha X_\beta).
\]

Then, with \( \lambda = \mathbb{E}X \) and \( Z \sim \text{Po}(\lambda) \) a Poisson random variable with mean \( \lambda \), the total variation distance between \( X \) and \( Z \) is at most \( 2(b_1 + b_2) \). (An additional term \( b_3 \) is involved if there are weak dependencies.)

\(^1\)We say that a sequence of events \( \mathcal{E}_n, n \geq 1 \) occurs with high probability (w.h.p.) if \( \lim_{n \to \infty} P(\mathcal{E}_n) = 1 \).
Here, we define $B_\alpha$ to be the set of cliques that share an edge with clique $\alpha$. Then,

$$X = \sum_{\alpha \in I} X_\alpha$$

$$\lambda = EX = \binom{n}{k} p^{(k)} \approx \frac{\omega^{(k)}}{k!}$$, \hspace{1cm} (4)

$$b_1 = \binom{n}{k} p^{(k)} \sum_{\ell=2}^k u_\ell$$ \text{ where } u_\ell = \binom{k}{\ell} \left( \frac{n-k}{k} \right),$$

$$b_2 = \binom{n}{k} p^{(k)} \sum_{\ell=2}^{k-1} v_\ell$$ \text{ where } v_\ell = \binom{k}{\ell} p^{(k)} - \binom{\ell}{2}.$$

Note that with reference to (2), $\gamma > 0$ implies that w.h.p. $G_{n,p}$ contains a $k$-clique. We claim that for $k \geq 3$, we also have

$$\gamma < \frac{2k-4}{(k-1)(k(k-1)-1)} = \frac{2}{k-1} - \frac{2(k-1)}{2^{(k-1)} - 1},$$ \hspace{1cm} (5)

which will imply that w.h.p. no two $k$-cliques share an edge. Indeed, we check that

$$(2k-4)(3k^2 + k - 4) - 4(k-1)(k(k-1)-1) = 2k^3 - 2k^2 - 12k + 12 = k((k-1)^2 - 7) + 12 > 0,$$

for $k \geq 3$.

Now with $I$ as in (3), let $K_\alpha, \alpha \in I$, denote the $k$-cliques of $K_n$.

For $\alpha \in I$ let $X_\alpha = 1$ if $K_\alpha$ appears in $G_{n,p}$ and let $X_\alpha = 0$ otherwise. Let $X$ be the number of copies of $K_k$ in $G_{n,p}$. Then it follows from Theorem 1 of [1] that for $s \geq 0$

$$\left| P(X \leq s) - \sum_{r=0}^{s} \frac{\lambda^r e^{-\lambda}}{r!} \right| \leq b_1 + b_2.$$  

(Note that $b_3$ of [1] is 0 in our case. $B_\alpha$ for us is the set of cliques that share an edge with $K_\alpha$.)

Note now that if $\ell \leq k - 2$ then

$$\frac{v_{\ell+1}}{v_\ell} = \frac{(k-\ell)^2}{p^\ell(n-2k+\ell+1)(\ell+1)}.$$  

Thus if $2 < \ell < k - 2$ then, if $n \geq 5k$ we have

$$\frac{v_{\ell+1}}{v_\ell} \leq \frac{(k-\ell)^2}{p^\ell(n-2k+\ell+1)(\ell+1)} \leq \frac{4}{3} \cdot \left( \frac{4}{3} \right)^2 \cdot \frac{13}{12} \cdot p < 3p.$$  

Now we see that if $p < 1/3$ then the sequence $v_\ell, 2 \leq \ell \leq k - 2$ is log-convex and so is maximised at $\ell = 2$ or $\ell = k - 1$.

Furthermore,

$$\frac{v_2}{v_{k-1}} \leq \frac{(k-1)(n-k-1)!p^{(k-1)} - 1}{2(k-2)!(n-2k+2)!} \leq \frac{k-1}{2(k-2)!} n^\tau,$$  

4
where (see (5))
\[
\tau = (k - 3) - \left( \binom{k-1}{2} - 1 \right) \cdot \frac{2(k-1)}{k(k-1)-1} \leq 0.
\]

It follows that
\[
b_2 \leq \binom{n}{k} p^{(k)}(k-2) v_{k-1} = \binom{n}{k} \frac{\omega^{(k)}_2}{n^k} \cdot (k-2) \cdot k(n-k) \frac{\omega^{k-1}}{n^2} \leq \frac{\omega^{(k-1)(k+2)/2}}{(k-2)! n}.
\]

Now \( b_1 \leq b_2 \) since \( u_k < u_{k-1} \) and for \( 2 \leq \ell \leq k-1 \) we see that the \( \ell \)th summand in \( b_1 \) is equal to \( p^{(\ell)}_2\) times the \( \ell \)th summand in \( b_2 \). From \([1, \text{Theorem 1}]\) it follows that for \( \varepsilon > 0 \) we have
\[
P(X \notin [(1-\varepsilon)\lambda, (1+\varepsilon)\lambda]) \leq P(\text{Po}(\lambda) \notin [(1-\varepsilon)\lambda, (1+\varepsilon)\lambda]) + 2b_2.
\]

(6)

4 Clique separation in \( G_{n,p} \)

We next consider the probability that \( G_{n,p} \) contains a pair of cliques that share an edge. This can be bounded by
\[
\binom{n}{k} \binom{k}{2} \binom{n-k}{k-2} p^{(k)} v_{k-2} = \binom{n}{k} p^{(k)} v_2 \leq b_2.
\]

(7)

Now we let \( \mathcal{E} \) denote the event that \( X \in [(1-\varepsilon)\lambda, (1+\varepsilon)\lambda] \) AND that all \( k \)-cliques in \( G_{n,p} \) are edge-disjoint. The value of \( \varepsilon \) is given below in (9). From (6) and (7) we have
\[
P(\neg \mathcal{E}) \leq 2 \exp \left\{ -\frac{\varepsilon^2 \lambda}{3} \right\} + 3b_2 \leq 2 \exp \left\{ -\frac{\varepsilon^2 \omega^{(k)}_2}{3k!} \right\} + \frac{3\omega^{(k-1)(k+2)/2}}{(k-2)! n}.
\]

(8)

The expression \( 2 \exp \left\{ -\frac{\varepsilon^2 \lambda}{3} \right\} \) bounds the probability that \( \text{Po}(\lambda) \) deviates from its mean \( \lambda \) by an amount \( \varepsilon \lambda \).

We (approximately) minimise the RHS expression in (8) by choosing \( \varepsilon \) so that
\[
\varepsilon^2 \omega^{(k)}_2 = A_1 \log n
\]

for some constant \( A_1 > 0 \) that will be defined below.

In which case
\[
\varepsilon = \frac{(A_1 \log n)^{1/2}}{\omega^{(k)/2}} = \frac{(A_1 \log n)^{1/2}}{n^{(k)/2}}.
\]

We then have from (8) that
\[
P(\neg \mathcal{E}) \leq \frac{2}{n^{A_1/(3k!)}} + \frac{3\omega^{(k-1)(k+2)/2}}{(k-2)! n}.
\]

(10)
5 Uniform $[0,1]$ weights

5.1 Estimating probabilities

Suppose now that we give uniform $[0,1]$ weights to the edges of $K_n$. Fix $p, \omega$ as above. Let $W_k$ denote the minimum weight of a $k$-clique. Let 

$$x = \frac{\alpha}{n^{2/(k-1)}}.$$

We wish to estimate $P(W_k \leq x)$ where $x \leq p$ is not too large. Let $C = \{C_1, C_2, \ldots, C_m\}$ be the set of $k$-cliques in the graph induced by edges of weight at most $p$. Let $Z_i$ be independent copies of a random variable which is the sum of $\binom{k}{2}$ uniform distributions on $[0, p]$. In particular, each $Z_i$ is identical in distribution to the weight of $C_i$. Assuming that $\alpha \leq \omega$, we have from the definitions of $Z_i, x, p$ that 

$$P(Z_i \leq x) = \frac{\alpha^{\binom{k}{2}}}{(\binom{k}{2})! \omega^{\binom{k}{2}}}.$$ 

Now, from the definition of $E$, we have for $\ell \geq 1$ that 

$$P(W_k \geq x \mid X = \ell, E) = \left(1 - \frac{\alpha^{\binom{k}{2}}}{(\binom{k}{2})! \omega^{\binom{k}{2}}}\right)^\ell.$$ 

Note that since $x \leq p$, the given events $X \geq 1$ and $E$ imply that any cliques of $K_n$ containing edges not in $G_{n,p}$, i.e. not in $C$, are not of minimum weight.

Therefore, letting $\rho_{10}$ denote the RHS of (10), we have 

$$\left(1 - \frac{\alpha^{\binom{k}{2}}}{(\binom{k}{2})! \omega^{\binom{k}{2}}}\right)^{(1+\varepsilon)\lambda} - \rho_{10} \leq P(W_k \leq x) \leq \left(1 - \frac{\alpha^{\binom{k}{2}}}{(\binom{k}{2})! \omega^{\binom{k}{2}}}\right)^{(1-\varepsilon)\lambda} + \rho_{10}. \quad (11)$$ 

We will now use the following elementary inequalities:

$$1 - \xi \leq e^{-\xi} \text{ and } 1 - \xi \geq e^{-\xi/(1-\xi)} \text{ for } 0 \leq \xi \leq 1 \text{ and } \frac{1}{1 - \xi} \leq 1 + 2\xi \text{ for } 0 \leq \xi \leq \frac{1}{2}.$$ 

We will assume that $\alpha \leq \omega$. This implies that:

$$\frac{\alpha^{\binom{k}{2}}}{(\binom{k}{2})! \omega^{\binom{k}{2}}} \leq \frac{1}{3}. \quad (12)$$

Now we compute:

$$\text{LHS}(11) \geq \exp \left\{ - \frac{(1 + \varepsilon)\alpha^{\binom{k}{2}}}{(\binom{k}{2})! \omega^{\binom{k}{2}}} \left(1 - \frac{\alpha^{\binom{k}{2}}}{(\binom{k}{2})! \omega^{\binom{k}{2}}}\right) \right\} - \rho_{10}$$ 

$$\geq \exp \left\{ - \frac{(1 + \varepsilon)\alpha^{\binom{k}{2}}}{(\binom{k}{2})! \omega^{\binom{k}{2}}} \left(1 + \frac{2\alpha^{\binom{k}{2}}}{(\binom{k}{2})! \omega^{\binom{k}{2}}}\right) \right\} - \rho_{10}, \quad \text{using (4) and (12)}$$
We have used

At this point, we have not chosen values for $A$. For example, if (12) holds then

$$ \rho \leq (k!) \frac{1}{(k-2)!} n \log(\frac{n}{k!\omega(k)}) \leq 2 \frac{n^{2/(k-1)}}{n^{1/(3k!)}} + \frac{3\omega^{(k-1)(k+2)/2}}{(k-2)!n} + \frac{3(A_1 \log n)^{1/2} \alpha}{(k!) \omega^{(k+2)/2} k!} \exp \left\{ -\frac{\alpha^{(k)}_{\omega(k)}}{(k!) \omega^{(k)/2} k!} \right\} \leq \frac{2}{n^{A_1/(3k!)}} + \frac{3\omega^{(k-1)(k+2)/2}}{(k-2)!n} + \frac{3(A_1 \log n)^{1/2}}{e\omega^{(k)/2}}. $$

We have used $ye^{-y} \leq e^{-1}$ with $y = \frac{\alpha^{(k)}_{\omega(k)}}{(k!) \omega^{(k)/2} k!}$.

At this point, we have not chosen values for $A_1$. We will choose $A_1 = 3k!\xi_k$ for simplicity and then from (2) we see that if (12) holds then

$$ \left| P\left(W_k \geq \frac{\alpha}{n^{2/(k-1)}}\right) - \exp \left\{ -\frac{\alpha^{(k)}_{\omega(k)}}{(k!) \omega^{(k)/2} k!} \right\} \right| \leq \frac{2}{n^{A_1/(3k!)}} + \frac{3\omega^{(k-1)(k+2)/2}}{(k-2)!n\xi_k} + \frac{(3k!\xi_k \log n)^{1/2}}{en^{k\xi_k}}. $$

For example, if $k = 3$ (i.e. triangle case) then $\xi_3 = 3/13$ and so if (from (12)) $\alpha \leq n^{2/13}$ then

$$ \left| P\left(W_k \geq \frac{\alpha}{n}\right) - e^{-\alpha^3/36} \right| \leq \frac{5e + (5 \log n)^{1/2}}{en^{3/13}}. $$

We have now proved Theorem 3 and Theorem 1(a).

### 5.2 Estimating expectation

We now obtain an estimate of $E(W_k)$ for general $k$. Indeed,

$$ E(W_k) = \int_{x=0}^{\alpha^{(k)/2}} P(W_k \geq x)dx $$
\[
= \frac{1}{\binom{n}{k}} \int_{\alpha=0}^{\omega} \exp \left\{ -\frac{\alpha (k^2)}{2} \right\} d\alpha + O\left( e^{-n^{\Omega(1)}} \right)
\]

\[\approx \frac{\phi_k}{n^{2/(k-1)}},\]

where \(\phi_k\) is as given in (1).

This proves Theorem 1(b).

6 Extensions and Generalisations

6.1 Not just uniform \([0, 1]\)

The first extension to consider is to replace the uniform distribution by something else. Suppose then that \(X_e\) is a random variable and that it has a density \(f(x)\) with \(f(0) = D > 0\). Near the origin \(Y_e = DX_e\) is almost identically distributed as a uniform \([0, 1]\) random variable and so Theorem 3 can easily be modified to handle this situation. In particular to find the expectation of \(W_k\) asymptotically, all we have to do is multiply the expression in Theorem 3 by \(D^{-1}\).

6.2 Strictly balanced \(H\)

Erdős and Rényi [3] identified the threshold for the existence of \(H\) and it was further shown by Bollobás [2] and Karoński and Ruciński [6] that if \(W_H\) denotes the number of copies of \(H\) in \(G_{n,p}\) that if \(np^{\text{den}(H)} \to c\), \(c\) constant and that if \(\lambda = \frac{\nu(H)}{\text{aut}(H)}\) then

\[P(W_H = k) \approx \frac{\lambda^{k} e^{-\lambda}}{k!}.\]

Frieze [4] extended this to the case where \(\lambda \leq n^{\varepsilon H}\), for some positive constant dependent only on \(H\). We note that \(v(K_k) = k\), \(e(K_k) = \binom{k}{2}\) and so \(\text{den}(K_k) = (k-1)/2\). Furthermore, \(\text{aut}(K_k) = k!\).

To apply our approach, all we have to do is replace \(n^{-2/(k-1)}\) by \(n^{-1/\text{den}(H)}\) and \(\alpha^{\binom{k}{2}} / (\binom{k}{2}!k!)\) by \(\alpha^{\nu(H)} / (e(H)!\text{aut}(H))\). Lemma 6.1 (for example) of [4] shows that the density of the union of two copies of \(H\) that share at least one edge is strictly greater than \(\text{den}(H)\) and so we can restrict our attention to a collection of edge disjoint copies of \(H\), as we have done for complete graphs, to obtain Theorem 2.

References


