The distribution of minimum-weight cliques and other subgraphs in graphs with random edge weights

Alan Frieze[*], Wesley Pegden[†], Gregory Sorkin[‡]

July 3, 2017

Abstract

We determine, asymptotically in n, the distribution and mean of the weight of a minimum-weight k-clique (or any strictly balanced graph H) in a complete graph K_n whose edge weights are independent random values drawn from the uniform distribution or other continuous distributions. For the clique, we also provide explicit (non-asymptotic) bounds on the distribution's CDF in a form obtained directly from the Stein-Chen method, and in a looser but simpler form. The direct form extends to other subgraphs and other edge-weight distributions. We illustrate the clique results for various values of k and n. The results may be applied to evaluate whether an observed minimum-weight copy of a graph H in a network provides statistical evidence that the network's edge weights are not independently distributed but have some structure.

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; a natural special case that may be of particular interest is when H is a k-clique. This seemingly natural problem seems to be absent from the literature thus far. It can be viewed as a fixed-size version of the NP-complete Maximum Weighted Clique Problem (MWCP); the review article [3] on Maximum Clique includes discussion of MWCP algorithms (Section 5.3) and the performance of heuristics on random graphs (in Section 6.5). The same article reviews applications, including in telecommunications, fault diagnosis, and computer vision and pattern recognition. Wikipedia's article [5] on the Clique problem includes applications in chemistry, bioinformatics, and social networks, and many of these seem more naturally modelled as weighted than unweighted problems. Research on fast algorithms for MWCP in "massive graphs" arising in practice is ongoing; a recent example, [4], notes applications in telecommunications and biology (specifically, the study [11] of finding a 5-protein gene marker for Alzheimer's disease).

The weight distribution of a minimum-weight subgraph H of a randomly weighted network is a natural question in discrete mathematics and probability, and in addition has statistical ramifications that we think may be of importance in applied settings. Taking the clique for purposes of

[*]Department of Mathematical Sciences, Carnegie Mellon University. Research supported in part by NSF Grants DMS1362785, CCF1522984 and a grant(333329) from the Simons Foundation. alan@random.math.cmu.edu.

[†]Department of Mathematical Sciences, Carnegie Mellon University. Research supported in part by NSF Grant DMS1363136 and the Sloane Foundation. wes@math.cmu.edu.

[‡]Department of Mathematics, London School of Economics and Political Science. g.b.sorkin@lse.ac.uk

discussion, the distribution we seek would allow one to judge whether the weight of the smallest-weight clique in a given network provides statistical evidence that the network's edge weights are not independently, uniformly random. Because uniformity is not essential to our analysis (see Section 6), an anomalous weight is evidence that the edge weights are not independent, i.e., that the network's weights have some structure. The statistics could alternatively be performed by repeated simulations of random networks matching the null hypothesis but our approach is preferable for the usual reasons that simulation is cumbersome, it does not provide rigorous results, and the number of simulations must be more than the reciprocal of the desired significance level (potentially quite large), and additionally because finding a minimum-weight clique in a graph is an NP-hard problem that each simulation would have to solve.

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, and here we derive explicit (non-asymptotic) bounds. We also obtain asymptotic results for other graphs H, and we extend both sets of results to other distributions. We only consider finding subgraphs H of the complete graph K_n, but the same approach may be applicable to other networks of interest.

The *density* of a graph H is defined as $\operatorname{den}(H) = e(H)/v(H)$ where $e(H)$ and $v(H)$ denote the number of edges and vertices of H, and H is *strictly balanced* if

$$\operatorname{den}(H) > \operatorname{den}(H') \text{ for all strict subgraphs } H' \subset H.$$

Theorem 1 *Let H be a fixed strictly balanced graph with v vertices, m edges, a automorphisms, and density $d = m/v$. Let the edges of K_n be given independent uniform $[0,1]$ edge weights, and let W denote the minimum weight of a subgraph isomorphic to H. Then, for any non-negative $z = z(n)$ asymptotically in n,*

$$\mathsf{P}\left(W \geq \frac{z}{n^{1/d}}\right) = \exp\left\{-\frac{z^m}{m!\,a}\right\} + o(1). \tag{1}$$

Also,

$$\mathsf{E}(W) \sim \hat{\mu} := n^{-1/d}\frac{(m!\,a)^{1/m}}{m}\Gamma\left(\frac{1}{m}\right). \tag{2}$$

Here $f(n) \sim g(n)$ means that $f(n)/g(n) \to 1$ as $n \to \infty$, which we may also write as $f(n) = g(n)\,(1+o(1))$.

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

Theorem 2 *Fix $k \geq 3$, let the edges of K_n be given independent uniform $[0,1]$ edge weights, and let W denote the minimum weight of a clique K_k. Let*

$$w = \frac{z}{n^{2/(k-1)}} \text{ and } \lambda = \binom{n}{k}\frac{w^{\binom{k}{2}}}{\binom{k}{2}!} \sim \frac{z^{\binom{k}{2}}}{k!\binom{k}{2}!} \quad \text{(as in (9))}.$$

Then, for $z \leq \min\left\{n^{2/\binom{k}{2}}, n^{2/(k-1)}\exp(-\frac{k-1}{k-2})\right\}$ (equivalently, $w \leq \min\left\{n^{-2/k}, \exp(-\frac{k-1}{k-2})\right\}$),

$$\left|\mathsf{P}\left(W \geq w = \frac{z}{n^{2/(k-1)}}\right) - \exp(-\lambda)\right| \leq \frac{8}{7}\,\frac{(k-2)}{\binom{k}{2}!(k-1)!^2}\,\frac{z^{\binom{k}{2}+k-1}}{n}. \tag{3}$$

Theorem 3 derives tighter bounds for the clique than those of Theorem 2, but is presented later because it requires additional notation.

The proofs use the Stein-Chen method. Section 2 presents the method as applied to finding a low-weight clique, and establishes the explicit probability bounds of Theorem 3. Section 3 outlines how explicit bounds could be obtained for other distributions, and other subgraphs H. Section 4 simplifies (but loosens) the bounds to give Theorem 3.

Section 5 applies the Stein-Chen method to strictly balanced graphs H to obtain the asymptotic probability bounds and asymptotic expectation of Theorem 1. Section 6 generalizes Theorem 1 to non-uniform edge weight distributions, as Theorem 6,

Section 7 presents plots and tables for the application of Theorem 3 to various clique sizes k and graph sizes n, giving an indication of where our results are effective and where they need improvement. The Conclusions in Section 8 recapitulate the results achieved and discuss where they might be applied and how they might be extended.

2 Stein-Chen Method

We use a version of the Stein-Chen method given in [1, Theorem 1]. We will summarize the general formulation, which may become clearer when we then show how it applies in our case. The formulation characterizes the distribution of a value $X = \sum_{\alpha \in I} X_\alpha$ where I is an arbitrary index set and each X_α is a Bernoulli random variable, $X_\alpha \sim \text{Be}(p)$, that is, X_α is 1 with probability p and 0 otherwise. For each $z \in I$ there is a "neighborhood of dependence" $B_\alpha \subseteq I$ with the property that X_α is independent of all the X_β for β outside of B_α. With $\lambda = \mathsf{E}X$, $Z \sim \text{Po}(\lambda)$ a Poisson random variable with mean λ, and

$$b_1 := \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} p_\alpha p_\beta, \tag{4}$$

$$b_2 := \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B_\alpha} p_{\alpha\beta}, \text{ where } p_{\alpha\beta} = \mathsf{E}(X_\alpha X_\beta), \tag{5}$$

the conclusion of [1, Theorem 1] is that

$$\text{TVD}(X, Z) \leq b_1 + b_2, \tag{6}$$

where $\text{TVD}(X, Z)$ denotes the total variation distance between the two distributions, i.e., the maximum, over events E, of the difference in the probabilities of E under the two distributions. (The full theorem involves an additional term b_3 if there are weak dependencies, but we do not need this. Also, we have adjusted for our use of the standard definition of TVD, where [1] defines TVD as twice this.)

In our case, with $G = K_n$, we are interested in

$$X = \text{number of } k\text{-cliques of } G \text{ weighing} \leq w.$$

We are specifically interested in $\mathsf{P}(X = 0)$ i.e. the probability that there is no such clique. We have

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

where the index set I is the set of all k-cliques of G,

$$I = \binom{[n]}{k} \tag{7}$$

and

$$X_\alpha = \begin{cases} 1, & \text{if clique } \alpha \text{ has weight} \le w \\ 0, & \text{otherwise.} \end{cases}$$

Denoting the number of edges in a k-clique by

$$m = \binom{k}{2},$$

each X_α satisfies

$$X_\alpha \sim \mathrm{Be}(p)$$

where, for $w \le 1$,

$$p = p(w) := E(X_\alpha) = \frac{w^m}{m!}, \tag{8}$$

the probability that the sum of m i.i.d. uniform $[0, 1]$ variables is at most w. The sum of i.i.d. uniform random variables has an Irwin-Hall distribution, whose CDF is well known (more on this in Section 3.1), but it is not hard to see that when $w \le 1$ this probability is given by (8). For a clique to have total weight $\le w$, first, each of its m edges must have weight $\le w$. With i.i.d. uniform $[0, 1]$ random edge weights, conditioned on the vector of m edge weights lying in $[0, w]^m$, the vector is a uniformly random point in this m-dimensional cube, the event that the sum of its coordinates is $\le w$ is the event that the point lies in a standard m-dimensional simplex scaled by w, and the volume of this simplex is $w^m/m!$.

Immediately,

$$\lambda = \lambda(w) := \mathsf{E}X = \binom{n}{k} p = \binom{n}{k} \frac{w^m}{m!} \in \frac{n^k}{k!} \frac{w^m}{m!} \left[1 - \frac{k^2}{2n}, 1\right]. \tag{9}$$

The neighborhood of dependence B_α is the set of cliques that share an edge with clique α. Breaking this down more finely, into cliques sharing ℓ vertices with α, we see that

$$|B_\alpha| = \sum_{\ell=2}^{k} u_\ell \text{ where } u_\ell = \binom{k}{\ell}\binom{n-k}{k-\ell}. \tag{10}$$

Then,

$$b_1 = \binom{n}{k} \sum_{\ell=2}^{k} u_\ell p^2. \tag{11}$$

To calculate b_2, suppose that two cliques (or indeed copies of any graph H) α and β have a edges in common and each has b edges not in the other (with $a + b = m$). The event that $X_\alpha X_\beta = 1$, i.e., that both cliques weigh $\le w$, is equivalent to the shared edges weighing some $W_a \le w$, and both sets of unshared edges weighing $\leq w - W_a$. As in (8), the CDF for the sum of s edges to be at most r is $F_s(r) = r^s/s!$, so the corresponding density is $f_s(r) = r^{s-1}/(s-1)!$. Thus,

$$p_{\alpha\beta} = p_{\alpha\beta}(\alpha, \beta, w) = \int_0^w f_a(w_a)(F_b(w - w_a))^2 \, dw_a \tag{12}$$

$$= \int_0^w \frac{1}{(a-1)!}(w_a)^{a-1} \left(\frac{1}{b!}(w - w_a)^b \right)^2 \, dw_a \tag{13}$$

which by a change of variable to $t = w_a/w$ is

$$= \frac{w^{a+2b}}{(a-1)!\,(b!)^2} \int_0^1 t^{a-1}(1-t)^{2b} \, dt.$$

By Euler's integral of the first kind, $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt = \Gamma(x)\Gamma(y)/\,\Gamma(x+y)$, this is

$$= \frac{w^{a+2b}}{(a-1)!\,(b!)^2} \cdot \frac{(a-1)!(2b)!}{(a+2b)!} \tag{14}$$

$$= \frac{w^{a+2b}}{(a+2b)!} \binom{2b}{b}. \tag{15}$$

Cliques α and β sharing ℓ vertices share $a = \binom{\ell}{2}$ edges, while the number of edges unique to β is

$$b = m_\ell := \binom{k}{2} - \binom{\ell}{2}.$$

Thus, from (15), with (5) and (10),

$$b_2 = \binom{n}{k} \sum_{\ell=2}^{k-1} u_\ell \, p_{\alpha\beta}\left(\tbinom{\ell}{2}, \tbinom{k}{2} - \tbinom{\ell}{2}, w\right) \tag{16}$$

$$= \binom{n}{k} \sum_{\ell=2}^{k-1} u_\ell \, \frac{w^{m+m_\ell}}{(m+m_\ell)!} \binom{2m_\ell}{m_\ell}. \tag{17}$$

With

$$Z \sim \mathrm{Po}(\lambda)$$

the conclusion from [1, Theorem 1], per (6), is that

$$|\mathsf{P}(W \geq w) - \exp(-\lambda)| = |\mathsf{P}(X = 0) - \mathsf{P}(Z = 0)| \leq \mathrm{TVD}(X, Z) \leq b_1 + b_2.$$

We have thus proved the following theorem:

Theorem 3 *Fix $k \geq 3$, let the edges of K_n be given independent uniform $[0,1]$ edge weights, and let W denote the minimum weight of a clique K_k. Let $w = z/n^{2/(k-1)}$, $\lambda = \binom{n}{k} \frac{w^{\binom{k}{2}}}{\binom{k}{2}!} \sim \frac{z^{\binom{k}{2}}}{k!\binom{k}{2}!}$ (as in (9)), and b_1 and b_2 be as given in (11) and (17) (calling in turn on (8) and (10)). Then, for $z \leq n^{2/(k-1)}$ (equivalently, $w \leq 1$),*

$$\left| \mathsf{P}\left(W \geq w = \frac{z}{n^{2/(k-1)}} \right) - \exp(-\lambda) \right| \leq b_1 + b_2. \tag{18}$$

3 Extensions of Theorem 3

In this section, we illustrate how we could, with the aid of a computer, extend Theorem 3 and obtain precise values for b_1, b_2 in more general circumstances.

Given any edge weight distribution, we can at least in principle know the CDF $F_m(w)$ for a set of m edges to have total weight at most w, and the corresponding PDF $f_m(w)$. We may then generalise (8) to $p = F_m(w)$, where as usual $m = \binom{k}{2}$, as before define λ by (9) and b_1 by (11), and compute b_2 through (16), with $p_{\alpha\beta}$ given by (12).

3.1 Uniform edge weights

With edge weights uniformly distributed as before, but removing Theorem 3's restriction to $w \leq 1$, the sum of m uniform weights has Irwin-Hall distribution, with distribution and density functions

$$F_m(w) = \frac{1}{m!} \sum_{i=0}^{\lfloor w \rfloor} (-1)^i \binom{m}{i} (w-i)^m$$
$$f_m(w) = \frac{1}{(m-1)!} \sum_{i=0}^{\lfloor w \rfloor} (-1)^i \binom{m}{i} (w-i)^{m-1};$$

see [7, eq(12)], [8], [10, eq(26.48)].

Given k (thus m) and w it is straightforward to calculate $F_m(w)$ and thus p, λ, and b_1. Calculating b_2 reduces, for each ℓ in (16), to calculating $p_{\alpha\beta}$ through the integral in (12). Break the range of integration into intervals within which neither w nor w_a takes an integral value, by splitting at the points where either does take an integral value; since $w \leq m$ there are at most $2m$ such points. Within each subintegral, the integrand is the product of a polynomial in w_a (from $f_a(w_a)$) and a polynomial in $(w - w_a)$ (from the square of $F_b(w - w_a)$). Each term of this can be integrated as an Euler integral of the first kind, as was done in going from (13) to (14), or indeed expanded to a polynomial in w_a (the powers are all bounded in terms of k) where each term can be integrated straightforwardly (even as an indefinite integral).

In principle, then, we can extend Theorem (3) to all $w \leq m$. Indeed, for each w we can produce p and $p_{\alpha\beta}$, thus giving b_1 and b_2 as explicit functions of n. (We cannot get an explicit function of w, at least by the method above, because the partition of the integral into sub-integrals is different for each w.) In practice, a naive implementation in Maple struggles with $k = 10$ both in computation time and numerical stability.

3.2 Exponential edge weights

We may also consider exponentially distributed edge weights. Without loss of generality we assume rate 1; anything else is a simple rescaling. The sum of m rate-1 exponentials has Erlang distribution with well known distribution and density functions that can be stated in a variety of forms including

$$F_m(w) = 1 - e^{-w} \sum_{i=0}^{m-1} \frac{w^i}{i!}$$
$$f_m(w) = \frac{w^{m-1}e^{-w}}{(m-1)!};$$

see for example [6] and [9, eq(17.2)].

Again given w it is straightforward to compute p, λ, and b_1, while computing b_2 requires computing $p_{\alpha\beta}$ for each ℓ in the sum in (16) and the key is to evaluate (12). In this case the form of $F_b(w)$ means that the integrand is a finite sum (with length a function of k), each term of which has form a constant (with respect to w) times $(w_a)^a e^{-w_a}$ (coming from $f_a(w_a)$) times $(w-w_a)^r e^{-s(w-w_a)}$ (from the square of $F_b(w-w_a)$), for some integers r and s. These may in turn be expanded to terms of form $(w_a)^r e^{s\,w_a}$. Each of these is integrable (even as an indefinite integral); alternatively, each definite integral (over w_a from 0 to w) is a lower incomplete gamma function.

It thus seems feasible, if unenviable, to extend Theorem 3 to provide explicit bounds for exponential edge weights. In practice, Maple has little difficulty with the calculations through $k = 6$ but they quickly get more difficult: at $k = 10$ is is challenging to calculate even a single ℓ the corresponding function $p_{\alpha\beta}(w)$.

3.3 Subgraphs H other than cliques

Generalising Theorem 3 to subgraphs H other than cliques appears straightforward. The neighborhood of dependence of a given copy of H needs more careful treatment, but this can be done in this non-asymptotic setting precisely as presented in Section 5 for asymptotic calculations. The explicit calculations here do not even require that H be strictly balanced, but it can be expected that the error term b_2 will be large if it is not (for the same reasons that the strict balance is generally required in application of the second moment method).

4 Calculating bounds

Given values of k and w, in practice one would apply Theorem 3 using (18), calculating $b_1 + b_2$ exactly from (11) and (17), as indeed we do in Section 7. However, to characterize the quality of the estimate of $\mathsf{P}(W \geq w)$ we derive an upper bound on $b_1 + b_2$ as a relatively simple (summation-free) function of k and w.

We start with b_2, the more difficult and (as we will see) larger of these two parameters. First, in lieu of (15), we observe that for clique β to have weight at most w, the m_ℓ edges unique to it must have total weight $\leq w$, and therefore

$$\mathsf{P}(X_\beta = 1 \mid X_\alpha = 1) \leq \frac{w^{m_\ell}}{m_\ell!} \leq \frac{w^{m_\ell}}{(k-1)!}. \tag{19}$$

In the first inequality we have used that since these edges are unique to β, conditioning on the weight of α being at most w is irrelevant. The second is simply because, over the range of ℓ from

2 to $k-1$, $m_\ell \geq \binom{k}{2} - \binom{k-1}{2} = k-1$. Also,

$$u_\ell \leq \binom{k}{\ell} \frac{n^{k-\ell}}{(k-\ell)!}. \tag{20}$$

Substituting (19) and (20) into (16), it follows that

$$b_2 \leq b_2' := \binom{n}{k} \sum_{\ell=2}^{k-1} v_\ell \text{ where } v_\ell = \binom{k}{\ell} \frac{n^{k-\ell}}{(k-\ell)!}\, p\, \frac{w^{m_\ell}}{(k-1)!}. \tag{21}$$

Claim 4 *Assuming that* $w \leq \min\left\{n^{-2/k}, \exp(-\frac{k-1}{k-2})\right\}$, *over* $2 \leq l \leq k-1$, v_ℓ *is maximized at* $\ell = k-1$.

Proof We first claim that v_ℓ is log-convex over $2 \leq \ell \leq k-1$, so that the maximum occurs either at $\ell = 2$ or $\ell = k-1$. For $k = 3$ and 4 this is trivial. Otherwise, for $3 \leq \ell \leq k-2$,

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

and

$$\frac{v_\ell^2}{v_{\ell-1}v_{\ell+1}} = \frac{\ell+1}{\ell} \cdot \left(\frac{k-\ell+1}{k-\ell}\right)^2 w. \tag{22}$$

To establish that v_ℓ is log-convex over $2 \leq l \leq k-1$ it suffices to show that (22) is ≤ 1 over $3 \leq l \leq k-2$. Using $1 + x \leq \exp(x)$, from (22) we have

$$\frac{v_\ell^2}{v_{\ell-1}v_{\ell+1}} \leq w \exp\left\{\frac{1}{\ell} + \frac{2}{k-\ell}\right\} \leq w \exp\left\{\frac{k-1}{k-2}\right\} \leq 1,$$

where the final inequality is by hypothesis and the previous one because $\frac{1}{\ell} + \frac{2}{k-\ell}$ is convex, so its maximum occurs at one of the extremes, either $\ell = 3$ or (in fact) $\ell = k-2$.

Thus, v_ℓ is log-convex and its maximum occurs either at v_2 or v_{k-1}. However,

$$\frac{v_2}{v_{k-1}} = \frac{\binom{k}{2} n^{k-2} w^{\binom{k}{2}-\binom{2}{2}}/(k-2)!}{\binom{k}{k-1} n^1 w^{\binom{k}{2}-\binom{k-1}{2}}/(1!)} = \frac{k-1}{2(k-2)!}(nw^{\frac{1}{2}k})^{k-3} \leq 1,$$

by the hypothesis that $w \leq n^{-2/k}$. Thus, $v_2 \leq v_{k-1}$, proving the claim. □

It follows from (21) and Claim 4 that

$$\begin{aligned} b_2 \leq b_2' &\leq \binom{n}{k}(k-2) \cdot v_{k-1} \leq \binom{n}{k}(k-2) \cdot \binom{k}{k-1} \frac{n^1}{1!} \frac{w^{\binom{k}{2}}}{\binom{k}{2}!} \frac{w^{k-1}}{(k-1)!} \\ &\leq \frac{(k-2)}{\binom{k}{2}!(k-1)!^2}\, n^{k+1} w^{\binom{k}{2}+k-1}. \end{aligned} \tag{23}$$

Recalling (11), using (20), and by analogy with (21),

$$b_1 \leq b_1' := \binom{n}{k} \sum_{\ell=2}^{k} v_\ell' \text{ where } v_\ell' = \binom{k}{\ell} \frac{n^{k-\ell}}{(k-\ell)!}\, p^2. \tag{24}$$

From the definitions of v_ℓ and v'_ℓ in (21) and (24), for $2 \le \ell \le k-1$,

$$\frac{v'_\ell}{v_\ell} = \frac{p}{w^{m_\ell}/\,(k-1)!} = \frac{w^m/\,m!}{w^{m_\ell}/\,(k-1)!} \le \frac{1}{3} w^{\binom{\ell}{2}} \le \frac{1}{3} w \le \frac{1}{3} \exp\left(-\frac{k-1}{k-2}\right) \le \frac{1}{3e}.$$

It follows (referring now to the definitions of b'_2 and b'_1 in (21) and (24)) that the sum of all but the $\ell = k$ term in b'_1, which has no counterpart in b'_2, is at most $\frac{1}{3e} b'_2$. Also, from (24), the $\ell = k$ term of b'_1 is is small compared with its $\ell = k-1$ term:

$$\frac{v'_k}{v'_{k-1}} = \frac{\binom{k}{k} n^0/0!}{\binom{k}{k-1} n^1/1!} = \frac{1}{kn} \le \frac{1}{9}.$$

That is, the last term in the summation for b'_1 is at most $1/9$ times the second-last, therefore it is at most $1/9$ times the sum of all but the last, so that

$$b_1 \le b'_1 \le \left(1 + \frac{1}{9}\right) \frac{1}{3e} b'_2 < \frac{b'_2}{7}. \tag{25}$$

Equation (3) of Theorem 2 follows from (25), (23), and (18) of Theorem 3, since substituting $w = z/n^{2/(k-1)}$ into the term $n^{k+1} w^{\binom{k}{2}+k-1}$ of (23) gives $z^{\binom{k}{2}+k-1}/n$.

5 Strictly balanced H

Let H be a strictly balanced graph with v vertices, m edges, and automorphism group aut(H) of cardinality a. We apply the Stein-Chen method in parallel with Section 2. A bit more care is needed in defining the index set I. Think of a copy of H in G as defined by a set of v vertices of G, together with a 1-to-1 mapping from these vertices to those of H. The set S of vertices is drawn from the collection $\binom{[n]}{v}$, the set of all v-element subsets of $[n]$, with $\left|\binom{[n]}{v}\right| = \binom{n}{v}$. Taking the elements of S in lexicographic order, the mapping into $V(H)$ is given by a permutation π of the values $1 \dots v$, taken modulo the automorphism group of H. Thus we may draw π from a set L of permutations, with

$$|L| = \frac{v!}{a},$$

L consisting of one permutation from each equivalence class.

Then, in analogy with Section 2's equation (7), here

$$I = \binom{[n]}{v} \times L.$$

With no change from before, the number of copies of H of weight $\le w$ is given by a random variable $X = \sum_{\alpha \in I} X_\alpha$, the X_α Bernoulli random variables. In parallel with (8), each X_α has expectation

$$p := E(X_\alpha) = \frac{w^m}{m!},$$

assuming $w \le 1$, and in parallel with (9),

$$\lambda = \lambda(w) := \mathsf{E}X = |I| \cdot p = \binom{n}{v} \frac{v!}{a} \frac{w^m}{m!}. \tag{26}$$

As before, we focus on b_2 and then treat b_1. The structure of dependent events here is a bit subtle, and an example may be useful. Suppose H is the 2-path with edges $\{1,2\}$ and $\{2,3\}$. Its only automorphism is relabeling 123 to 321. On vertices, say $S = \{9, 11, 15\}$ of G, there are then 3 index sets: the set $\{9, 11, 15\}$ in combination with any permutation chosen from $L = \{123, 231, 312\}$, corresponding respectively to paths 9–11–15, 11–15–9, and 15–9–11. (For instance in 231, we take the 2nd, 3rd, and 1st elements of the set in that order.) The permutations 321, 132, and 213 are eliminated from L by automorphism, and correspond respectively to paths 15–11–9, 9–15–11, and 11–9–15 already listed. Suppose α is vertex set $\{9, 11, 15\}$ with permutation 123, giving path 9–11–15. Consider all possible dependent indices β on vertices $\{11, 15, 18\}$. With $\pi = 123$, β gives path 11–15–18, sharing edge 11–15 with α. With $\pi = 312$, β gives path 18–11–15, again sharing edge 11–15 with α. Finally, with $\pi = 231$, β gives path 15–18–11, sharing no edges with α, and thus $\beta \notin B(\alpha)$.

For $\beta \in B(\alpha)$, the pair (α, β) describes an overlapping pair of copies of H: a pair of labeled graphs (V_1, E_1) and (V_2, E_2), each isomorphic to H, $V_1, V_2 \in \binom{[n]}{v}$, with $|E_1 \cap E_2| \geq 1$ and thus $\ell := |V_1 \cap V_2| \geq 2$. Their union is a graph $F = (V_1 \cup V_2, E_1 \cup E_2)$, and for strictly balanced H, $\operatorname{den}(F) > \operatorname{den}(H)$. (This is easy to show, well known, and the reason for introducing strict balance; one early reference is [2, eq(4.3)]. Any such graph F has at most $2v - 2$ vertices, so up to the labeling of the vertices there are only finitely many possibilities. Let

$$d' = d'(H) \tag{27}$$

be the minimum density of all such graphs F. For example, if $H = K_k$ then d' is obtained for two k-cliques sharing $k-1$ vertices, and $d' = \frac{2\binom{k}{2} - \binom{k-1}{2}}{2k-(k-1)} = \frac{(k-1)(k+2)}{2(k+1)} > d = \frac{k-1}{2}$.

For copies α and β both to have weight $\leq w$, their union graph F must have weight $\leq 2w$, and as in (8) this event has probability $(2w)^{m(F)}/m(F)!$, assuming $2w \leq 1$. Thus, if the two copies overlap in ℓ vertices, implying that $m(F) \geq d'(2v - \ell)$, we have

$$\mathsf{E}(X_\alpha X_\beta) \leq \frac{(2w)^{m(F)}}{m(F)!} \leq \frac{(2w)^{d'(2v-\ell)}}{(d'(2v-\ell))!}. \tag{28}$$

It follows that

$$\begin{aligned} b_2 &\leq |I| \sum_{\ell=2}^{v} \binom{v}{\ell} \binom{n-v}{v-\ell} |L| \frac{(2w)^{d'(2v-\ell)}}{(d'(2v-\ell))!} \\ &= O\left(n^v \sum_{\ell=2}^{v} n^{v-\ell} w^{d'(2v-\ell)}\right) \\ &= O\left(\sum_{\ell=2}^{v} (n\, w^{d'})^{2v-\ell}\right) \\ &= O\left((n\, w^{d'})^v\right) = o(1), \end{aligned}$$

the last pair of inequalities holding subject to the condition that

$$w = o(n^{-1/d'}) \quad \text{or equivalently} \quad z = w\, n^{1/d} = o(n^{1/d - 1/d'}). \tag{29}$$

Note that unlike in (17) the sum here includes $\ell = v$ but nonetheless capitalizes on $\beta \neq \alpha$ from the definition of b_2 (see (5)): the vertex sets of β and α may be equal but the index sets themselves are different, so that the union graph F is not isomorphic to H, and therefore $\operatorname{den}(F) \geq d' > d$.

Now compare b_1 and b_2 from their definitions in (4) and (5). For each term (α, β) common to both sums, the summand in b_1 is $p_\alpha p_\beta = p^2 = O(w^{2m})$, and (with $w \le 1$) this is of smaller order than the corresponding summand in b_2 (see (5) and (28)), which is of order $w^{m(F)}$. (F is formed of two copies of H sharing at least one edge, thus $m(F) \le 2m-1$.) Only the terms (α, α) are unique to b_1, within b_1 they are fewer than the other terms, and all terms are equal, so they do not change the order of b_1. It follows that

$$b_1 = O(b_2).$$

We have established that, subject to (29),

$$|\mathsf{P}(W \ge w) - e^{-\lambda}| \le b_1 + b_2 = O\left((n\, w^{d'})^v\right) = o(1). \tag{30}$$

With $z = w\, n^{1/d}$, and using the usual falling-factorial notation, observe from (26) that

$$\lambda = \frac{(n)_v}{a} \frac{w^m}{m!} \sim \frac{n^v}{a} \frac{w^m}{m!} = \frac{z^m}{m!\, a} =: \lambda'.$$

By the intermediate value theorem, there is a point $\lambda'' \in [\lambda, \lambda']$ at which $\frac{d}{d\lambda} e^{-\lambda} = (\exp(-\lambda) - \exp(-\lambda'))/(\lambda - \lambda')$. It follows that

$$\begin{aligned} e^{-\lambda'} - e^{-\lambda} &= (\lambda' - \lambda) \cdot \frac{d}{d\lambda} \exp(-\lambda)\big|_{\lambda = \lambda''} \\ &= o(1)\lambda'' \cdot \exp(-\lambda'') = o(1), \end{aligned} \tag{31}$$

using that both λ and λ are $\lambda''(1+o(1))$ and that $\lambda'' \exp(-\lambda'') \le 1/e$ for any $\lambda'' \ge 0$. Now, in (30) substitute $w = z/n^{1/d}$, yielding

$$\left|\mathsf{P}\left(W \ge \frac{z}{n^{1/d}}\right) - e^{-\lambda'}\right| \le \left|\mathsf{P}\left(W \ge \frac{z}{n^{1/d}}\right) - e^{-\lambda}\right| + \left|e^{-\lambda} - e^{-\lambda'}\right| = o(1). \tag{32}$$

This completes the proof of (1) subject to (29), i.e., for $z = o(n^{1/d - 1/d'})$.

To extend this to all $z = z(n)$, we will observe that there is a weight threshold w_0 (see (33) below) where w_0 is large compared with $n^{-1/d}$ so that a cheap copy of H (cheaper than w_0) is almost certainly present, but w_0 is small compared with $n^{-1/d'}$ so that an overlapping pair of cheap copies is almost certainly not present and thus the error bound $b_1 + b_2$ is small. Values $w < w_0$ are controlled by the previous case, while for values $w > w_0$, (1) holds trivially because all its terms are $o(1)$. The same thresholding around w_0 will be used shortly in proving (2). We now implement this idea.

For any $0 < \alpha < 1$ (throughout, $\alpha = 1/2$ will do), define

$$w_0 = n^{-(1-\alpha)(1/d) - (\alpha)(1/d')} \quad \text{and} \quad z_0 = w_0\, n^{1/d} = n^{\alpha(1/d - 1/d')}. \tag{33}$$

By construction,

$$w_0\, n^{1/d'} = n^{(1-\alpha)(1/d' - 1/d)} = o(1) \tag{34}$$

so that (29) is satisfied, while at the same time

$$z_0 = w_0\, n^{1/d} = n^{(\alpha)(1/d - 1/d')} = \omega(1). \tag{35}$$

A putative counterexample to Theorem 1 equation (1) consists of an infinite sequence $z = z(n)$ for which the error terms are not $o(1)$. Divide such a sequence into two subsequences according to whether $z \leq z_0$ or $z > z_0$. We have just established that the subsequence with $z \leq z_0$ must give error terms $o(1)$, so consider the subsequence with $z > z_0$. Here,

$$\begin{aligned}\left|\mathsf{P}\left(W \geq \frac{z}{n^{1/d}}\right) - \exp\left\{-\frac{z^m}{m!\,a}\right\}\right| &\leq \mathsf{P}\left(W \geq \frac{z_0}{n^{1/d}}\right) + \exp\left\{-\frac{z^m}{m!\,a}\right\} \\ &\leq \left(\exp\left\{-\frac{z_0^m}{m!\,a}\right\} + o(1)\right) + \exp\left\{-\frac{z^m}{m!\,a}\right\} = o(1),\end{aligned}$$

where the application of (32) is justified by (34) and both exponential terms are small because of (35) and $z \geq z_0$. This completes the proof of (1) for all z.

We now turn to Theorem 1, Equation (2). We have

$$\mathsf{E}(W) = \int_0^m \mathsf{P}(W \geq w)\,dw = \int_0^{w_0} \mathsf{P}(W \geq w)\,dw + \int_{w_0}^m \mathsf{P}(W \geq w)\,dw.$$

Setting

$$c = \frac{(n)_v}{a\,m!}$$

and substituting (30) into the first integral (as justified by (34) and (29)) gives

$$\begin{aligned}\mathsf{E}(W) &= \int_0^{w_0} \left(e^{-cw^m} + O\left(\left(n\,w^{d'}\right)^v\right)\right)\,dw + \int_{w_0}^m \mathsf{P}(W \geq w)\,dw \\ &= \int_0^{w_0} e^{-cw^m}\,dw + \int_0^{w_0} O\left(\left(n\,w^{d'}\right)^v\right)\,dw + \int_{w_0}^m \mathsf{P}(W \geq w)\,dw \qquad (36) \\ &= (1+o(1))n^{-1/d}\frac{(a\,m!)^{-1/m}}{m}\,\Gamma\left(\frac{1}{m}\right) + o(n^{-1/d}) + O(\exp(-n^{\Omega(1)})); \qquad (37)\end{aligned}$$

we will prove (37) by considering each of the three integrals in (36) in turn.[1] From (37), Theorem 1, Equation (2) follows immediately.

The first integral in (36) is the principal one. Let $x = cw^m$ so that $w = (x/c)^{1/m}$ and $dw = c^{-1/m}\frac{1}{m}x^{\frac{1}{m}-1}dx$. Then

$$\begin{aligned}\int_0^{w_0} e^{-cw^m}\,dw &= \frac{c^{-1/m}}{m}\int_0^{cw_0^m} e^{-x}\,x^{\frac{1}{m}-1}dx \\ &\sim n^{-1/d}\frac{(a\,m!)^{-1/m}}{m}\,\Gamma\left(\frac{1}{m}\right).\end{aligned}$$

The asymptotic equality above follows from considering the two multiplicands separately. For the first multiplicand, $c^{-1/m}/m$, we just observe that c's term $((n)_v)^{-1/m} \sim n^{-v/m} = n^{-1/d}$. For the second multiplicand, the integral, the upper limit of integration is tending to infinity: $c\,w_0^m$ is of order $n^v w_0^m = (n^{1/d}w_0)^m \to \infty$, by (35). Thus the integral is asymptotic to $\int_0^\infty e^{-x}\,x^{\frac{1}{m}-1}dx$, which is equal to $\Gamma\left(\frac{1}{m}\right)$: it is an example of Euler's integral of the second kind, $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$.

[1] Landau notation does not normally presume the sign of the quantity in question, but in error expressions like (38) and (37) we mean for Ω to denote a positive quantity.

For the second integral in (36),

$$\int_0^{w_0} O\left(\left(n\, w_0^{d'}\right)^v\right)\, dw = O\left(w_0 \cdot (n^{1/d'} w_0)^{d'v}\right)$$

which, from $n^{1/d'} w_0 = o(1)$ by (34) and $v \geq 2$ is

$$= O\left(w_0 \cdot (n^{1/d'} w_0)^{2d'}\right) = O\left(w_0 \cdot (n\, w_0^{d'})^2\right).$$

To show that this is $o(n^{-1/d})$ as claimed in (37) means showing that, when multiplied by $n^{1/d}$, it is $o(1)$. This follows from

$$n^{1/d} \cdot w_0 \cdot (n\, w_0^{d'})^2 = n^{(1/d - 1/d')\,(2\alpha d' + \alpha - 2d')} = n^{-\Omega(1)} = o(1), \tag{38}$$

the final two inequalities holding if $2\alpha d' + \alpha - 2d' < 0$, i.e., if $\alpha < 2d'/(2d'+1)$. Recall from (27) that d' is the density of a graph F describing an overlapping pair of copies of H; say F has m' edges and v' vertices. Since F is connected, $v' \leq m'+1$ and $d' = m'/v' \geq m'/(m'+1)$. Since there is at least one edge shared between the two copies and one edge unique to each copy, $m' \geq 3$, so $d' \geq 3/4$. Thus (38) holds for any $\alpha < (2 \cdot 3/4)/(2 \cdot 3/4 + 1) = 3/5$. Fixing for example $\alpha = 1/2$ (in all other parts of the proof, any α strictly between 0 and 1 will do), the integrated error term is indeed $o(n^{-1/d})$.

For the third integral in (36), while $\mathsf{P}(W \geq w)$ of course decreases with w, it is difficult for us to capitalize on this since our estimates cannot be applied for $w > n^{1/d'}$ where condition (29) is violated. If as for the second integral we reason through $\mathsf{P}(W \geq w) \leq \mathsf{P}(W \geq w_0)$, the estimate is not good enough: we get an expression like (38) but with its integration range of w_0 replaced by $\Theta(1)$, giving $n^{-\frac{1}{d}[(d'-d)(2-2\alpha)-1]}$, and if d' and d are nearly equal the exponent is not negative for any α between 0 and 1.

Claim 5 *For any $0 < \alpha < 1$, with w_0 given by* (33)*, $\mathsf{P}(W > w_0) \leq \exp(-n^{\Omega(1)})$.*

Proof First, we claim that an Erdős–Rényi random graph $G \sim \mathcal{G}(n, w_0)$ contains a copy of H w.p. $> 1/2$. This can be obtained as a classical application of the second-moment method, but it also follows trivially from (32): The set of edges of weight $\leq w_0$ forms a random graph G, the claim is that this subgraph includes a copy of H w.p. $> 1/2$, and (32) says that with even higher probability (namely $1 - o(1)$) there exists such a copy with additional properties (not only is each edge weight $\leq w_0$, but the total is also $\leq w_0$).

In particular, the set of edges of weight $\leq w_0$ forms a random graph $G_1 \sim \mathcal{G}(n, w_0)$, and G_1 contains a copy of H w.p. $> 1/2$.

Now form a second, independent random graph $G_2 \sim \mathcal{G}(n, w_0)$, each of whose edges has weight $\leq 2w_0$, by the following standard trick. For an edge appearing in G_1, accept it into G_2 w.p. w_0. For an edge not appearing in G_1, accept it into G_2 with probability $1 - w_0$ *if* its weight is between w_0 and $2w_0$, and otherwise reject it. Note that in the second case, the weight is in the range $(w_0, 2w_0)$ with probability $w_0/(1-w_0)$, and then we take it only w.p. $1 - w_0$, for a net probability of w_0. Thus, each edge appears in G_2 with probability exactly w_0 independent of G_1 and the other edges.

As an Erdős–Rényi random graph, G_2 contains a copy of H w.p. $> 1/2$, independently of G_1, and in such a copy every edge has weight $\leq 2w_0$.

Repeat this process for graphs $G_3, \ldots, G_k$. With probability $\geq 1-2^{-k}$ at least one of these graphs contains a copy of H, and if so all its edges have weight $\leq kw_0$ for total weight $\leq k\,m\,w_0$.

To get the claim, given α, choose smaller constants $0 < \alpha'' < \alpha' < \alpha$. These give rise to corresponding values $w_0'' < w_0' < w_0$, and the ratios $\Delta'' = w_0'/w_0''$ and $\Delta' = w_0/w_0'$ are both of order $n^{\Omega(1)}$. Given $G_1, G_2, \ldots G_k$ with $k = \Delta''$, we look for H in the k copies of $G(n, w_0'')$: we find such a copy of H w.p. $1 - 2^{-\Delta''} = 1 - \exp(-n^{\Omega(1)})$, and any such copy has weight at most $k \cdot m \cdot w_0'' \leq \Delta'' \cdot \Delta' \cdot w_0'' = w_0$. $\square$

From Claim 5 it is immediate that

$$0 < \int_{w_0}^{m} \mathsf{P}(W \geq w)\, dw \leq m \cdot \mathsf{P}(W \geq w_0) = m \cdot \exp(-n^{\Omega(1)}),$$

and we absorb the constant m into the Ω. This concludes analysis of the third integral in (36), and thus concludes the proof of Theorem 1, Equation (2).

6 Extension of Theorem 1

Theorem 1 extends to distributions other than uniform on $[0, 1]$; such extensions are common in situations where, intuitively, only edges with very small weights are relevant.

Theorem 6 *The conclusions of Theorem 1 hold under the same hypotheses except that now the edge weights are i.i.d. copies of any non-negative random variable X with finite expectation and a continuous distribution function F that is differentiable from the right at 0, with slope $F'(0) = 1$.*

The assumption that $F'(0) = 1$ is without loss of generality. As is standard, it can be extended to a variable X for which $F'(0) = c$, for any $c > 0$, simply by rescaling: applying the theorem to $c\,X$.

To prove the theorem, couple X with a random variable $U = F(X)$. As the quantile of X, U is distributed uniformly on $[0, 1]$.

As in (33), fix $0 < \alpha < 1/2$ and define w_0 accordingly; recall that $w_0 \to 0$ as $n \to \infty$. Any edge of weight $w_U \leq 2w_0 \to 0$ in model U has weight $w_X = F^{-1}(w_U) = w_U\,(1 + o(1))$ in model X, since $F'(w_0) \to 1$. Symmetrically, any edge with weight $w_X \leq 2w_0$ in model X has weight $w_U = F(w_X) = w_X\,(1 + o(1))$ in model U.

Let H_U and H_X denote the lowest-weight copies of H in the two models, and W_U and W_X the corresponding optimal weights. If $W_U \leq w_0$ then $W_X \leq W_U(1 + o(1))$, since H_U would give such an X-weight (each of its constituent edges has weight $w_U \leq w_0$, thus asymptotically equal weight w_X) and the weight of H_X may be even smaller. Taking the same hypothesis not the symmetric one as might be expected, if $W_U \leq w_0$, then $W_X \leq 2\,w_0$ (this is why we introduced the factor of 2), in which case (now symmetrically) $W_U \leq W_X(1 + o(1))$.

Thus, if $W_U \leq w_0$ — call this "event E" — then

$$W_X/L \leq W_U \leq W_X L \tag{39}$$

for some $L = L(w_0) = 1 + o(1))$ in the limit $n \to \infty$ and thus $w_0 \to 0$. Let $\bar{E}$ be the complementary event and recall from Claim 5 that $\mathsf{P}(E) = 1 - \exp(n^{-\Omega(1)})$.

We now prove the distributional result (1) for $X \sim F$. Rewrite (1) as

$$\mathsf{P}(W_U \ge w) = f(w) + o(1), \tag{40}$$

where $f(w) = \exp\left\{-\frac{(wn^{1/d})^m}{m!\,a}\right\}$. Note that if we change w by a factor $L = 1 + o(1)$ then the argument of the exponential changes by a factor $L^m = 1 + o(1)$ and thus, by (31), f changes by an additive $o(1)$, i.e., $f(wL) = f(w) + o(1)$.

Given any w, and conditioning on event E,

$$\begin{aligned}
\mathsf{P}(W_X \ge w \mid E) &\ge \mathsf{P}(W_U/L \ge w) \quad \text{from (39)} \\
&= \mathsf{P}(W_U \ge w\,L) \\
&= f(w\,L) + o(1) \quad \text{from (1) and (40)} \\
&= f(w) + o(1) \quad \text{by the argument below (40).}
\end{aligned}$$

Symmetrically, $\mathsf{P}(W_X \ge w \mid E) \le \mathsf{P}(W_U \ge w/L) = f(w) + o(1)$ and thus $\mathsf{P}(W_X \ge w \mid E) = f(w) + o(1)$. For any event A, and any event E of probability $1 - o(1)$ it holds that $\mathsf{P}(A) = \mathsf{P}(A \mid E) + o(1)$, so here it follows that $\mathsf{P}(W_X \ge w) = \mathsf{P}(W_X \ge w \mid E) + o(1) = f(w) + o(1)$.

This completes the proof of the distributional result. We now prove the expectation result (2) for $X \sim F$.

Let w_0 and event E be as above. Recall from Claim 5 that $\mathsf{P}(\bar{E}) = \exp(-n^{\Omega(1)})$, and from Theorem 1 that $\mathsf{E}(W_U) = \Theta(n^{-1/d})$. By the law of total expectation,

$$\mathsf{E}(W_U) = \mathsf{E}(W_U \mid E)\mathsf{P}(E) + \mathsf{E}(W_U \mid \bar{E})\mathsf{P}(\bar{E})$$

which in this case gives $\mathsf{E}(W_U) = \mathsf{E}(W_U \mid E)(1 + o(1))$ and thus

$$\mathsf{E}(W_U \mid E) = \mathsf{E}(W_U)(1 + o(1)). \tag{41}$$

Also by the law of total expectation,

$$\mathsf{E}(W_X) = \mathsf{E}(W_X \mid E)\mathsf{P}(E) + \mathsf{E}(W_X \mid \bar{E})\mathsf{P}(\bar{E}). \tag{42}$$

As previously established, under event E, $W_X = W_U(1 + o(1))$. It follows that

$$\mathsf{E}(W_X \mid E) = \mathsf{E}(W_U \mid E)\,(1 + o(1)) = \mathsf{E}(W_U)\,(1 + o(1)), \tag{43}$$

where in the second equality we have used (41). In the event $\bar{E}$, W_X lies between 0 and the weight of a prescribed copy of H (say, on vertices $1, \ldots, v$, in that order). We may test for event E by revealing edge weights up to w_0 in model U, so for each edge we know either the exact weight (at most w_0), or know that the weight is $\ge w_0$. In the X model, correspondingly, in the prescribed copy we know the edge weights up to $x_0 = F^{-1}(w_0)$, or that the weight is $\ge x_0$. The expected weight of each edge is larger in the case that it is known to be $\ge x_0$ and, even making this pessimistic assumption for every edge in the prescribed copy, we have

$$0 < \mathsf{E}(W_X \mid \bar{E}) \le m \cdot \mathsf{E}(X \mid X > x_0)) = O(1). \tag{44}$$

(The conditional expectation $\mathsf{E}(X \mid X > x_0)$ cannot be infinite, as then by the law of total expectation the expectation of X itself would be infinite, contradicting our hypothesis.)

Substituting (43) and (44) into (42) gives

$$\mathsf{E}(W_X) = \mathsf{E}(W_U)\,(1 + o(1)) + O(1)\exp(-n^{\Omega(1)}) \sim \mathsf{E}(W_U).$$

This completes the proof of the extended expectation result.

7 Sample Results and Discussion

In this section we discuss the quality of the results provided by Theorem 3, the lower and upper bounds — call them respectively $F^-(w)$ and $F^+(w)$ — on the CDF $F(w)$ of the weight W of a minimum-weight k-clique in a randomly edge-weighted complete graph of order n.

Figure 1 shows F^- and F^+ for $k = 3$, $n = 100$ (left) and $k = 3$, $n = 1,000$ (right). The vertical axis indicates cumulative probability; the horizontal axis indicates w and is given in units of the estimated mean $\hat{\mu}$ given by (2) of Theorem 1. Here $m = \binom{k}{2}$, $d = m/k = (k-1)/2$, and $a = k!$, and by (2) of Theorem 1,

$$\mathsf{E}(W) \sim \hat{\mu} = \frac{1}{n^{2/(k-1)}} \frac{\left(\binom{k}{2}!k!\right)^{1/\binom{k}{2}}}{\binom{k}{2}} \Gamma\left(\frac{1}{\binom{k}{2}}\right).$$

For $k = 3$ this gives $\mathsf{E}(W) \sim \hat{\mu} = \frac{1}{n}\frac{36^{1/3}}{3}\Gamma(1/3) = (1.2878\ldots)\, n^{-1}$.

Looking at Figure 1, for $k = 3$, $n = 100$ we are getting good estimates in the lower tail, mediocre estimates for values of w near the (estimated) mean, and poor estimates in the upper tail. For $k = 3$, $n = 1,000$ we get good results in the lower tail and through the mean, but still poor results in the upper tail.

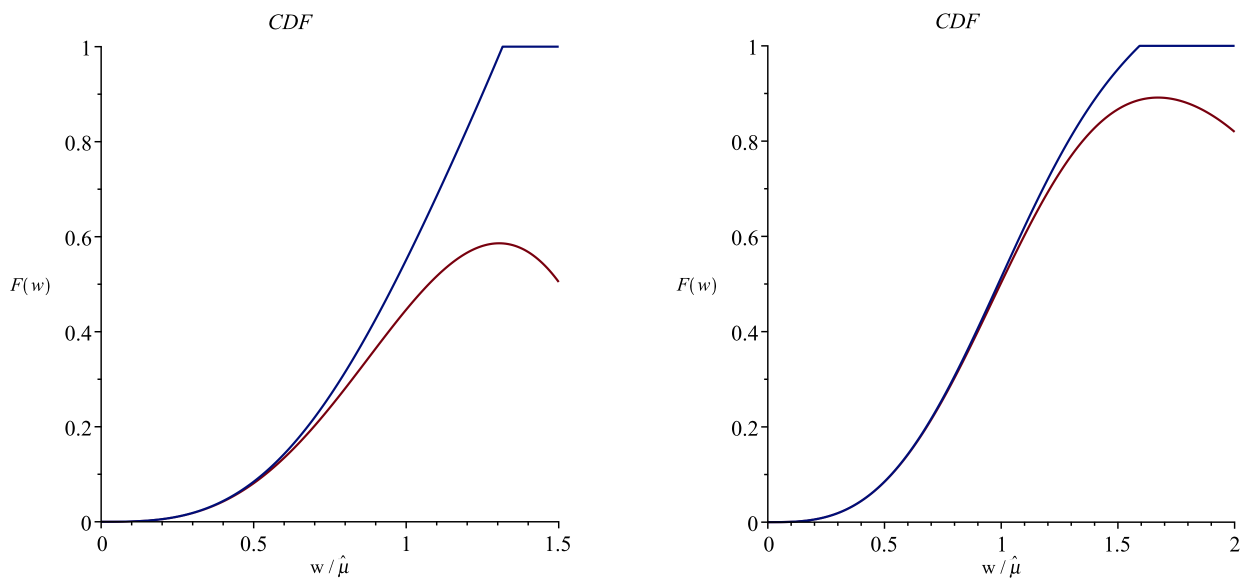

Figure 1: CDFs for $k = 3$, $n = 100$ (left) and $k = 3$, $n = 1,000$ (right).

The probability bounds from the theorem can be less than 0 or greater than 1. In the following discussion and in Figure 1, we have truncated both bounds to the range $[0, 1]$, in particular capping F^+ at 1. Also, because the error terms increase with W, F^- is not monotone increasing. In the following discussion we artificially force it to be (weakly) increasing, by replacing $F^-(w)$ with $\max\{w' \leq w \colon F^-(w')\}$. (To show the nature of the calculated bounds, though, this was not done in Figure 1.)

k	n	0.05	$\hat{\mu}$	LB	UB	0.95	max gap
3	100	0.04862	0.02949	0.44556	0.55215	—	0.41804
3	1,000	0.04986	0.00295	0.50278	0.51387	—	0.11282
3	10,000	0.04999	0.00029	0.50871	0.50983	0.96326	0.02310
3	100,000	0.05	0.00003	0.50931	0.50942	0.95124	0.00403
4	100	0.04470	0.21895	0.31442	0.58856	—	0.66515
4	1,000	0.04947	0.04717	0.45684	0.48193	—	0.19120
4	10,000	0.04995	0.01016	0.47005	0.47236	0.98093	0.03674
4	100,000	0.05	0.00219	0.47127	0.47149	0.95236	0.00594
10	3,000,000	0.04998	0.88550	0.43948	0.43978	0.95220	0.00409
10	10,000,000	0.05	0.67763	0.44100	0.44109	0.95077	0.00131

Table 1: Measurements of the quality of lower and upper bounds on the CDF for various values of k and n. The column "0.05" gives the value of $F^-(w)$ at the w where $F^+(w) = 0.05$. The next three columns give the estimated mean $\hat{\mu}$ of W along with $F^-(\hat{\mu})$ and $F^+(\hat{\mu})$. The column "0.95" gives $F^+(w)$ where $F^-(w) = 0.95$. The column "max gap" is the largest difference, over all w, of $F^+(w) - F^-(w)$; in all cases, it was equal to the gap between the maximum of F^- and 1.

These preliminary observations on Figure 1 suggest a few measures of interest, compiled in Table 1. Let us explain the table and the results observed.

Lower tail tests The most natural application of our results is to perform lower tail tests. It is easy to imagine contexts which would result in smaller-weight cliques than i.i.d. edge weights would produce, for example social networks in which if there is an affinity (modeled as a small weight) between A and B, and an affinity between B and C, then there is likely also to be an affinity between A and C.

If we wish to show that values of W as small as one observed occur with probability less than (say) 5% under the null hypothesis (that weights are i.i.d. uniform $(0, 1)$ random variables), then that observation must be at or below the point w where $F^+(w) = 0.05$. If at this point the lower bound is, say, $F^-(w) = 0.02$, and if the latter happens to be the truth (if F^- rather than F^+ is a good approximation to F here), then we require an observation at the 2% level to demonstrate significance at the 5% level, and this may prevent our doing so.

We therefore take as a measure of lower-tail performance the value of $F^-(w)$ at the point w where $F^+(w) = 0.05$, that is, $F^-(w)|_{F^+(w)=0.05}$. Show in the table column "0.05", if this value is close to 0.05 our estimates have given up little, and this is seen largely to be the case throughout the table.

Of course we may be interested in one-tail significance tests at confidence levels $\alpha < 0.05$ (0.05 being the largest threshold in common use). In this case we would hope for a small gap $(F^+(w) - F^-(w))|_{F^+(w)=\alpha}$ or small ratio $F^+(w)/F^-(w)|_{F^+(w)=\alpha}$. Experimentally, both of these measures appear to be increasing functions of α (i.e., decreasing as α decreases), and thus the high quality of our bounds at $\alpha = 0.05$ implies the same for any $\alpha \leq 0.05$. Our bounds thus appear to be quite useful for lower tail tests.

Mid-range values It is natural to check how good our probability bounds are for typical values of W. Taking the estimated mean $\hat{\mu}$ of W to stand in for a typical value, the table reports $\hat{\mu}$ and the lower and upper bounds $F^-(\hat{\mu})$ and $F^+(\hat{\mu})$ on the CDF $F(w)$ at this point. It can be seen that the quality of this mid-range estimate is poor for $k = 3$, $n = 100$, where the gap $F^+(\hat{\mu}) - F^-(\hat{\mu})$

is above 0.1, but considerably better for $n = 1,000$ and $n = 10,000$, where the gap is only about 0.01 or 0.001 respectively.

Upper tail estimates Our results might also be applied to perform upper tail tests. This would be appropriate for contexts that would produce larger-weight cliques than i.i.d. edge weights would produce, perhaps a social "enmity" network in which if there is enmity (modeled as a small weight) between A and B, and enmity between B and C, then (on the basis that "the enemy of my enemy is my friend") there is likely to be less enmity (larger weight) between A and C.

Our first measure here is the obvious analogue of the lower-tail one: the value of F^+ at the point where F^- is 0.95, $F^+(w)|_{F^-(w)=0.95}$. This is shown in the table in the column "0.95". In many cases F^- never even reaches 0.95, this measure is undefined, and the implication is that it is impossible to establish upper-tail significance with our method even at the 5% confidence level. However, once n is large enough that the measure is defined, larger values of n quickly lead to a small gap $(Fp(w) - F^-(w))|_{F^+(w)=0.95}$: if we can in principle establish upper-tail significance, we can often do so fairly efficiently.

A second measure relevant here is the maximum gap between our lower and upper bounds, $\max_{w>0}(F^+(w) - F^-(w))$. Typically the maximum gap is achieved at the smallest point where $F^+(w) = 1$, and thus if the gap is larger than 0.05 (as for instance for $k = 3$, $n = 1,000$, for which the maximum gap is around 0.11) an upper-tail significance at the 5% level cannot possibly be established. We chose this measure rather than the maximum of F^- because this has the stronger interpretation that our probability estimates are this accurate across the range: for any observed W, we can report lower and upper bounds on the corresponding CDF value (under the null hypothesis) no further apart than this gap.

Parameter values and potential improvements For $k = 3$, values of n as small as 100 give good estimates in the lower tail, and modestly good ones for typical values of W, but no upper tail results. By $n = 10,000$, results are good across the range, with a maximum gap of around 0.02; with $n = 100,000$ this decreases to 0.004.

For $k = 4$ there is a similar pattern, with only slightly less sharp results for the same n.

For $k = 10$ the picture is significantly different. Recall that our methods restrict us to estimating $F(w)$ for $w \leq 1$ and here that leaves us hopelessly far into the left tail. With $k = 10$ and $n = 100,000$, the estimated mean of W is $\hat{\mu} = 1.8856$, while $F^+(1)$ is less than 10^{-12}. For $n = 1,000,000$, with $w \leq 1$ we still cannot access the estimated mean $\hat{\mu} = 1.13036$, and $F^+(1) \sim 0.00231$: our methods would be useful for observations anomalously small at the 0.1% confidence level (to name a standard value near 0.00231), but nothing much above that.

However, with $k = 10$, W is concentrated near $\hat{\mu}$, and thus, *once n is large enough that $\hat{\mu}$ falls below 1*, our methods give good results across the range, as shown in the table. So, for $k = 10$, the problem we observe is with the *range of validity* of our estimates rather than their quality. If Theorem 3 is extended as outlined in Section 3.1, the results might well be adequately tight.

One other weakness of our methods is that the lower bound F^- falls significantly short of 1 for $k = 3$ and $k = 4$ with $n = 100$ and $n = 1,000$, revealed in the table's large "maximum gap" measures. It might be possible to improve this by applying the method used in proving Claim 5, but we have not attempted this.

8 Conclusions

The object studied in this work is the distribution of a minimum-weight clique, or copy of a strictly balanced graph H, in a complete graph G with i.i.d. edge weights. Theorem 1 provides asymptotic characterizations of the distribution and its mean for any strictly balanced graph H, while Theorems 2 and 3 provide explicit (non-asymptotic) descriptions of the distribution for cliques.

This distribution is a natural object of mathematical study, but also likely to have practical relevance, particularly for statistical determination that a given network's weights are *not* i.i.d. We look forward to seeing such applications of the work. Some potential applications would involve networks that are not complete graphs, and extending our results to such cases seems challenging, whether by extending to other infinite graph classes of graphs or by including a particular graph as part of the input.

As presented, our explicit methods are for cliques, the uniform distribution and clique weights at most 1. However, as discussed in Section 3, it is easy to write down calculations for the uniform distribution and all clique weights, the exponential distribution, and probably other common distributions. In doing so we encountered computational challenges, but these seem surmountable if the incentive is more than just fleshing out a table. As noted in Section 3.3, extending explicit results to subgraphs H other than cliques is straightforward.

The quality of the results from our methods, as discussed in Section 7, is largely good, especially when we are interested in lower-tail results and relatively small cliques, or larger cliques in very large graphs. To improve the probability bounds in the middle range, near the mean, would seem to require an approach other than Stein-Chen, but we have no concrete alternative suggestions. Improving the bounds in the upper tail might be done, as suggested earlier, by applying the ideas of Claim 5.

References

[1] R. Arratia, L. Goldstein and L. Gordon, Two Moments Suffice for Poisson Approximations: The Chen-Stein Method, *Annals of Probability* 17 (1989) 9-25.

[2] B. Bollobás, Threshold Functions for Small Subgraphs, *Mathematical Proceedings of the Cambridge Philosophical Society*, 90 (1981) 197–206.

[3] I. Bomze, M. Budinich, P. M. Pardalos, M. Pelillo, The Maximum Clique Problem, in *Handbook of Combinatorial Optimization*, eds. D.-Z. Du and P. M. Pardalos, Kluwer, 1999.

[4] S. Cai and J. Lin, Fast Solving Maximum Weight Clique Problem in Massive Graphs, *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16).*

[5] Clique problem, *Wikipedia*, revised 20 May 2017 08:40 UTC.

[6] Erlang distribution, *Wikipedia*, revised 2 June 2017 20:04 UTC.

[7] P. Hall, The distribution of means for samples of size N drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable, *Biometrika*, XIX (1927) 240–244.

[8] J. O. Irwin, On the frequency distribution of the means of samples from a population having any law of frequency with finite monments, with special reference to Pearson's Type II, *Biometrika*, XIX (1927) 225–239.

[9] N. L. Johnson, S. Kotz, N. Balakrishnan, eds., Continuous Univariate Distributions Vol. 1, 2nd ed., Wiley, 1995.

[10] N. L. Johnson, S. Kotz, N. Balakrishnan, eds., Continuous Univariate Distributions Vol. 2, 2nd ed., Wiley, 1995.

[11] M. G. Ravetti and P. Moscato, Identification of a 5-protein biomarker molecular signature for predicting Alzheimer's disease, *PLoS ONE*, 3(9):e3111, 2008.