# Connectivity of random cubic sum graphs

Andrew Beveridge*
Department of Mathematics, Statistics and Computer Science
Macalester College
Saint Paul, MN 55105

June 1, 2010

**Abstract**

Consider the set $\mathcal{SG}(Q_k)$ of all graphs whose vertices are labeled with non-identity elements of the group $Q_k = \mathbb{Z}_2^k$ so that there is an edge between vertices with labels $a$ and $b$ if and only if the vertex labeled $a + b$ is also in the graph. Note that edges always appear in triangles, since $a + b = c$, $b + c = a$ and $a + c = b$ are equivalent statements for $Q_k$. We define the *random cubic sum graph* $\mathcal{SG}(Q_k, p)$ to be the probability space over $\mathcal{SG}(Q_k)$ whose vertex sets are determined by $\Pr[x \in V] = p$ with these events mutually independent. As $p$ increases from 0 to 1, the expected structure of $\mathcal{SG}(Q_k, p)$ undergoes radical changes. We obtain thresholds for some graph properties of $\mathcal{SG}(Q_k, p)$ as $k \to \infty$. As with the classical random graph, the threshold for connectivity coincides with the disappearance of the last isolated vertex.

**Key words.** Random graph, sum graph.

**AMS Subject Classifications.** 05C80, 05C78.

## 1 Introduction

The study of random graphs began with the landmark papers of Erdős-Rényi [10, 11]. They introduced the (uniform) random graph $\mathcal{G}(n, p)$, which is the probability space over every graph $G$ on the vertex set $[n]$ where $\Pr[\{i, j\} \in G] = p$, with these events mutually independent. In recent years, there has been much interest in alternative random graph models, particularly those inspired by so-called real world networks such as social networks, the world wide web and wireless networks. This has lead to the development of random graph models with dependencies between the appearance of edges. Some models achieve a prescribed (exact or expected) degree sequence, such as the configuration model [17] and the preferential attachment model [4, 7, 8]. Other models reflect an underlying geometric structure, such as random geometric graphs [18]. In this paper, we study a random graph structure that obeys an underlying *algebraic* structure.

Introduced by Frank Harary [14], a *sum graph* is a graph $G$ with a labeling $\ell : V(G) \to \mathbb{Z}^+$ such that $(x_i, x_j) \in E(G)$ if and only if there exists $x_k \in V(G)$ such that $\ell(x_i) + \ell(x_j) = \ell(x_k)$. Sum

*Supported in part by NSA Young Investigator Grant H98230-08-1-0064.

graphs have been well studied. A typical result identifies the minimum number of isolated vertices that must be added to a graph so that the resulting structure has a sum graph labeling [5, 9, 13, 16]. Sum graphs have potential applications for data compression: specifically for compressed storage of graphs. They can also be used for secure communication of information [19]. See [12] for a summary of known results on sum graphs.

Many generalizations of sum graphs have also been studied. A graph is a *mod sum graph* [6] if the labels are drawn from $\mathbb{Z}_m$. Another generalization to $f$-graphs [2] replaces the sum $x_i + x_j$ with an arbitrary symmetric polynomial $f(x_i, x_j)$. Consider a further generalization of mod sum graphs. Let $H$ be any group with identity element $e$ and let $H^* = H \backslash \{e\}$ denote all the non-identity elements of $H$. An *$H$-sum graph* is a directed graph with a labeling $\ell : V(G) \to H^*$ such that $(x_i, x_j) \in E(G)$ if and only if there exists $x_k \in V(G)$ such that $\ell(x_i) + \ell(x_j) = \ell(x_k)$ for some $k$. We exclude $e$ as a potential label since this vertex would (inconveniently) be adjacent to all other vertices. An $H$-sum graph is undirected if and only if $H$ is abelian. We denote the set of all $H$-sum graphs by $\mathcal{SG}(H)$. Abusing notation slightly, we will identify the elements of $H^*$ with the set of potential vertices so that $V(G) \subset H^*$. When clarification is needed, we will refer to $x \in H^*$ as "the element $x$" and $x \in V(G)$ as "the vertex $x$."

We define the *random $H$-sum graph* $\mathcal{SG}(H, p)$ to be the probability space over the set of $H$-sum graphs whose vertex sets $V \subset H^*$ are determined by $\Pr[x \in V] = p$ with these events mutually independent. For convenience, we define $n := |H^*| = |H| - 1$. For $G \in \mathcal{SG}(H)$ with $v(G) = r$, the probability of obtaining $G$ via this process is $p^r(1-p)^{n-r}$. Note that the labeled vertices are independent but the induced edges are not.

A second model for a random $H$-sum graph is to specify the desired number of vertices $r \leq |H^*|$ and then choose a random $r$-subset of $H^*$. The resulting probability space is denoted by $\mathcal{SG}(H, r)$. These two models are quite compatible. Indeed, the number of vertices in $\mathcal{SG}(H, p)$ is governed by the binomial distribution $\text{Bin}(n, p)$. The following Chernoff bounds (cf. [15], p. 26) will be useful here (and in the sequel). If $X \in \text{Bin}(n, p)$ and $E[X] = \mu = np$ then with $\phi(x) = (1+x)\log(1+x) - x$, $x \geq -1$,

$$\Pr[X \geq (1+\delta)\mu] \leq \exp\left(-\mu\phi\left(\delta\right)\right), \qquad t \geq 0; \tag{1}$$

$$\Pr[X \leq (1-\delta)\mu] \leq \exp\left(-\mu\phi\left(-\delta\right)\right), \qquad t \geq 0. \tag{2}$$

Let the acronym *whp* (with high probability) mean "with probability $1 - o(1)$ as $n \to \infty$." If $p = r/n$ and $G \in \mathcal{SG}(H, p)$ then for any fixed arbitrarily small $\delta > 0$ these bounds imply that

$$(1-\delta)r = (1-\delta)np \leq E[v(G)] \leq (1+\delta)np = (1+\delta)r \tag{3}$$

whp provided that $r \to \infty$ and $n - r \to \infty$.

Since these two models are roughly equivalent we will focus on the model $\mathcal{SG}(H, p)$, appealing to the model $\mathcal{SG}(H, r)$ when appropriate. As with the classical random graph, we study the threshold functions for random $H$-sum graphs. These results should be viewed in the context that $H$ is a finite group that is part of a sequence of finite groups with $|H|$ growing to infinity. Fixing a graph theoretic property $\mathcal{A}$, we can determine the threshold for $\lim_{|H^*|\to\infty} \Pr[\mathcal{SG}(H,p) \text{ satisfies } \mathcal{A}]$.

Herein we restrict ourselves to the family of cubic groups $Q_k = \mathbb{Z}_2^k$, whose edge dependencies remain manageable. We define $n = n(k) = |Q_k^*| = 2^k - 1$ to be the number of possible vertices in our graph. For convenience, we adopt the composition notation $ab$ rather than the addition notation $a+b$. Furthermore, we denote the edge joining vertices $a$ and $b$ as $(a,b)$ to avoid confusion with the vertex labeled $ab$. Note that the edges of $G \in \mathcal{SG}(Q_k,p)$ always appear in triangles: $ab = c$, $bc = a$ and $ac = b$ are equivalent statements for $a, b, c \in Q_k$. We refer to the set $\{a, b, ab\}$ as an *elementary triangle*. Finally for any $v \in V(G)$, $E[\deg(v) \mid v \in V(G)] = 2 \cdot E[\text{Bin}((n-1)/2, p^2)] = (n-1)p^2$ and the Chernoff bounds guarantee tight concentration around this mean value when $p(n) = \omega(\sqrt{1/n})$.

As an aside, we can view a cubic sum graph as a 3-regular hypergraph by considering the elementary triangle $\{a, b, ab\}$ as a hyperedge rather 3 separate edges (since they can only appear together). In the sum graph corresponding to the full vertex set $Q_k^*$, we find that every pair $a, b$ of vertices appears in the unique triple $\{a, b, ab\}$. In other words, we have a Steiner system of order $2^k - 1$.

We characterize the thresholds for the following graph properties. We start with the very early stages of the random process.

**Theorem 1** *Let $p = c(n)/n^{2/3}$ where $0 \le c(n) = o(\log n)$. Then*

$$\Pr[\mathcal{SG}(Q_k,p) \text{ contains an elementary triangle}] \to \begin{cases} 0 & \text{if } c(n) \to 0, \\ e^{-c^3/6} & \text{if } c(n) \to c, \\ 1 & \text{if } c(n) \to \infty, \end{cases}$$

*where $c$ is an arbitrary constant. Moreover, if $c(n) \to c$ then the distribution governing the number of elementary triangles converges to* $\text{Po}(c^3/6)$.

In other words, whp $\mathcal{SG}(Q_k,r)$ contains only isolated vertices until $r \approx n^{1/3}$. The next graph structure to appear is a *butterfly* which consists of two triangles that share a vertex. The vertex set of a butterfly has the form $\{a, b, ab, c, ac\}$. Butterflies appear in $\mathcal{SG}(Q_k,r)$ when $r \approx n^{2/5}$.

**Theorem 2** *Let $p = c(n)/n^{3/5}$ where $0 \le c(n) = o(\log n)$. Then*

$$\Pr[\mathcal{SG}(Q_k,p) \text{ contains a butterfly}] \to \begin{cases} 0 & \text{if } c(n) \to 0, \\ e^{-c^5/8} & \text{if } c(n) \to c, \\ 1 & \text{if } c(n) \to \infty, \end{cases}$$

*where $c$ is an arbitrary constant. Moreover, if $c(n) \to c$ then the distribution governing the number of butterflies converges to* $\text{Po}(c^5/8)$.

Moving forward in the random process, we determine the threshold for the disappearance of isolated vertices.

**Theorem 3** *Consider $\mathcal{SG}(Q_k, p)$ for $p = \sqrt{\frac{\log n + \log\log n + c(n)}{n}}$ where $|c(n)| = o(\log\log n)$. Then*

$$\Pr[\mathcal{SG}(Q_k, p) \text{ contains an isolated vertex}] \to \begin{cases} 1 & \text{if } c(n) \to -\infty, \\ e^{-e^{-c/2}} & \text{if } c(n) \to c, \\ 0 & \text{if } c(n) \to \infty, \end{cases}$$

*where $c$ is an arbitrary constant. More specifically, if $c(n) \to c$ then the distribution governing the number of isolated vertices converges to $\mathrm{Po}(e^{-c/2})$.*

We can view a random cubic sum graph as a dynamic evolution over time. For each element $a \in Q_k^*$ let $y_a$ be selected uniformly and independently from $[0, 1]$. We increase $p$ from 0 to 1 and include the vertex $a$ when $p \geq y_a$. The graph evolves as we increase from $p = 0$ (where the graph has no vertices) up to $p = 1$ (where the graph is $K_n$). For $0 < p < 1$, the graph of all included vertices has distribution $\mathcal{SG}(Q_k, p)$. Alternatively, we can view this evolution as a random process. A *sum graph process* is a sequence $\tilde{G} = (G_t)_{t=1}^n$ where $G_t$ is a sum graph on $t$ vertices, $V(G_t) \subseteq Q_k^*$, and $G_{t-1} \subset G_t$ for $1 \leq t \leq n$. Using the uniform distribution on the set of all such processes gives the *random sum graph process* $\widetilde{\mathcal{SG}}$. In this interpretation, $G_t \in \tilde{G}$ has distribution $\mathcal{SG}(Q_k, t)$.

As with the classical random graph, the threshold for the disappearance of the last isolated vertex coincides with the threshold for connectivity. However, the connectivity of our sum graph process is more complicated than for the classic model. Indeed, once the classic random graph process becomes connected, it remains connected. The connectivity of a random sum graph may oscillate. Consider the following simple example: if our first four included vertices are $a, b, ab, c$, Then the graph becomes connected at $t = 3$ and disconnected again at $t = 4$. Our next theorem ensures that once the random sum graph becomes connected, it remains connected whp.

If $Q(G)$ is a graph property, then the *hitting time* $\tau$ at which $Q$ appears is $\tau(Q(\tilde{G})) = \min\{t \geq 0 \mid G_t \text{ has } Q\}$ whp where $\tilde{G} \in \widetilde{\mathcal{SG}}$. This hitting time formulation allows us to simply express the relationship between minimum degree $\delta(\tilde{G})$ and connectivity $\kappa(\tilde{G})$.

**Theorem 4** *Consider $\tilde{G} = (G_t)_{t=1}^n$, and let $1 \leq t_1 \leq t_2$. If $\delta(G_{t_1}) > 0$ then $\delta(G_{t_2}) > 0$ whp. If $\kappa(G_{t_1}) > 0$ then $\kappa(G_{t_2}) > 0$ whp. Finally, $\tau(\delta(\tilde{G}) > 0) = \tau(\kappa(\tilde{G}) > 0)$ whp.*

Consider the following algebraic interpretation of the connectivity threshold of $\mathcal{SG}(Q_k, p)$. Suppose that we choose a random subset $S \subset Q_k^*$ of size $r$. How large must $r$ be to ensure that for any two elements $a, b \in S$, whp there exists some subset $\{c_1, c_2, \ldots, c_t\} \subset S$ so that $a = b + \sum_{i=1}^t c_i$? The answer is given by Theorem 4: we must take $r \geq \sqrt{n(\log n + \log\log n + \omega(n))}$.

We compare the threshold of connectivity for the random cubic sum graph with that of the classical Erdős-Rényi random graph model. The threshold for connectivity for $\mathcal{SG}(Q_k, p)$ occurs when $p = \Theta(\sqrt{\log n / n})$. This graph will have $\Theta(\sqrt{n \log n})$ vertices and $\Theta(\sqrt{n(\log n)^3})$ edges. For comparison, the connectivity threshold for the Erdős-Rényi random graph $\mathcal{G}(n', p')$ occurs at

$p' = \log n'/n'$. In order to have graphs of similar sizes, we take $n' = \sqrt{n \log n}$, giving a comparable connectivity threshold of $p' = \Theta(\sqrt{\log n/n})$. At this point, the classical random graph also has $\Theta(\sqrt{n(\log n)^3}\,)$ edges. While the size of the edges sets are similar, the structure of these random graphs are not. For example, every vertex of the random cubic sum graph has even degree and is contained in at least one cycle.

Finally we remark that there are many open questions concerning random cubic sum graphs. The two most natural problems would be to identify the threshold for the emergence of a giant component and to characterize the diameter for various values of $p$ above the threshold for connectivity.

## 2 Random Cubic Sum Graphs for Small $p$

In this section, we prove Theorems 1 and 2. They characterize the thresholds for the appearance of our earliest nontrivial graph structures: elementary triangles and butterflies. We start with an intuitive motivation each threshold. For a given $p$, the expected number of elementary triangles is $\frac{1}{3}\binom{n}{2}p^3$: an elementary triangle is of the form $\{a, b, ab\}$, so choosing a pair of elements determines the triangle. Each elementary triangle can be chosen in three different ways. This calculation leads to our threshold function of $p = n^{-2/3}$. The expected number of butterflies is $\frac{1}{8}n(n-1)(n-3)p^5$. Indeed, we have $n$ choices for the shared vertex, $\frac{n-1}{2}$ ways to complete the first triangle and then $\frac{n-3}{2}$ ways to create the second triangle. We must multiply by an additional $\frac{1}{2}$ since the order we choose these triangles does not matter. This gives an intuitive threshold function of $p = n^{-3/5}$ for the appearance of butterflies.

To prove each theorem, we need both the Janson Inequality and the method of moments. Let $\{B_i\}_{i\in I}$ be a set of random events with all $\Pr[B_i] \le \epsilon$. Set $\triangle = \sum_{\tau_i \sim \tau_j} \Pr[B_i \wedge B_j]$ where we sum over all unordered pairs $i, j$ and $M = \prod_{i\in I} \Pr[\overline{B_i}]$. The Janson Inequality (c.f. Theorem 8.1.1 in [3]) states that $M \le \Pr[\wedge_{i\in I}\overline{B_i}] \le M \exp\left(\frac{1}{1-\epsilon}\frac{\triangle}{2}\right)$.

Now explicitly consider the events $\{B_{n,i}\}_{i\in I}$ as part of a sequence depending on $n \to \infty$. Let $X_{n,i}$ be the indicator function for the event $B_{n,i}$ and set $X_n = \sum_{i\in I} X_{n,i}$. The method of moments (c.f. Corollary 6.8 in [15]) states that if $\lambda \ge 0$ is such that for every $k \ge 1$, as $n \to \infty$ we have $E[X_n]_k = \sum^*_{i_1,\ldots,i_k} \Pr[I_{n,i_1} = \cdots = I_{n,i_k} = 1] \to \lambda^k$ (where this sum is taken over all distinct ordered $k$-tuples $i_1, \ldots, i_k$) then $X_n$ converges to $\mathrm{Po}(\lambda)$ in distribution.

**Proof of Theorem 1.** For $G \in \mathcal{SG}(Q_k, p)$, let $A_\tau$ be the event that the elementary triangle $\tau$ is present. We write $\tau_i \sim \tau_j$ if $\tau_i \ne \tau_j$ and the events $A_{\tau_i}, A_{\tau_j}$ are not independent and set $\triangle = \sum_{\tau_i\sim\tau_j} \Pr[A_{\tau_i} \wedge A_{\tau_j}]$ where we sum over all unordered pairs $i, j$.

We have $\Pr[A_\tau] = p^3 = c(n)^3/n^2 = o(1)$ and $M = \prod_\tau \Pr[\overline{A_\tau}] = (1-p^3)^{n(n-1)/6}$. Furthermore, $\triangle = \frac{1}{3}\binom{n}{2}p^3 \cdot \frac{3}{2}(n-3)p^2 \le n^3p^5 = c(n)^5/n^{1/3} = o(1)$. Indeed, each of the $\frac{1}{3}\binom{n}{2}$ elementary triangles is present with probability $p^3$. The dependence $A_{\tau_i} \sim A_{\tau_j}$ holds if and only if $|A_{\tau_i} \cap A_{\tau_j}| = 1$.

We have 3 choices for this common vertex and $(n-3)/2$ choices for the additional pair. Since $\Pr[A_\tau] = o(1)$ and $\triangle = o(1)$, the Janson Inequality gives $\Pr[\wedge_\tau \overline{A}_\tau] \to M \to \exp(-c(n)^3/6)$. In particular, when $c(n) \to 0$, there are no triangles whp and when $c(n) \to \infty$, there are triangles whp.

We analyze the case $c(n) \to c$ using the method of moments. Let $X_\tau$ denote the indicator function for the event $A_\tau$ and set $X = \sum_\tau X_\tau$ where the sum is taken over all possible elementary triangles. We show that $E[X]_k \to (c^3/6)^k$ for all $k \geq 1$. We have $E[X]_1 = \sum_\tau \Pr[X_\tau = 1] = \frac{1}{3}\binom{n}{2}p^3 \to c^3/6$. For $k \geq 2$, let $B$ be the set of all ordered $k$-tuples $(\tau_1, \tau_2, \ldots, \tau_k)$ of elementary triangles. Partition $B$ into $B_1 = \{(\tau_1, \tau_2, \ldots, \tau_k) \mid \tau_i \cap \tau_j = \emptyset \text{ for all } i, j\}$ and $B_2 = B \backslash B_1$. We have

$$\sum_{(\tau_1, \ldots \tau_k) \in B_1} \Pr[X_{\tau_1} = \cdots = X_{\tau_k} = 1] = \prod_{i=0}^{k-1} \left( \frac{1}{3} \frac{(n-3i)(n-6i-1)}{2} \right) p^3 \to \left( \frac{c^3}{6} \right)^k.$$

When choosing our $(i+1)$th isolated triangle, we have already chosen $3i$ vertices. After choosing one of the remaining $n - 3i$ vertices, we have $n - 6i - 1$ choices for our second vertex to ensure that our triangle is disjoint from the previous $i$ triangles. The order we choose these two vertices does not matter, and each triangle is chosen in three different ways.

To complete the proof, we show that the contribution from $k$-tuples in $B_2$ tends to zero. For any $k$-tuple of distinct elementary triangles $(\tau_1, \ldots \tau_k)$ with $k \geq 2$, we have

$$\sum_\tau \Pr[X_\tau = 1 \mid X_{\tau_1} = \cdots = X_{\tau_k} = 1] \leq \left( \binom{3k}{3} + \binom{3k}{2} p + \binom{3k}{1} \frac{n}{2} p^2 + \frac{1}{3} \binom{n}{2} p^3 \right) = O(1)$$

where the sum is taken over all elementary triangles $\tau \notin \{\tau_1, \ldots, \tau_k\}$. The four terms in the upper bound correspond to the number of vertices that $\tau$ shares with the previous $k$ triangles. Fixing $\tau_1, \tau_2$ and using induction gives

$$\sum_{\tau_3, \ldots, \tau_k} \Pr[X_{\tau_1} = \cdots = X_{\tau_k} = 1] = O(\Pr[X_{\tau_1} = X_{\tau_2} = 1])$$

where the sum is taken over all possible $k-2$ tuples of elementary triangles.

For each $(\tau_1, \ldots, \tau_k) \in B_2$, there is at least one pair of these elementary triangles that share a vertex. Let $B_2' \subset B_2$ denote the subset such that $|\tau_1 \cap \tau_2| = 1$, so that the first pair of triangles in our $k$-tuple create a butterfly. We have $|B_2| \leq k^2 |B_2'|$ and $\Pr[X_{\tau_1} = X_{\tau_2} = 1] = p^5$. This leads to our bound

$$\begin{aligned}
\sum_{(\tau_1, \ldots \tau_k) \in B_2} \Pr[X_{\tau_1} = \cdots = X_{\tau_k} = 1] &\leq k^2 \sum_{(\tau_1, \ldots \tau_k) \in B_2'} \Pr[X_{\tau_1} = \cdots = X_{\tau_k} = 1] \\
&= O\left( \sum_{\tau_1, \tau_2 : |\tau_1 \cap \tau_2| = 1} \Pr[X_{\tau_1} = X_{\tau_2} = 1] \right) \\
&= O(\triangle) = O\left(n^3 p^5\right) = O(n^{-1/3}) \to 0.
\end{aligned}$$

□

**Proof of Theorem 2.** After isolated triangles, the next graph structure to appear is a butterfly of the form $\beta = \{a, b, ab, c, ac\}$. Let $A_\beta$ be the event that butterfly $\beta$ appears in the graph, so $\Pr[A_\beta] = p^5 = c(n)^5/n^3 = o(1)$ and $M = \prod_\beta \Pr[\overline{A_\beta}] = (1-p^5)^{n(n-1)(n-3)/8}$.

Calculating $\triangle$ is more involved. Events $A_{\beta_1}$, $A_{\beta_2}$ are only dependent when $\beta_1 \cap \beta_2 \neq \emptyset$. Let $\beta_1 = \{a, b, ab, c, ac\}$ and $\beta_2 = \{x, y, xy, z, xz\}$ be butterflies with $|\beta_1 \cap \beta_2| = k$ where $1 \leq k \leq 4$, so that $\Pr[A_{\beta_1} \wedge A_{\beta_2}] = p^{10-k}$. As calculated at the start of this section, we may choose $\beta_1$ in $n(n-1)(n-3)/8 < n^3/8$ distinct ways. Fixing $\beta_1$, let $U_k$ be the set of all butterflies intersecting $\beta_1$ in exactly $k$ vertices, $1 \leq k \leq 4$.

For $\beta_2 \in U_1$, we have 5 choices for this common vertex $u \in \beta_1$ and $(n-5)/2$ ways to complete the first triangle in $\beta_2$. One of these first 3 vertices must the center of $T$, and we complete the butterfly in $(n-7)/2$ ways. Therefore $|U_1| \leq 15(n-5)(n-7)/4 < 4n^2$.

If $\beta_2 \in U_2$ then there are $\binom{5}{2}$ ways to determine $\{x, y\} = \beta_1 \cap \beta_2$, and we pick an arbitrary third vertex $z$ in $n-5$ ways. There are at most 6 possibly distinct butterflies we can construct from these vertices, each of which is entirely determined by our choice of center from among $\{x, y, z, xy, yz, xz\}$, and so $|U_2| \leq 60(n-5) < 60n$.

Next, we claim that if $T \in U_3$ then $a \in \beta_2$. Indeed, otherwise we can assume that $\beta_1 \cap \beta_2 = \{b, c, ab\}$ without loss of generality. Since $a \notin \beta_2$, either $\beta_2 = \{b, c, bc, ab, ac\}$ or $\beta_2 = \{b, ac, abc, c, ab\}$. In either case, $|\beta_1 \cap \beta_2| = 4$, a contradiction. A similar argument shows that if $\beta_2 \in U_3$ then either $\beta_1 \cap \beta_2 = \{a, b, ab\}$ or $\beta_1 \cap \beta_2 = \{a, c, ac\}$. In each case, we have three choices for the center vertex and $(n-5)/2$ choices for the final two vertices. So $|U_3| \leq 3(n-5) < 3n$.

Finally, we claim that $|U_4| = 2$. Indeed, if $\beta_2 \in U_4$ then $a \notin \beta_2$ (otherwise this forces $\beta_2 = \beta_1$), giving only two possible choices for $\beta_2$: $\{bc, b, c, ac, ab\}$ and $\{abc, b, ac, c, ab\}$.

Putting this all together for $p = \omega n^{-3/5}$, we have

$$\begin{aligned}\triangle &= \sum_{i \sim j} \Pr[A_{\beta_i} \wedge A_{\beta_j}] \leq \frac{1}{2} \cdot \frac{n^3}{8} \sum_{k=1}^{4} |U_i|\, p^{10-k} \leq \frac{n^3}{16}\left(4n^2p^9 + 20np^8 + 3np^7 + 2p^6\right) \\ &\leq \frac{1}{16}\left(4c(n)^9 n^{-2/5} + 20c(n)^8 n^{-4/5} + 3c(n)^7 n^{-1/5} + 2c(n)^6 n^{-3/5}\right) = O(c(n)^7 n^{-1/5}) = o(1).\end{aligned}$$

Because $\Pr[A_\tau] = o(1)$ and $\triangle = o(1)$, the Janson Inequality gives $\Pr[\wedge_\beta \overline{A}_\beta] \to M \to \exp(-c(n)^5/8)$. In particular, when $c(n) \to 0$, there are no butterflies whp and when $c(n) \to \infty$, there are butterflies whp.

Our analysis for $c(n) \to c$ is similar to the previous proof, and we sketch the argument. Let $X_\beta$ denote the indicator function for the event $A_\beta$ and set $X = \sum_\beta X_\beta$ where the sum is taken over all possible butterflies. We have $E[X]_1 = \sum_\beta \Pr[X_\tau = 1] = \frac{1}{8} n(n-1)(n-3)p^5 \to c^5/8$. For $k \geq 2$, let $B$ be the set of all ordered $k$-tuples $(\beta_1, \ldots, \beta_k)$ of distinct butterflies. Partition $B$ into

$B_1 = \{(\beta_1, \ldots, \beta_k) \mid \beta_i \cap \beta_j = \emptyset \text{ for all } i, j\}$ and $B_2 = B \backslash B_1$. We have

$$\sum_{(\beta_1,\ldots\beta_k)\in B_1} \Pr[X_{\beta_1} = \cdots = X_{\beta_k} = 1] = \left(\frac{n^3}{8} p^5 (1 - o(1))\right)^k \to \left(\frac{c^5}{8}\right)^k.$$

As in the previous proof, the contribution from $k$-tuples in $B_2$ tends to zero. For any $k$-tuple of distinct butterflies $(\beta_1, \ldots \beta_k)$ with $k \geq 2$, we have $\Pr[X_\tau = 1 \mid X_{\tau_1} = \cdots = X_{\tau_k} = 1] = O(1)$. As before, the two dominant terms correspond to choosing $\beta \subset \cup_{i=1}^k \beta_i$ and choosing $\beta \cap \left(\cup_{i=1}^k \beta_i\right) = \emptyset$. The proof from this point forward is analogous to the proof of Theorem 1, using $\beta$ in place of $\tau$ and using $\triangle = O(c(n)^7 n^{-1/5}) = o(1)$. □

## 3 Threshold for Isolated Vertices

In this section, we prove Theorem 3.

**Proof of Theorem 3.** Let $G \in \mathcal{SG}(Q_k, p)$ where $p = \sqrt{(\log n + \log\log n + c(n))/n}$ and let $X$ be the number of isolated vertices in $G$. Then

$$\begin{aligned} E[X] &= \sum_a \Pr[a \text{ isolated} \mid a \in G] \Pr[a \in G] \;=\; np(1-p^2)^{(n-1)/2} \\ &= np \exp\left(\frac{(n-1)}{2} \log(1-p^2)\right) \;=\; np \exp\left(-\frac{n-1}{2} p^2\right) \exp\left(-\frac{n-1}{2} \sum_{k=2}^{\infty} \frac{p^{2k}}{k}\right). \end{aligned}$$

We have

$$\begin{aligned} E[X] &= \sqrt{n(\log n + \log\log n + c(n))} \exp\left(-\frac{(n-1)}{2n}(\log n + \log\log n + c(n))\right) \exp\left(O\left(\frac{(\log n)^2}{n}\right)\right) \\ &= \sqrt{\frac{n(\log n + \log\log n + c(n))}{n \log n}}\, e^{-c(n)/2}(1+o(1)) \;=\; e^{-c(n)/2}(1+o(1)) \to e^{-c(n)/2}. \end{aligned} \tag{4}$$

If $c(n) = \omega(n)$ is a slowly growing function then $E[X] \to 0$, so $X = 0$ whp.

Next, we consider the case $c(n) = -\omega(n)$ where $\omega(n)$ is a slowly growing function. We have $E[X] \to \infty$, and we must show that $X > 0$ whp. Let $X_a$ be the indicator function for element $a$ being an isolated vertex. Fixing elements (and potential vertices) $a \neq b$, we determine $\Pr[X_a \wedge X_b]$. If both $a$ and $b$ are isolated vertices, then vertex $ab$ cannot be present. We analyze the remaining $n-3$ elements by constructing a dependency graph $G'$ with $V(G') = Q_k^* \backslash \{a, b, ab\}$ and $E(G') = \{(x, y) \,|\, xy = a\} \cup \{(x, y) \,|\, xy = b\}$. $G'$ consists of disjoint 4-cycles. Indeed, $E(G')$ is naturally partitioned into two perfect matchings, so $G'$ is 2-regular. If $u, v, x, y \in V(G')$ with $uv = xy = a$ and $ux = b$ then $vy = (uu)vy = u(uv)y = u(xy)y = (ux)(yy) = ux = b$, so these four vertices create a cycle.

Given that $a$ and $b$ are isolated vertices in $G$, for each of the 4-cycles of $G'$ we can have zero vertices, one vertex or two antipodal vertices present in $G$. The combined probability of these events is $(1-p)^4 + 4p(1-p)^3 + 2p^2(1-p)^2 = 1 - 4p^2 + 4p^3 - p^4$. This gives

$$E[X^2] = \sum_{a,b} \Pr[X_a \wedge X_b] = n(n-1)p^2(1-p)\left(1 - 4p^2 + 4p^3 - p^4\right)^{(n-3)/4} \tag{5}$$

where the sum is taken over all ordered pairs of elements $a, b$. We have

$$\begin{aligned}
\Pr[X > 0] \quad &\geq \quad \frac{E[X]^2}{E[X^2]} = \frac{n^2p^2(1-p^2)^n}{n(n-1)p^2(1-p)(1-4p^2+4p^3-p^4)^{(n-3)/4}} \\
&\geq \quad \frac{(1-p^2)^n}{(1-4p^2+4p^3)^{n/4}} \geq \left(\frac{1-p^2}{1-p^2+p^3}\right)^n \geq \left(1+p^3\right)^{-n} \geq \exp(-np^3) \to 1.
\end{aligned}$$

Therefore $\Pr[X > 0]$ with high probability, meaning that there are isolated vertices below the threshold.

Finally, suppose that $c(n) \to c$ where $c$ is a constant. We prove that $X$ converges to $\text{Po}(e^{-c/2})$ via the method of moments. In particular, for $s > 0$ we show that

$$\sum_{a_1, a_2, \ldots, a_s} \Pr[X_{a_1} \wedge X_{a_2} \wedge \cdots \wedge X_{a_s}] = \left(e^{-c/2}\right)^s$$

where the sum is taken over all ordered $s$-tuples of elements. The case $s = 1$ is shown by equation (4). For $s = 2$, we consider the limiting behavior of equation (5):

$$\begin{aligned}
\sum_{a,b} \Pr[X_a \wedge X_b] \quad &= \quad n^2p^2\left(1 - 4p^2 + O(p^3)\right)^{n/4}(1+o(1)) \\
&= \quad \left(n^2p^2 \exp(-np^2 + O(np^3))\right)(1+o(1)) \\
&= \quad e^{-c(n)}\,\frac{n(\log n + \log\log n + c(n))}{n \log n}(1+o(1)) \to e^{-c}.
\end{aligned}$$

For fixed $s > 2$, we must create a dependency graph $G'(a_1, a_2, \ldots, a_s)$ for having isolated vertices $a_1, a_2, \ldots, a_s$. This dependency graph is constructed similarly to the 2 vertex case above. First, we consider the set $B \subset Q_k^*$ of size $2^s - 1$ consisting of elements corresponding to all possible combinations of $a_1, a_2, \ldots, a_k$. Among the members of $B$, the $s$ elements $a_i$ must be present and the $\binom{s}{2}$ elements $a_ia_j$ must be absent. As for the remaining $2^s - 1 - s - \binom{s}{2}$ elements in $A$, one acceptable configuration is the absence of all of them which occurs with probability $(1-p)^{2^s-1-(s+s^2)/2} = 1 - o(1)$ since $s$ is fixed. There are other allowable configurations, but of course the total probability is at most 1. In summary, the probability that the elements of $A$ do allow for the $a_i$ to be isolated is $p^s(1-p)^{s(s-1)/2}(1-o(1))$.

We consider the remaining $n - (2^s - 1) = 2^k - 2^s$ elements. Generalizing the 2-vertex case, our dependency graph partitions these elements into disjoint $s$-hypercubes. Indeed, we first observe that the graph $H$ with $V(H) = B \cup \{e\}$ and $E(H) = \{(b_1, b_2) \mid b_1b_2 \in B\}$ is the $s$-hypercube. This graph is the prototype for the remaining components of $G'(a_1, a_2, \ldots, a_s)$. Indeed, let $C$ be a component of $G'(a_1, a_2, \ldots, a_s)$ and let $x \in V(C)$. Then $V(C) = \{xy \mid y \in B \cup \{e\}\}$ and for $xv_1, xv_2 \in V(C)$, we have $(xv_1, xv_2) \in E(C)$ if and only if $v_1v_2 \in B$.

Respecting the dependency graph of $C$, we can only allow an independent set of vertices of $C$ to be present in our graph. Let $Z_C$ denote the event that the vertices in $C$ that are present in the graph form an independent set. Then

$$\begin{aligned}\Pr[Z_C] &= (1-p)^{2^s} + 2^s p(1-p)^{2^s-1} + \frac{1}{2}(2^s)(2^s-(s+1))p^2(1-p)^{2^s-2} + O(p^3) \\ &= 1-(2^{s-1}s)p^2 + O(p^3)\end{aligned}$$

We now calculate

$$\begin{aligned}&\sum_{a_1,a_2,\ldots,a_s} \Pr[X_{a_1} \wedge X_{a_2} \wedge \cdots \wedge X_{a_s}] \\ &= (n)_s p^s (1-p)^{s(s-1)/2}(1-2^{s-1}sp^2+O(p^3))^{(n-2^s+1)/2^s}(1-o(1)) \\ &= n^s p^s (1-2^{s-1}sp^2+O(p^3))^{n/2^s}(1-o(1)) \;=\; n^s p^s \exp\left(-\frac{s}{2}np^2\right)(1+o(1)) \\ &= \left(e^{-c(n)/2}\right)^s \left(\frac{n(\log n + \log\log n + c(n))}{n\log n}\right)^{s/2}(1+o(1)) \to \left(e^{-c/2}\right)^s,\end{aligned}$$

completing the proof by the method of moments. □

## 4 Threshold for Connectivity

In this section, we prove Theorem 4. As per equation (3), the models $\mathcal{SG}(Q_k, p)$ and $\mathcal{SG}(Q_k, r)$ where $r = np$ are essentially equivalent. We make most of our arguments using the former model, which is easier to work with. In particular, we can define an analogous random sum graph process for this model as follows. We assign every element $a$ an i.i.d. weight $p_a$ drawn uniformly from $[0,1]$. We increase $p$ from 0 to 1, including vertex $a$ when $p \geq p_a$. We define the resulting random process as $\hat{G} = (G_p)_{p\in[0,1]}$, where $G_p \in \mathcal{SG}(Q_k, p)$.

In order to prove that the graph becomes connected when the last isolated vertex disappears (and remains connected thereafter), we prove a series of lemmas. We focus on the range $\sqrt{\log n/n} \leq p \leq \sqrt{(\log n + \log\log n + c)/n}$. For $p = \sqrt{\log n/n}$, we show that whp $\mathcal{SG}(Q_k, p)$ consists of isolated vertices and a constant number of large components of size at least $\sqrt{n\log n}/100$. Next, as $p$ increases from $\sqrt{\log n/n}$ to $\sqrt{(\log n + \log\log n + c)/n}$, we show that whp the graph consists of a small number of isolated vertices along with a single connected component. These isolated vertices are joined to the connected component at the threshold provided by Theorem 3. Our final lemma proves that whp no additional isolated vertices are introduced beyond this point.

Observe that

$$\Pr[G_p \text{ is not connected}] = \Pr[\cup_{t=1}^{n/2}\{G_p \text{ has a component of size } t\}]$$

and therefore

$$\Pr[G_p \text{ has an isolated vertex}] \leq \Pr[G_p \text{ is not connected}]$$

$$\leq \Pr[G_p \text{ has an isolated vertex}] + \sum_{t=3}^{n/2} \Pr[G_p \text{ has a component of size } t]$$

$$\leq \Pr[G_p \text{ has an isolated vertex}] + \sum_{t=3}^{n/2} E[\text{number of components of size } t].$$

Note that we cannot have a component of size 2 as edges appear in triangles. Define $p_- = \sqrt{\log n/n}$ and $p_+ = \sqrt{(\log n + \log\log n + c)/n}$.

**Lemma 5** *Let $c > 0$ be a constant. If $p_- \leq p \leq p_+$ then whp the only components of $G \in \mathcal{SG}(Q_k, p)$ with size smaller than $\frac{1}{100}\sqrt{n \log n}$ are isolated vertices.*

**Proof.** We write $p = \sqrt{g(n) \log n/n}$ where $1 \leq g(n) \leq 1 + (\log\log n + c)/\log n = 1 + o(1)$. For $3 \leq t \leq n/2$, let $X_t$ be the number of components of size $t$. We determine an upper bound for $E[X_t]$. There are $\binom{n}{2}/3$ potential components of size 3. A component of size 4 is impossible. For $t \geq 5$, an upper bound on the number of possible components of size $t$ is $\binom{n}{t-2}$. Indeed, if $a$ and $b$ are in the component, then $ab$ must also be in the component. A vertex $c$ connected to this triangle requires adding one of $ac, ab, abc$, so that 2 of the first 5 vertices are determined by the other 3.

Let $H$ be a potential component of size $t$ and let $A$ be the event that all of the vertices in $H$ are present, so that $\Pr[A] = p^t$. Triangles $\{a, b, ab\}$ that cross from $H$ to $G - H$ are of two forms: either $a, b \in V(G - H)$ and $ab \in V(H)$ or $a, b \in V(H)$ and $ab \in V(G - H)$. Let $\{B_i\}$ be the set of all events of the former type. We will ignore the second type as they can only decrease the probability that $H$ is isolated: $\Pr[H \text{ is an isolated component}] \leq \Pr[\wedge \overline{B}_i \mid A] \Pr[A] = p^t \cdot \Pr[\wedge \overline{B}_i \mid A]$.

We bound $\Pr[\wedge \overline{B}_i \mid A]$ for three cases: $t = 3$, $5 \leq t \leq \sqrt{n/\log n}$ and $\sqrt{n/\log n} \leq t \leq \frac{1}{100}\sqrt{n \log n}$. We use the following version of Janson's inequality [3, p. 116, Theorem 8.1.1] for cases 1 and 2. We write $i \sim j$ when $i \neq j$ and $B_i$ and $B_j$ are not independent. If $\triangle = \sum_{i \sim j} \Pr[B_i \wedge B_j]$ where this sum is over all ordered pairs, and $\mu = \sum_i \Pr[B_i]$ then $\Pr\left[\bigwedge_i \overline{B}_i\right] \leq \exp\left(-\mu + \triangle/2\right)$.

Case 1: $t = 3$. For $H$ to be an isolated triangle, we have $\mu = \frac{3}{2}(n-3)p^2$ and $\triangle = 6(n-3)p^3$. Indeed, for $\mu$, we have 3 choices for the vertex in $H$ and $(n-3)/2$ distinct choices for a vertex in $G - H$ (which uniquely determines the second vertex in $G - H$). As for $\triangle$, we choose an ordered pair of vertices from $H$ (one for $B_i$ and one for $B_j$) and one vertex from $G - H$ that is in both $B_i$ and $B_j$. Janson's inequality yields $\Pr[\wedge \overline{B}_i] \leq \exp(-\frac{3}{2}(n-3)p^2(1-2p))$ and therefore

$$\begin{aligned} E[X_3] &\leq \frac{1}{3}\binom{n}{2} p^3 \exp\left(-\frac{3}{2}(n-3)p^2(1-2p)\right) \\ &< \frac{1}{6} n^{1/2} (g(n)\log n)^{3/2} \cdot \exp\left(-\left(1 - 2\sqrt{\frac{g(n)\log n}{n}}\right)\log n\right) \to 0. \end{aligned}$$

Case 2: $5 \leq t \leq \alpha\sqrt{n/\log n}$ for $0 < \alpha < 1$. For $t$ in this range,

$$\frac{1}{2}t(n-2t)p^2 \leq \mu \leq \frac{1}{2}t(n-t)p^2 \qquad (6)$$

$$\binom{t}{2}(n-2t)p^3 \;=\; \left(\binom{t}{2}(n-t)-\binom{t}{2}t\right)p^3 \;\leq\; \tfrac{1}{2}\triangle \;\leq \binom{t}{2}(n-t)p^3 \;<\; \frac{1}{2}t^2(n-t)p^3. \quad (7)$$

For $\mu$, we have $t$ choices for the vertex in $H$ and at most $(n-t)/2$ choices for the vertex in $G-H$. In the lower bound, we account for triangles with two vertices in $H$ and one in $G-H$. As for $\triangle/2$, $\binom{t}{2}(n-t)$ is an upper bound on the number pairs of triangles containing distinct vertices in $H$ and sharing one vertex in $G-H$. Finally, $\binom{t}{2}t$ is an upper bound on the number of pairs of intersecting triangles with at least one triangle having two vertices in $H$ (pick 2 vertices in H for one triangle and another vertex in $H$ for the second triangle). We have

$$E[X_t] \;\leq\; \binom{n}{t-2}p^t\exp\left(-\mu+\triangle/2\right) \leq \left(\frac{npe}{t-2}\right)^{t-2} p^2\exp\left(-\frac{tp^2}{2}(n-2t)+\frac{t^2p^3}{2}(n-t)\right).$$

Substituting $p=\sqrt{g(n)\log n/n}$ and taking the logarithm yields

$$\begin{aligned}\log(E[X_t]) \;&\leq\; -2\log n+\frac{t}{2}\log\log n+(t-2)\log\left(\frac{e}{t-2}\right)\\ &\quad+\frac{t^2g(n)\log n}{2n}\left(2+\sqrt{ng(n)\log n}-t\sqrt{\frac{\log n}{n}}\right). \qquad (8)\end{aligned}$$

We consider the endpoints. Taking $t=5$, we find $\log(E[X_5])=-2\log n+O(\log\log n)\to-\infty$, so that $E[X_5]\to 0$ and therefore there are no components of size 5 whp. Now consider $t=\alpha\sqrt{n/\log n}$ where $0<\alpha<1$. In this case,

$$\log(E[X_{\alpha\sqrt{n/\log n}}]) \;\leq\; \left(-\frac{\alpha}{2}+\frac{\alpha^2}{2}(g(n))^{3/2}\right)\sqrt{n\log n}+O\left(\sqrt{\frac{n}{\log n}}\log\log n\right) \;\to\; -\infty,$$

confirming that whp there are no components of size $\alpha\sqrt{n/\log n}$.

We now show that when $5<t\ll\sqrt{n/\log n}$, whp equation (8) tends to $-\infty$. Taking its derivative with respect to $t$ and solving for any critical points gives the equation

$$\frac{1}{2}\log\log n=\log(t-2)-\frac{t\,g(n)\log n}{n}\left(2+\sqrt{ng(n)\log n}\right)+\frac{3}{2}t^2g(n)\left(\frac{\log n}{n}\right)^{3/2}.$$

The left hand side is independent of $t$ and the only possible critical point in the given interval is roughly $t\approx\sqrt{\log n}+2$. Therefore this critical point is of the form $t=\beta\sqrt{\log n}$ where $|\beta-1|<\epsilon$ for arbitrarily small $\epsilon$. Near this critical point, equation (8) becomes

$$\log(E[X_{\beta\sqrt{\log n}}]) \;\leq\; -2\,\log n+O(\sqrt{\log n}\log\log n)\to-\infty.$$

This guarantees that there are no components of size $5<t\ll\sqrt{n/\log n}$.

Case 3: $\sqrt{n/\log n}\leq t\leq\frac{1}{100}\sqrt{n\log n}$. Just as in Case 2, equations (6) and (7) hold. Therefore, $\mu\leq\frac{1}{2}(n-t)tp^2=\frac{1}{2}np^2t(1-o(1))$ and $\triangle\geq 2\binom{t}{2}(n-2t)p^3\geq np^2t(1-o(1))$. For large $n$, $\triangle\geq\mu$ and the extended form of Janson inequality [3, p. 117, Theorem 8.1.2] holds: $\Pr\left[\bigwedge_i\overline{B}_i\right]\leq\exp\left(-\mu^2/2\triangle\right)$. Using the complementary bounds in equations (6) and (7), the inequality yields

$$\Pr\left[\bigwedge_i\overline{B}_i\right]\leq\exp\left(-\frac{(n-2t)^2p}{4(n-t)}\right)\leq\exp\left(-\frac{1}{5}(n-2t)p\right).$$

This gives

$$E[X_t] \le \binom{n}{t} p^t \exp\left(-\frac{1}{5}(n-2t)p\right) \le \left(\frac{npe}{t}\right)^t \exp\left(-\frac{1}{5}(n-2t)p\right)$$

and taking the log gives

$$\log(E[X_t]) \le t \log\left(\frac{npe}{t}\right) - \frac{1}{5}(n-2t)p =: f(t). \tag{9}$$

We consider the endpoints. First, $t = \sqrt{n/\log n}$:

$$\log(E[X_{\sqrt{n/\log n}}]) \le -\frac{1}{10}\sqrt{n\, g(n) \log n} + O\left(\sqrt{\frac{n}{\log n}} \log\log n^{2e}\right) \to -\infty$$

for $n$ sufficiently large. Next, $t = \frac{1}{100}\sqrt{n \log n}$:

$$\log(E[X_t]) \le \left(-\frac{1}{10} + \frac{\log(100e)}{100} + \frac{1}{2}\log g(n)\right)\sqrt{n\log n} + O(\log n) \to -\infty$$

for $n$ sufficiently large. To settle all values in between, the derivative $f'(t)$ of the right hand side of equation (9) is

$$\frac{d}{dt} f(t) = \log\left(\frac{np}{t}\right) + \frac{2}{5}p \ge \log(100\, g(n)) + \frac{2}{5}\sqrt{\frac{g(n)\log n}{n}} > 0.$$

That $f'(t)$ is positive on this interval ensures $E[X_t] \to 0$ for all such $t$. □

**Lemma 6** *Consider the evolution from $G_{p_-}$ to $G_{p_+}$. At the conclusion of this phase, $G_{p_+}$ consists one a large connected component along with some (perhaps zero) isolated vertices whp.*

**Proof.** Lemma 5 ensures that when $p = p_-$, the only components of size smaller than $\sqrt{n \log n}/100$ are isolated vertices. We now show that as $p$ increases from $p_-$ to $p_+$, any components of size at least $\sqrt{n\log n}/100$ must coalesce so that the graph becomes one connected component along with some isolated vertices.

As shown in equation (3), whp $\mathcal{SG}(Q_k, p)$ looks like $\mathcal{SG}(Q_k, r)$ where $r = np$ (here we choose $r$ random vertices). We consider this model for convenience. Let $r_1 = np_- = \sqrt{n\log n}$ and $r_2 = np_+ = \sqrt{n(\log n + \log\log n + c)}$. Then whp

$$\begin{aligned} a = r_2 - r_1 &= \sqrt{n\log n}\left(\left(1 + \frac{\log\log n + c}{\log n}\right)^{1/2} - 1\right) \\ &\ge \frac{1}{2}\sqrt{n\log n}\left(\frac{\log\log n + c}{\log n}\right) \\ &= \frac{\log\log n + c}{2}\sqrt{\frac{n}{\log n}} \end{aligned}$$

is the number of vertices that will be added to arrive at the final graph.

Let $G \in \mathcal{SG}(Q_k, r_1)$ so that $v(G) = r_1 = \sqrt{n \log n}$. There are at most 100 connected components of size at least $\sqrt{n\log n}/100$. Let $H_1, H_2, \ldots, H_s$ be these large components where $1 \le s \le 100$.

For any pair $H_i$, $H_j$ there must be at least $n \log n/10000$ potential edges between them. Label each of these potential edges by the element $v$ whose addition will add that edge to the graph. For each element $v \notin V(G)$, let $X_v$ denote the number of potential edges with label $v$. Then $E[X_v] = \frac{1}{2}(n-1)p_1^2 = \frac{1}{2}\log n(1-o(1))$. We use Chernoff's inequality (1) to show that whp every $v \in Q_k^* \backslash V(G)$ has at most $\frac{3}{2}\log n$ associated labeled edges. Let $Y_i$ denote the event that $X_i \geq 3E[X_i] = \frac{3}{2}(1-o(1))\log n$. Then

$$\Pr[\vee Y_i] \leq \sum_{v_i \notin G} \Pr[X_i > 3E[X_i]] \leq (1-o(1))n \exp(-5(1-o(1))\log n) \leq n^{-3} \to 0.$$

It follows that whp there are at least $n/15000$ distinct labels between any two components. Let $Z_{ij}$ denote the number of edges induced between distinct components $H_i$ and $H_j$ via the addition of the $a$ random vertices to $G$. We have

$$\mu_{ij} = E[Z_{ij}] \geq a \cdot \frac{n}{15000} \cdot \frac{1}{n - \sqrt{n \log n}} \geq \frac{a}{15000} \geq \frac{\log\log n}{30000}\sqrt{\frac{n}{\log n}}.$$

Once again we combine a union bound along with the Chernoff inequality to show that all these large components must coalesce into a single component whp. Indeed,

$$\begin{aligned} \Pr[\vee(Z_{ij}=0)] &\leq \sum_{i,j}\Pr[Z_{ij}=0] \leq \sum_{i,j}\Pr[Z_{ij} < \frac{1}{2}\mu_{ij}] \leq 10000\exp\left(-\mu_{ij}\phi(1/2)\right) \\ &\leq 10000\exp\left(-\frac{1-\log 2}{60000}\log\log n\sqrt{\frac{n}{\log n}}\right) \to 0. \end{aligned}$$

Finally, we claim that any newly created component must be an isolated vertex. Indeed, by Lemma 5, all non-isolated vertices must be in components of size at least $\sqrt{n\log n}/100$. We have only added $a \ll \sqrt{n \log n}/100$ vertices, so these vertices cannot be part of a new small component. Therefore whp we have one large component along with (perhaps) some isolated vertices. □

Let $q = \sqrt{(\log n + \log\log n - \log\log\log n)/n}$. By Theorem 3, the graph $G_q$ has isolated vertices whp. We now show that for $p > q$, whp no isolated vertices are added in $G_p$. In other words, the set of isolated vertices in $G_p$ decreases monotonically for $p > q$.

**Lemma 7** *Whp, every element $a \in Q_k^*$ such that $p_a > q$ has $\deg_{G_{p_a}}(a) > 0$.*

**Proof.** We partition $[0,1]$ into four intervals, each representing a phase in the evolution our random cubic sum graph. We will show that every vertex added in the latter three stages will be adjacent to existing vertices. Let $q_0 = 0$, $q_1 = q = \sqrt{(\log n + \log\log n - \log\log\log n)/n}$, $q_2 = \sqrt{(\log n + 2\log\log n - \log\log\log n)/n}$, $q_3 = \sqrt{(2\log n + 2\log\log n - \log\log\log n)/n}$ and $q_4 = 1$.

Let $W_i = \{a \in Q_k^* \mid p_a \in [q_{i-1}, q_i]\}$ for $1 \leq i \leq 4$. Consider the sequence of graphs $G_{q_1}$, $G_{q_2}$, $G_{q_3}$, and $G_{q_4} = K_n$. Note that $V(G_{q_i}) = \cup_{j=1}^{i} W_j$. We claim that for $2 \leq i \leq 4$, whp every vertex in $W_i$ is adjacent to some pair of vertices that already appear in $G_{q_{i-1}} \subset G_{p_a}$. Indeed, for $a \in W_i$, let $Z_a$ denote the indicator for event that $a$ does not form a triangle with pairs of vertices in $G_{q_{i-1}}$. We have

$$\begin{aligned}\Pr\Big[\bigcup_{a:p_a>q}(a \text{ is an isolated vertex in } G_{p_a})\Big] &\leq \Pr\Big[\vee_{a\in W_2\cup W_3\cup W_4} Z_a\Big] \\ &\leq \sum_{a\in W_2} E[Z_a] + \sum_{a\in W_3} E[Z_a] + \sum_{a\in W_4} E[Z_a].\end{aligned}$$

Using $q_2 = \sqrt{q_1^2 + \log\log n/n} \leq q_1 + \sqrt{\log\log n/n}$, we can bound

$$\begin{aligned}\sum_{a\in W_2} E[Z_a] &\leq (q_2 - q_1)n(1-q_1^2)^{n/2} \\ &\leq \sqrt{n\log\log n}\,\exp\left(-\frac{\log n + \log\log n - \log\log\log n}{2}\right) \\ &= \sqrt{n\log\log n}\sqrt{\frac{\log\log n}{n\log n}} = \frac{\log\log n}{\sqrt{\log n}} \to 0.\end{aligned}$$

Similarly, we have $q_3 = \sqrt{q_2^2 + \log n/n} \leq q_2 + \sqrt{\log n/n}$, so that

$$\begin{aligned}\sum_{a\in W_3} E[Z_a] &= (q_3 - q_2)n(1-q_2^2)^{n/2} \\ &\leq \sqrt{n\log n}\,\exp\left(-\frac{\log n + 2\log\log n - \log\log\log n}{2}\right) \\ &= \sqrt{n\log n}\sqrt{\frac{\log\log n}{n\,(\log n)^2}} = \sqrt{\frac{2\log\log n}{\log n}} \to 0.\end{aligned}$$

Finally,

$$\begin{aligned}\sum_{c\in W_4} E[Z_a] &\leq n(1-q_3^2)^{n/2} \\ &\leq n\,\exp\left(-\frac{2\log n + 2\log\log n - \log\log\log n}{2}\right) \\ &= n\frac{\sqrt{\log\log n}}{n\,\log n} = \frac{\sqrt{\log\log n}}{\log n} \to 0.\end{aligned}$$

Hence $\Pr[\bigvee_{a\in W_2\cup W_3\cup W_4} Z_a] \to 0$. □

**Proof of Theorem 4.** Equation (3) guarantees the equivalence of the random sum graph processes $\tilde{G} = (G_t)_{t=1}^n$ and $\hat{G} = (G_p)_{p\in[0,1]}$. We prove the theorem for the latter formulation. We consider the evolution starting from $G_q$. By Theorem 3, there are isolated vertices whp when $p = q$. By Lemma 7, whp no new isolated vertices are introduced beyond this point. Therefore, once the minimum degree is nonzero, it remains nonzero thereafter whp.

We now turn our attention to the connectivity of the graph. First we consider the evolution from $G_q$ to $G_{p_+}$. By Lemma 6, $G_{p_+}$ consists of one large connected component plus some isolated vertices. Moreover, the proof of Lemma 6 shows that no new small components are created for $q < p < p_+$.

Therefore every vertex $v$ added during this phase is contained in the unique large component of $G_{p_v}$. Now consider the random sum graph process starting from $G_{p_+}$. The only obstruction to connectivity is the existence of isolated vertices that were included for $p < q$. When these isolated vertices disappear, the graph becomes connected, so that $\tau(\delta(\hat{G})) > 0) = \tau(\kappa(\hat{G}) > 0)$. Once the graph becomes connected, it remains connected whp since no new isolated vertices are introduced for $p > q$. □

**Acknowledgment.** The author thanks the anonymous referee for the careful reading and the suggestions for refining the results herein.

## References

[1] N. Alon and D. J. Kleitman, Sum-free subsets, in, A tribute to Paul Erdős, A. Baker, B. Bollobás and A. Hajnál, eds., Cambridge University Press, Cambridge, England, 1990, pp. 13–26.

[2] N. Alon and E. Scheinerman, Generalized sum graphs, Graphs and Combin., 8 (1992), pp. 23–29.

[3] N. Alon and J. Spencer, The Probabilistic Method, John Wiley & Sons, New York, 2000.

[4] A. Barbási and R. Albert, Emergence of scaling in random networks, Science 286, pp. 509-512.

[5] D. Bergstrand, F. Harary, K. Hodges, G. Jennings, L. Kuklinski, and J. Wiener, The sum numbering of a complete graph, Bull. Malaysian Math Soc., 12, 1989, pp. 25–28.

[6] J. Boland, R. Laskar, C. Turner and G. Domke, On mod sum graphs, Congress. Numer., 70, 1990, pp. 131–135.

[7] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády, The degree sequence of a scalefree random graph process, Random Structures Algorithms, 18, No. 3, 2001, pp. 279–290.

[8] C. Cooper and A. Frieze, A general model of web graphs Random Structures and Algorithms, 22, 2001, pp. 500-511.

[9] M. N. Ellingham, Sum graphs from trees, Ars Combin., 35, 1993, pp. 335–349.

[10] P. Erdős and A. Renyi, On random graphs, Publ. Math. Debrecen., 6, 1959, pp. 290–297.

[11] P. Erdős and A. Renyi, On the evolution of random graphs, Magyar Tud. Akad. Mat. Kutató Int. Közl., 5, 1960, pp. 17–61.

[12] J. Gallian, A Dynamic Survey of Graph Labeling, Electronic Journal of Combinatorics, Dynamic Survey 6, revision #9, July 2005, `http://www.combinatorics.org`.

[13] T. Hao, On sum graphs, J. Combin. Math. Combin. Computing, 6, 1989, pp. 207–212.

[14] F. Harary, Sum Graphs and Difference Graphs, Congress. Numer., 72, 1990, pp. 101–108.

[15] S. Janson, T. Łuczak and A. Ruciński, Random Graphs, John Wiley & Sons, New York, 2000.

[16] M. Miller, J. Ryan, K. Slamin, and W. Smyth, Labelling wheels for the minimum sum number, J. Combin. Math. Combin. Computing, 28, 1998, pp. 289–297.

[17] M. Molloy and B. Reed, A critical point for random graphs with a given degree sequence, in Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science, "Random Graphs '93" (Poznan 1993), 6, 1995, pp. 161–179.

[18] M. Penrose, Random Geometric Graphs, Oxford University Press, Oxford, England, 2003.

[19] S. Slamet, K. A. Sugeng and M. Miller, Sum graph based access structure in a secret sharing scheme, Journal of Prime Research in Mathematics, Vol.2, 2006, pp. 113-119.