Random graphs in a neural computation model

Alexandros V. Gerbessiotis

CS Department

New Jersey Institute of Technology

Newark, NJ 07102.

July 2, 2002

Abstract

We examine in this work the following graph theory problem that arises in neural computations that involve the learning of boolean expressions by studying the asymptotic connectivity properties of $G_{n,1/(kn)^{1/2}}$ random graphs, where k is a fixed positive integer. For an undirected graph $G = (V, E)$ let $N(X, Y) = \{v \in V - (X \cup Y) \mid \exists x \in X \text{with } (v, x) \in E\}$. For fixed k construct an undirected graph $G = (V, E)$ such that for all disjoint sets $A, B \subseteq V$ such that $|A| = |B| = k$, and $C = N(A, B) \cap N(B, A)$, set C is such that $|C|$ is either exactly k or as close to k as possible. Asymptotic results for large values of k are also presented.

Keywords: Random graphs, connectivity properties, neural networks.

1 Introduction

Let $G = (V, E)$ be an undirected connected graph on a set of n vertices V and let k be a constant. We define set $N(X, Y)$, where $X, Y \subseteq V$, to be the set of vertices in $V - X$ excluding vertices in Y that are adjacent to a vertex in X, in other words,

$$N(X, Y) = \{v \in V - (X \cup Y) \mid \exists x \in X \text{ with } (v, x) \in E\}.$$

We investigate in this work solutions to the following graph construction problem posed by L. G. Valiant [11, 12].

Problem 1 *For fixed k construct an undirected graph $G = (V, E)$ with the following connectivity property (CP). For all disjoint sets $A, B \subseteq V$ such that $|A| = |B| = k$, and $C = N(A, B) \cap N(B, A)$ set C is such that $|C|$ is either exactly k or as close to k as possible, in other words, one of the following two properties must be satisfied.*

$$\begin{aligned} |C| &= k \qquad \text{(Property CP1) or} \\ |C| &\approx k \qquad \text{(Property CP2).} \end{aligned}$$

We are interested in the behavior of the construction, if such exists, for n asymptotically tending to infinity. It has been shown in [8] that if such a graph G exists so that Property CP1 always holds, then graph G must have exactly $3k$ vertices. We are thus more interested in cases where Property CP2 applies. We examine in this work the connectivity properties of $G_{n,1/(kn)^{1/2}}$ random graphs with respect to this problem since for such graphs Valiant observed in [11] that for random choices of A and B of size k each, the expectation and the variance of $|C|$ are both approximately k asymptotically for large n.

In [11] the functional capabilities of sparse networks of neurons in accumulating knowledge through interactions with the outside world were investigated. In the neural network model proposed in [11, 12] the learning task in the sense of [10] is to establish in the network a circuit that computes a boolean expression. The structure of the underlying network of neurons plays an important role in the various learning tasks supported by the model. Neurons A, B, C for example, may store information related to some concepts or attributes c_A, c_B, and c_C respectively. When the neural network learns for example simple conjunctions, information related to conjunction $c_C = c_A \wedge c_B$ needs to be stored in a neuron C that is connected to both A and B [12]. Various modes of learning thus require that the neural network maintain at minimum the property that for every two neurons A and B there exists a third neuron C which is connected to both A and B. In addition redundancy requirements in modeling the human brain [12] stipulate that a concept or attribute be stored not in a single neuron but in a collection of some k neurons; k is called the *redundancy* parameter of the neural network. If we generalize the conjuntion example (a $k = 1$ case) to take into consideration this redundancy requirement, a concept or attribute becomes available not in a single neuron but in a collection of neurons. Let now $A, B, C, \ldots$ be sets of neurons storing information related to concepts or attributes $c_A, c_B, c_C, \ldots$ respectively. Let sets A and B be of size k each. Then, the requirement for set C of neurons that will store concept $c_C = c_A \wedge c_B$ is that $|C|$ be k (Property CP1) or very close to k (Property CP2), so that concept C becomes also available in about k neurons. This redundancy parameter k links neural networks to graphs and a solution to Problem 1 will provide a solution to the problem of organizing a neural network in such a way that would facilitate simple learning tasks (i.e. learning boolean expressions) and also maintain some form of redundancy in the sense that if a neuron storing concept c_C is knocked out, knowledge about c_C is still available in the network. Note that if neither Property CP1 nor CP2 are satisfied, then successive learning tasks that utilize concepts like C could cause instabilities in the neural network by trigerring large number of neurons (if $|C|$ is much higher than k) during the learning process of such tasks or diminishing number of neurons (if $|C|$ is much less than k). For example the successive learning task of learning $c_G = c_A \wedge c_B \wedge c_D \wedge c_E$, after learning concepts c_C and c_F, where $c_C = c_A \wedge c_B$ and $c_F = c_D \wedge c_E$, could cause further instability if c_C and c_F are available to more/less than about k neurons thus causing c_G to be available in even more/less neurons.

2 Definitions and Contents of the paper

We present below the following definitions for distinct vertices, a, b, and c and distinct sets of vertices A, B, and C of an undirected graph G with reference to Problem 1.

Definition 1 *A vertex* c *is* common *to vertices* a *and* b *if* c *is adjacent to both* a *and* b.

Similarly, a vertex c is common to two sets of vertices A and B if there exist vertices $a \in A$, and $b \in B$ such that c is common to a and b, and thus a set C is common to two sets of vertices A and B, if for every vertex $c \in C$, vertex c is common to A and B. The notion of "commonness" in this paper is related to the adjacency properties of a vertex or a set of vertices rather than to the intersection of sets of vertices.

Definition 2 *A set* C *is "bad" (or a-bad) if* $|C| = 0$ *or* $|C| \geq ak$ *for some positive constant* $a > 1$.

Definition 3 *A* $G_{n,p}$ *random graph is a undirected graph on* n *labeled vertices such that every edge, among the* $n(n-1)/2$ *possible ones, is included in the graph with probability* p *independently of the other ones.*

All results presented in this paper hold for $G_{n,p}$ random graphs, where $p = 1/(k \cdot n)^{1/2}$. When we claim that a property holds with high probability it means that this probability tends to one as $n \to \infty$ or, in other words, the probability that this property fails to hold is bounded above by a function $\epsilon(n) \to 0$ for $n \to \infty$. All logarithms in this manuscript are to base e, unless otherwise stated; in such exceptions $\lg x$ will denote the logarithm of x to base two. When the approximation $\approx$ symbol, it means that the ratio of the quantities on the two sides of this symbol tends to one as $n \to \infty$. The notation $X \sim N(\mu, \sigma^2)$ indicates that random variable X follows a normal distribution with expectation μ and variance σ^2. The probability of having $S_{n,p}$ successes in n independent Bernoulli trials of individual success probability p is given by a binomial term $B(S_{n,p}; n, p)$. The following bounds [2], for the tails of the binomial distribution will be used.

$$Pr(S_{n,p} \geq (1+\epsilon) \cdot n \cdot p) \leq e^{-\frac{1}{3} \cdot \epsilon^2 \cdot n \cdot p},$$

and

$$Pr(S_{n,p} \leq (1-\epsilon) \cdot n \cdot p) \leq e^{-\frac{1}{2} \cdot \epsilon^2 \cdot n \cdot p},$$

where $0 < \epsilon < 1$.

We examine in this paper Problem 1 in the context of $G_{n,1/(kn)^{1/2}}$ random graphs under an iterative procedure outlined below for iteration $r \geq 0$.

Procedure 1 *In a* $G_{n,1/(kn)^{1/2}}$ *random graph, the following binary-tree oriented iterative procedure forms a set of vertices at some iteration* $r \geq 0$*; in the implied binary tree a node* C *with two children* A *and* B *indicates that set* C *is common to* A *and* B*.*

- *At iteration 0,* 2^r *sets of vertices are formed by choosing arbitrarily at random* k*-subsets among the vertices of the graph not previously selected. Each such set becomes a leaf in a complete binary tree of height* r*.*

- *A set of vertices formed at iteration i is common to the two sets of vertices formed at iteration $i-1$ that are the children of the set in the implied binary tree depicting the iterative procedure. The number of sets thus formed at iteration i is 2^{r-i}.*

Figure 1 illustrates this iterative procedure for $r = 3$. We call the implied binary tree of the iterative procedure, *tree of iteration r*.

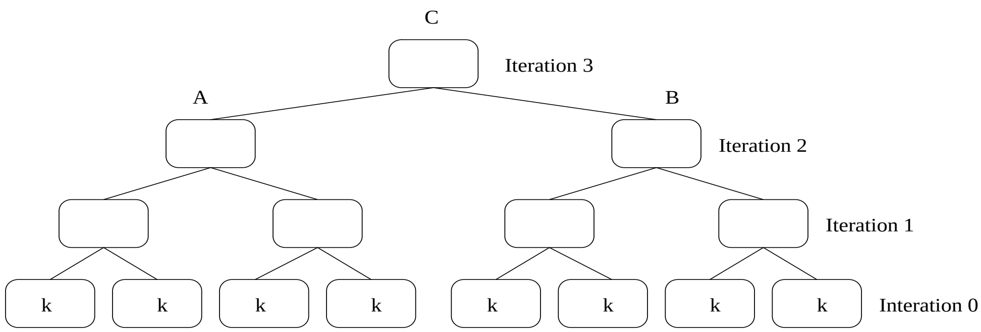

Figure 1: Tree of iteration $r = 3$.

We investigate for any constant iteration r the distribution of $|C|$, where C is a set formed at iteration r; with reference to Figure 1 such a set C becomes the root of the tree of iteration r or a node of height r in a tree of iteration $m > r$. Let Y_r be the random variable that represents $|C|$.

Proposition 1 establishes bounds for the expectation of Y_r, in terms of the expectation of Y_{r-1} representing $|A|$ and $|B|$. We show this proposition in a more general form where C is common to two sets A and B and the expectation of $|A|$ and $|B|$ are μ_A and μ_B respectively, by showing that $E[Y_r] \approx \mu_A \mu_B / k$ thus deriving by induction that $E[Y_i] \approx k$ for every $i \leq r$.

Proposition 1 *Consider Procedure 1, and a set C formed at iteration r common to sets A, B formed at iteration $r-1 \geq 0$ such that $|A|$ and $|B|$ have expectations and variances μ_A, v_A and $\mu_B (= \mu_A)$, $v_B (= v_A)$ respectively. Then asymptotically for large n*

$$\mathcal{E}[Y_r] \approx \frac{\mu_A \cdot \mu_B}{k} = \frac{\mathcal{E}[Y_{r-1}] \cdot \mathcal{E}[Y_{r-1}]}{k} = \frac{\mu_A^2}{k}.$$

Therefore, $\mathcal{E}[Y_r] \approx k$ for every constant iteration $r \geq 1$.

Proposition 2 shows the corresponding result for the variance of Y_r. We can prove a tighter bound for the variance as expressed by Proposition 3 after we establish Theorem 3 in section 4.

Proposition 2 *Consider Procedure 1, and a set C formed at iteration r common to sets A, B formed at iteration $r-1 \geq 0$ such that $|A|$ and $|B|$ have expectations and variances μ_A, v_A and*

$\mu_B(=\mu_A)$, $v_B(=v_A)$ respectively. Then asymptotically for large n

$$var(Y_r) = \mathcal{E}[Y_r^2] - \mathcal{E}^2[Y_r] \leq \frac{\mu_A \cdot \mu_B}{k} + \frac{4 \cdot v_A \cdot v_B}{k^2} + \frac{2 \cdot {\mu_A}^2 \cdot v_B + 2 \cdot {\mu_B}^2 \cdot v_A}{k^2}.$$

Since $\mu_A = \mu_B = \mathcal{E}[Y_{r-1}] \approx k$, and for $v_A = v_B = var(Y_{r-1}) = v$ we have that

$$var(Y_r) \leq k + 4v^2/k^2 + 4v.$$

Proposition 3 *Consider Procedure 1, and a set C formed at iteration r common to sets A, B formed at iteration $r-1 \geq 0$ such that $|A|$ and $|B|$ have expectations and variances μ_A, v_A and $\mu_B(=\mu_A)$, $v_B(=v_A)$ respectively. Then asymptotically for large n*

$$var(Y_r) = \mathcal{E}[Y_r^2] - \mathcal{E}^2[Y_r] \approx \frac{\mu_A \cdot \mu_B}{k} + \frac{v_A \cdot v_B}{k^2} + \frac{{\mu_A}^2 \cdot v_B + {\mu_B}^2 \cdot v_A}{k^2}.$$

Since $\mu_A = \mu_B = \mathcal{E}[Y_{r-1}] \approx k$, and for $v_A = v_B = var(Y_{r-1}) = v$ we have that

$$var(Y_r) \approx k + 2v + v^2/k^2.$$

In section 4 we use bounds for the tails of the binomial distribution to derive bounds for the tails of Y_r. For small values of k these bounds are rather trivial. The DeMoivre-Laplace bound combined with the techniques in the proofs of Theorems 1 and 2 to follow is applied to examine pieces of the binomial tail in order to derive the bound in Theorem 3 that is necessary to show Proposition 3. The results obtained can be summarized below. In these theorems, the probability that a set formed at some iterations has size y is bounded above by the probability that a set formed at some iteration is declared to have size y.

Theorem 1 *Consider Procedure 1, and a set C formed at iteration r. Then, in the limit $n \to \infty$, set C is declared to have size at most $(1-a)^{2^r-1} \cdot k$, $0 < a < 1$ with probability at most*

$$\sum_{\lambda=0}^{r-1} 2^\lambda \cdot e^{-\frac{1}{2} \cdot a^2 \cdot (1-a)^{2^{r-\lambda}-2} \cdot k}.$$

Theorem 2 *Consider Procedure 1, and a set C formed at iteration r. Then, in the limit $n \to \infty$, set C is declared to have size at most $(1+a)^{2^r-1} \cdot k$, $0 < a < 1$ with probability at most*

$$\sum_{\lambda=0}^{r-1} 2^\lambda \cdot e^{-\frac{1}{3} \cdot a^2 \cdot (1+a)^{2^{r-\lambda}-2} \cdot k}.$$

Theorem 3 *Consider Procedure 1, and a set C formed at iteration r. Then, in the limit $n \to \infty$, set C is declared to have size at most $u^{2^r-1} \cdot k$, with probability at most*

$$\sum_{\lambda=0}^{r-1} 2^\lambda \cdot (e/u)^{u^{2^{r-\lambda}-1} \cdot k}$$

where $u \cdot (1 - \frac{u^{2^l-2} \cdot k}{n}) > 2$ and $\frac{u^{2^l-2} \cdot k}{n} \cdot n > 1$, for any l such that $1 \leq l \leq r$.

The results described so far and to be proved in sections 3 through 4 hold for constant k and $n \to \infty$. In sections 5, 6 we examine the case where in addition to having $n \to \infty$, k is also large but independent of n. We show, through the use of exponential generating functions, that Y_r can be approximated by a normal random variable with mean k and variance approximately $(2^r - 1)k + O(1)$. More precisely, the value of the variance is that obtained, though through other techniques, in Proposition 3. In order to achieve this result we develop a new way to examine the variance of $|C|$; we first approximate the product $|A||B|$ by a normal random variable and then express the probability of having a set C of some size with respect to this newly introduced random variable.

Theorem 4 *If $X_1, X_2 \sim N(\mu, \sigma^2)$ and independent, then the random variable $X = X_1 \cdot X_2$ follows, for $\mu = \Theta(k)$ and $\sigma^2 = \Theta(k)$, for some parameter k independent of n, in the limit for large k, a normal approximation $N(\mu^2, \sigma^4 + 2\,\mu^2\,\sigma^2)$.*

Theorem 5 *Consider a $G_{n,1/(kn)^{1/2}}$ random graph, where k is independent of n. At iteration r, where r is constant, random variable Y_r approximates, for large k and $n \to \infty$, a normal distribution $N(k, (2^r - 1)k + O(1))$.*

3 Iterated Mean and Variance

We now proceed to examining the expectation and the variance of Y_r under Procedure 1. Suppose we have two distinct sets in a $G_{n,1/(kn)^{1/2}}$ random graph, one of size i and another of size j. The probability that a vertex c of the graph is connected to some vertex a of the first set and some vertex b of the second set is equal to r_{ij}

$$r_{ij} = (1 - (1-p)^i) \cdot (1 - (1-p)^j) \approx\; i \cdot j \cdot p^2 = \frac{i \cdot j}{k \cdot n}, \tag{1}$$

where $p = 1/\sqrt{nk}$ denotes the edge probability of $G_{n,1/(kn)^{1/2}}$. The approximation in Equation (1) holds provided $pi \ll 1$ (similarly for pj). We shall select sets of common vertices according to a process outlined in [1]. This way with high probability for large n and constant r a set formed at iteration r is disjoint from sets previously formed.

Procedure 1 is modified as follows and this modified version is referenced in all propositions and theorems. Note that in the description below edges of the random graph are revealed one by one as necessary during the course of execution of Procedure 1. Let V be the set of vertices of the $G_{n,p}$ random graph, F a random subset of V of size say $n^{1/5}$, $P = V - F$, and $n_0 = |P| = n - n^{1/5} - 2^r k$. The term $2^r k$ in the expression for $|P|$ counts the vertices of 2^r sets of k vertices each at iteration zero. In the steps of Procedure 1 we select the vertices that will form a set at some iteration m among the vertices of P rather than V. Such a set formed at iteration m will thus be common to two sets formed at iteration $m-1$. If the two sets of iteration $m-1$ have sizes i and j respectively then, the size of the set at iteration m is equal to l with probability given by a binomial distribution $B(l; n_0, r_{i,j})$. If the set of iteration m thus formed is of size indeed l we choose l vertices of P uniformly at random without replacement. We select and then remove the so chosen vertices from P and refill P to size n_0 by moving l arbitrary vertices from F to P. If the size of F is less than l then we declare failure of this selection process and subsequently of the formation of a set at iteration r. One can prove that with high probability in the construction of a tree of iteration r the number of vertices that will be selected from P will be overall less than the cardinality of set F. Thus this selection procedure will succeed with high probability. We call this method of selecting common vertices Select and its correctness is proved by Claim 1.

One needs to show that any result that will be obtained in this paper by restricting the choice of vertices to set P rather than V in Select could have been obtained, with probability $1-o(1)$, without the use of Select. This is proved as follows. The probability that any of at most $n^{1/5}$ vertices that can be selected from F by Select is connected to any vertex of a set of size at most $n^{1/5}$, which is the maximum size of F during Select, is at most $n^{1/5}(1-(1-p)^{n^{1/5}}) \leq n^{1/5}(1-e^{-n^{1/5}/((1/p)-1)}) \leq n^{2/5}/(\sqrt{kn}-1) = o(1)$. We used here the inequalities $(1-1/n)^n \leq 1/e \leq (1-1/n)^{n-1}$, $e^{-x} \geq 1-x \quad (x \geq 0)$. Hence, with probability $1-o(1)$, no two vertices among the ones included in a set formed during the construction of a tree of iteration r will have a vertex in common in F.

The claim below summarizes the properties of Select that will be of interest in this paper. This claim is to be proved at the end of this section. We note that the requirement that r be a constant can be replaced by one that restricts r to be independent of n only.

Claim 1 *If Select is used in Procedure 1 to form a set at some constant iteration r, then Select* *will reveal overall edge dependencies of at most $n^{1/5}$ vertices with high probability.*

3.1 Expectation of Y_r

The proofs of Proposition 1 and 2 are inductive and one uses the other to establish the inductive step.

Proof of Proposition 1: In the base case $r = 1$, two sets A and B of size k each have a vertex in common with probability r_{kk} given by Equation (1). Because of Select, the number of vertices in P common to these two sets follows a binomial distribution $B(k; n_0, r_{kk})$. Since $r_{kk} \to 0$ and $n_0 r_{kk} \to k$, as $n \to \infty$, the binomial distribution can be approximated by a Poisson distribution with mean k. Then the random variable Y_1 is such that $\mathcal{E}[Y_1] \approx k = k^2/k$ and $var(Y_1) \approx k$, as required.

For any iteration t, where $1 < t \leq r-1$, we assume that Proposition 2 is true and Proposition 1 is also true thus implying that $\mathcal{E}[Y_t] \approx \mathcal{E}[Y_{t-1}]\mathcal{E}[Y_{t-1}]/k$, and $\mathcal{E}[Y_t] \approx k$ since $\mathcal{E}[Y_{t-1}] = \mu_A = \mu_B \approx k$. Let the expectations and variances of $|A|, |B|$ be μ_A, v_A and μ_B, v_B respectively. By the inductive hypothesis $\mu_A = \mu_B = \mathcal{E}[Y_{r-1}] \approx k$ and by Proposition 2, $v_A = v_B = var(Y_{r-1})$ is also bounded above by a term dependent on k and the iteration and independent of n. We then prove the statement of Proposition 1 for iteration r. Let set A be of size i with probability p_i and B be of size j with probability q_j. We then get the following.

$$\mathcal{E}[Y_r] = \sum_{l=0}^{n_0} l \cdot \sum_{i,j} p_i \cdot q_j \cdot \tbinom{n_0}{l} {r_{ij}}^l (1 - r_{ij})^{n_0 - l}.$$

We now take into consideration the properties of the binomial distribution and get the following three cases, depending on the range of the indices i, j with respect to $\sqrt{nk}$:

1. i and j are less than $\sqrt{nk}$,
2. i and j are both at least $\sqrt{nk}$,
3. otherwise.

Let $\mathcal{E}_m[Y_r]$ ($m = 1, 2, 3$) be the contribution to the expectation $\mathcal{E}[Y_r]$ of each of the first two Cases, and for Case 3, for each of its two subcases so that

$$\mathcal{E}[Y_r] = \mathcal{E}_1[Y_r] + \mathcal{E}_2[Y_r] + 2 \cdot \mathcal{E}_3[Y_r].$$

We use the following inequalities.

$$(1 - 1/n)^n \leq 1/e \leq (1 - 1/n)^{n-1} \text{ }, e^{-x} \geq 1 - x \quad (x \geq 0) \text{ }, e^{-x} \leq 1 - x + x^2/2 \quad (0 \leq x < 1).$$

Case 1: Since

$$r_{ij} = (1-(1-p)^i)(1-(1-p)^j) \leq \frac{(i\cdot p)}{(1-p)}\frac{(j\cdot p)}{(1-p)} = \frac{i\cdot j}{k\cdot n}\cdot\frac{1}{(1-p)^2} \approx \frac{i\cdot j}{k\cdot n},$$

and

$$\begin{aligned} r_{ij} &= (1-(1-p)^i)(1-(1-p)^j) \geq (1-e^{-p\cdot i})\cdot(1-e^{-p\cdot j}) \\ &\geq (i\cdot p - i^2\cdot p^2/2)\cdot(j\cdot p - j^2\cdot p^2/2) = \left(\frac{i\cdot j}{k\cdot n} - \frac{i^2\cdot j}{2\cdot(k\cdot n)^{3/2}} - \frac{j^2\cdot i}{2\cdot(k\cdot n)^{3/2}} + \frac{i^2\cdot j^2}{4\cdot(k\cdot n)^2}\right), \end{aligned}$$

we get

$$\begin{aligned} \mathcal{E}_1[Y_r] &= \sum_{i<\sqrt{nk}, j<\sqrt{nk}} p_i\cdot q_j \sum_l l\cdot \tbinom{n_0}{l} {r_{ij}}^l (1-r_{ij})^{n_0-l} = \sum_{i,j} p_i\cdot q_j\cdot n_0\cdot r_{ij} \\ &\leq \sum_{i,j} p_i\cdot q_j\cdot n_0\cdot\frac{i\cdot j}{kn}\cdot\frac{1}{(1-p)^2} \approx \sum_{i,j} p_i\cdot q_j\cdot\frac{i\cdot j}{k} = \frac{1}{k}\cdot\sum_i i\cdot p_i\cdot\sum_j j\cdot q_j \leq \frac{\mu_1\cdot\mu_2}{k}. \end{aligned}$$

We examine the lower bound for r_{ij} and get that

$$\begin{aligned} \mathcal{E}_1[Y_r] &= \sum_{i<\sqrt{nk}, j<\sqrt{nk}} p_i\cdot q_j \sum_l l\cdot \tbinom{n_0}{l} {r_{ij}}^l (1-r_{ij})^{n_0-l} = \sum_{i,j<\sqrt{kn}} p_i\cdot q_j\cdot n_0\cdot r_{ij} \\ &\geq \sum_{i,j<\sqrt{kn}} p_i\cdot q_j\cdot n_0\cdot\left(\frac{i\cdot j}{k\cdot n} - \frac{i^2\cdot j}{2\cdot(k\cdot n)^{3/2}} - \frac{j^2\cdot i}{2\cdot(k\cdot n)^{3/2}} + \frac{i^2\cdot j^2}{4\cdot(k\cdot n)^2}\right) \\ &\approx \sum_{i,j<\sqrt{kn}} p_i\cdot q_j\cdot\left(\frac{i\cdot j}{k} - O(1/\sqrt{n})\right) \\ &\geq \sum_{i,j} p_i\cdot q_j\cdot\left(\frac{i\cdot j}{k} - O(1/n^{1/4})\right) \approx \frac{1}{k}\cdot\sum_i i\cdot p_i\cdot\sum_j j\cdot q_j = \frac{\mu_1\cdot\mu_2}{k}. \end{aligned}$$

We explain below how the various inequalities were derived. We used the inductive assumption for the mean and the variance to get

$$\begin{aligned} \sum_{i<\sqrt{nk},\ j<\sqrt{kn}} ijp_iq_j &= \sum_{i,j} ijp_iq_j - \sum_{i\geq\sqrt{kn}, j\geq\sqrt{kn}} ijp_iq_j \\ &\quad - \sum_{i\geq\sqrt{kn}, j<\sqrt{kn}} ijp_iq_j - \sum_{i<\sqrt{kn}, j\geq\sqrt{kn}} ijp_iq_j \geq \sum_{i,j} ijp_iq_j - O(1/n^{1/4}). \end{aligned}$$

In order to show this last derivation we need to show the case for

$$\begin{aligned} \sum_{i\geq\sqrt{kn}, j<\sqrt{kn}} ijp_iq_j &= \sum_{i\geq\sqrt{kn}} ip_i \sum_{j<\sqrt{kn}} jq_j \leq \sum_{i\geq\sqrt{kn}} ip_i \sum_j jq_j \leq \sum_{i\geq\sqrt{kn}} ip_i\mu_2 \\ &= \sum_{n\geq i>n^{3/4}} ip_i\mu_2 + \sum_{n^{3/4}\geq i\geq\sqrt{kn}} ip_i\mu_2 \leq n\sum_{i\geq n^{3/4}} p_i\mu_2 + n^{3/4}\sum_{i\geq\sqrt{kn}} p_i\mu_2 \\ &\leq n\frac{v_1+\mu_1^2}{n^{1.5}}\mu_2 + n^{3/4}\frac{v_1+\mu_1^2}{kn}\mu_2 = O(1/n^{1/2}) + O(1/n^{1/4}), \end{aligned}$$

where we used Chebyshev's inequality [7] to get the fourth inequality. The other two cases can be derived similarly. The following is also true.

$$\sum_{i,j<\sqrt{kn}} p_i \cdot q_j \cdot n_0 \cdot \left(-\frac{i^2 \cdot j}{2 \cdot (kn)^{3/2}} - \frac{j^2 \cdot i}{2 \cdot (kn)^{3/2}} + \frac{i^2 \cdot j^2}{4 \cdot (kn)^2} \right) \geq -O(1/\sqrt{n}). \tag{2}$$

In order to prove this claim we need to use the inductive hypothesis and the fact that $var(X) = \mathcal{E}[X^2] - (\mathcal{E}[X])^2$, so that $\sum_i i^2 p_i$ is bounded above by a term dependent on μ_1 and v_1. A similar claim can be made for the sum $\sum_j j^2 q_j$. Then the first term of the sum in Equation (2) gives

$$\begin{aligned}
\sum_{i,j<\sqrt{kn}} p_i \cdot q_j \cdot n_0 \cdot \left(-\frac{i^2 \cdot j}{2 \cdot (kn)^{3/2}} \right) &= \left(-\sum_{i<\sqrt{nk}} p_i \cdot i^2 \right) \cdot \left(\frac{1}{2 \cdot (kn)^{3/2}} \cdot \sum_{j<\sqrt{kn}} j \cdot q_j \right) \\
&\geq \left(-\sum_{i<\sqrt{nk}} p_i \cdot i^2 \right) \cdot \left(\frac{1}{2 \cdot (k^3 \cdot n)^{1/2}} \right) \cdot \mu_2 \\
&\geq -O(1/\sqrt{n}) \approx 0.
\end{aligned}$$

The second term of the sum in Equation (2) is treated similarly and the last one is at least zero.

Case 2: We now consider the expression

$$\begin{aligned}
\mathcal{E}_2[Y_r] &= \sum_l l \sum_{i\geq\sqrt{nk}, j\geq\sqrt{nk}} p_i q_j \tbinom{n_0}{l} {r_{ij}}^l (1 - r_{ij})^{n_0 - l} = \sum_{i,j} p_i q_j n_0 r_{ij} \\
&\leq \sum_{i,j\geq\sqrt{nk}} p_i q_j r_{ij} n_0 = n_0 \sum_{i\geq\sqrt{nk}} p_i \sum_{j\geq\sqrt{nk}} q_j \\
&\leq \frac{{\mu_1}^2 + v_1}{nk} \frac{{\mu_2}^2 + v_2}{nk} n_0 \leq c_1 \frac{1}{n} \quad \text{for some constant } c_1 \\
&\leq c_1 \frac{1}{n} \approx 0,
\end{aligned}$$

where we claimed Chebyshev's inequality to get an upper bound for $\sum_{i\geq\sqrt{nk}} p_i = Pr(i \geq \sqrt{nk})$. By the inductive hypothesis, the contribution of this term to the expectation $\mathcal{E}[Y_r]$ is negligible, thus deriving the last two inequalities above. We now show that the contribution to the expectation when the indices are constrained as in Case 3 is also negligible.

Case 3: For Case 3 suppose $i < \sqrt{kn}$ and $j \geq \sqrt{nk}$. The other case is the one with $j < \sqrt{kn}$ and $i \geq \sqrt{nk}$, which can be treated similarly. Then we have

$$pi - p^2 \cdot i^2/2 \leq r_{ij} = (1 - (1-p)^i) \cdot (1 - (1-p)^j) \leq p \cdot i \cdot \frac{1}{1-p} = \frac{i}{\sqrt{nk}} \cdot \frac{1}{1-p},$$

and therefore, following the steps of the previous cases we get

$$\mathcal{E}_3[Y_r] = \sum_l l \sum_{i,j} p_i \cdot q_j \cdot \tbinom{n_0}{l} {r_{ij}}^l (1 - r_{ij})^{n_0 - l} \leq \sum_{i<\sqrt{nk}, j\geq\sqrt{nk}} p_i \cdot q_j \cdot n_0 \cdot \frac{i}{\sqrt{nk}} \cdot \frac{1}{1-p}$$

$$= \sum_{i<\sqrt{nk}} i \cdot p_i \sum_{j\geq\sqrt{nk}} q_j \cdot \frac{1}{\sqrt{nk}} \cdot n_0 \cdot \frac{1}{1-p} \leq \mu_1 \cdot \frac{v_2 + {\mu_2}^2}{kn} \cdot n_0 \cdot \frac{1}{\sqrt{nk}} \cdot \frac{1}{1-p} \to 0 \text{ as } n \to \infty.$$

We then combine the various upper and lower bounds for $\mathcal{E}_i[Y_r]$, $i = 1, 2, 3$, to get that the iterated mean is asymptotically for large n

$$\mathcal{E}[Y_r] \approx \frac{\mu_1 \cdot \mu_2}{k}.$$

We also have by the inductive hypothesis that $\mu_1 = \mu_2 \approx k$ so that $\mathcal{E}[Y_r] \approx k$, asymptotically for large n, for every constant iteration r. $\square$

3.2 Variance of Y_r

We follow the terminology and techniques of the previous section to find the variance of Y_r, $r \geq 1$. The proof of Proposition 2 follows.

Proof of Proposition 2: It is true by definition [7] that $var(Y_r) = \mathcal{E}[Y_r^2] - \mathcal{E}^2[Y_r]$. The three cases distinguished in the Proof of Proposition 1 depending on the ranges of i, j are also considered here. Case 1, following the steps of the proof of Proposition 1 and using the same approximations for r_{ij}, gives

$$\begin{aligned}
\mathcal{E}_1[Y_r^2] &= \sum_l l^2 \sum_{i,j<\sqrt{kn}} \tbinom{n_0}{l} {r_{ij}}^l (1-r_{ij})^{n_0-l} p_i q_j = \sum_{i,j<\sqrt{kn}} p_i q_j \sum_l l^2 \tbinom{n_0}{l} {r_{ij}}^l (1-r_{ij})^{n_0-l} \\
&\approx \sum_{i,j<\sqrt{kn}} p_i q_j (n_0^2 r_{ij}^2 + n_0 r_{ij}(1-r_{ij})) \\
&\leq \sum_{i,j} p_i q_j \left(\frac{n_0^2 i^2 j^2}{(k^2 n^2)(1-p)^4} + \frac{n_0 ij}{(kn)(1-p)^2} \right) \\
&\approx \frac{\sum i^2 p_i \sum j^2 q_j}{k^2} + \frac{\sum i p_i \sum j q_j}{k} \approx \frac{(v_1 + {\mu_1}^2)(v_2 + {\mu_2}^2)}{k^2} + \frac{\mu_1 \mu_2}{k}.
\end{aligned}$$

Cantelli's inequality [7] yields for a non-negative random variable X with mean μ and variance v the following bound:

$$Pr(X > a + \mu) \leq \frac{v}{v + a^2} \quad (a > 0)\ .$$

We apply Cantelli's inequality in Case 2 to get the second inequality below.

$$\begin{aligned}
\mathcal{E}_2[Y_r^2] &= \sum_l l^2 \sum_{i,j\geq\sqrt{kn}} \tbinom{n_0}{l} {r_{ij}}^l (1-r_{ij})^{n_0-l} p_i q_j \approx \sum_{i,j\geq\sqrt{kn}} p_i q_j (n_0^2 r_{ij}^2 + n_0 r_{ij}(1-r_{ij})) \\
&\leq \sum_{i\geq\sqrt{nk}, j\geq\sqrt{kn}} p_i q_j (n_0^2 + n_0) = \sum_{i\geq\sqrt{nk}} p_i \sum_{j\geq\sqrt{nk}} q_j n_0^2 + \sum_{i\geq\sqrt{nk}} p_i \sum_{j\geq\sqrt{nk}} q_j n_0 \\
&\leq \frac{v_1}{v_1 + kn} \frac{v_2}{v_2 + kn} n_0^2 + \frac{v_1}{v_1 + kn} \frac{v_2}{v_2 + kn} n_0 \approx \frac{v_1 v_2}{k^2}.
\end{aligned}$$

For $i < \sqrt{nk}$ and $j \geq \sqrt{kn}$ of Case 3 we have that $r_{ij} \leq i/\sqrt{kn}$ and subsequently

$$\begin{aligned}
\mathcal{E}_3[Y_r^2] &= \sum_l l^2 \sum_{i<\sqrt{kn}, j \geq \sqrt{kn}} \tbinom{n_0}{l} {r_{ij}}^l (1-r_{ij})^{n_0-l} p_i q_j \approx \sum_{i<\sqrt{kn}, j \geq \sqrt{kn}} p_i q_j (n_0^2 r_{ij}^2 + n_0 r_{ij}(1-r_{ij})) \\
&\leq \sum_{i<\sqrt{nk}} p_i \cdot \left(\frac{i}{\sqrt{nk}}\right)^2 \cdot \sum_{j \geq \sqrt{nk}} q_j \cdot n_0^2 + \sum_{i<\sqrt{nk}} p_i \cdot \frac{i}{\sqrt{nk}} \sum_{j \geq \sqrt{nk}} q_j \cdot n_0 \\
&\leq \frac{v_1+\mu_1^2}{kn} \cdot n_0^2 \cdot \frac{v_2}{v_2+kn} + \frac{v_1+\mu_1^2}{kn} \cdot n_0 \cdot \frac{v_2}{v_2+kn} \approx \frac{(v_1+\mu_1^2)v_2}{k^2}.
\end{aligned}$$

The symmetric case $j < \sqrt{nk}$ and $i \geq \sqrt{kn}$ of Case 3 yields a similar bound.

Adding all contributions in the limit $n \to \infty$ we get

$$var(Y_r) \;=\; \mathcal{E}[Y_r^2] - \mathcal{E}^2[Y_r] \leq \frac{\mu_1\mu_2}{k} + \frac{4v_1v_2}{k^2} + \frac{2{\mu_1}^2 v_2 + 2{\mu_2}^2 v_1}{k^2},$$

as desired. By the inductive hypothesis of Propositions 1 and 2, μ_1, μ_2 and v_1, v_2 depend on k and the iteration only and not on n therefore $var(Y_r)$ is bounded above by a term that depends on k and the iteration only and not on n. □

The Proof of Claim 1 that utilizes Propositions 1 and 2 follows.

Proof of Claim 1: In Propositions 1 and 2 we showed that the mean and variance of Y_r, for constant r, depends on k and the iteration only and is thus constant for a constant r.

The probability of having a set of common vertices formed at some iteration $i \leq r$ of size greater than $n^{1/5}/2^{r+1}$ by Chebyshev's inequality (see [7]) and Propositions 1 and 2 is at most $1/\Theta(n^{2/5})$. There are at most 2^r sets formed at iterations $1 \leq i \leq r$, therefore the probability that the vertices of all these sets is more than $(n^{1/5}/2^{r+1})\ 2^r \leq n^{1/5}$ overall is also, by subadditivity, $1/\Theta(n^{2/5})$. This proves Claim 1. □

4 Left and Right tails of Y_r

We examine first the left tail of the distribution of Y_r. The following definition is needed.

Definition 4 *In Procedure 1 a set formed at iteration r is* declared *to be of size at most* $(1-a)^{2^r-1}k$*, where* $0 < a < 1$*, if either this set is of size at most* $(1-a)^{2^r-1}k$*, or any of the sets formed at some intermediate iteration* $m < r$ *is of size at most* $(1-a)^{2^m-1}k$*.*

The probability that a set is of size at most $(1-a)^{2^r-1}k$ is bounded above by the probability that this set is declared to be of size at most $(1-a)^{2^r-1}k$. The latter probability will be easier to find.

In the tree of iteration r, at iteration zero, 2^r sets each of size k, are paired in 2^{r-1} pairs to form 2^{r-1} sets of vertices at iteration one. Sets are formed according to Select and Procedure 1. Then, a vertex of set P (of Select) is common to two such sets of size exactly k with probability r_{kk} given by Equation (1). The size of a set formed at iteration one follows a binomial distribution with parameters n_0 and r_{kk}. The bounds for the left tail of the binomial distribution [2] are applied, since $r_{kk} \cdot n_0 \approx k$, to get that the probability of having a set of size less than $(1-a)k$ formed at iteration one is at most $exp(-a^2k/2)$, for large n. Hence, the probability that any of the 2^{r-1} sets of iteration one is of size at most $(1-a)k$ is, for large n, at most $2^{r-1}exp(-a^2k/2)$. We shall call this term p_1.

Similarly for two sets formed at iteration one whose sizes are at least $(1-a)k$ each, the probability that a vertex in P is common to both sets is at least, by Equation (1), $r_{(1-a)k\ (1-ak)}$. Thus at iteration two the mean $r_{(1-a)k\ (1-a)k}n_0$ of the size of the set common to the two sets of iteration one is by Proposition 1 and in the limit $n \to \infty$ at least $(1-a)^2 \cdot k$. Thus by the bounds in [2] the set formed at iteration two is of size at most $(1-a)^3k$, which is at most $(1-a)$ times the lower bound of its mean, with probability $exp(-a^2(1-a)^2k/2)$ in the limit $n \to \infty$. Since there are 2^{r-2} such sets in the tree of iteration r, we conclude that the probability that any of these sets is of size at most $(1-a)^3k$ is at most $2^{r-2} \cdot exp(-a^2(1-a)^2k/2)$. We call this last term p_2. Therefore, with probability at least $1-p_1-p_2$ for $n \to \infty$ all sets formed at iteration one are of size at least $(1-a)k$ and all sets formed at iteration two are of size at least $(1-a)^2k$. Otherwise we declare the set that is to be formed at iteration r to be of size at most $(1-a)^{2^r-1}k$. Inductively at iteration $l \leq r$, given that all 2^{r-l+1} sets formed at iteration $l-1$ are of size at least $(1-a)^{2^{l-1}-1} \cdot k$, the size of a set common to two such sets and thus formed at iteration l has mean at least $(1-a)^{2^l-2}k$. Hence, the probability that this set is of size at most $(1-a)^{2^l-1} \cdot k$ is at most $exp(-a^2(1-a)^{2^l-2}k/2)$. We then conclude that in the tree of iteration r, the probability that any of the 2^{r-l} sets formed at iteration l is of size at most $(1-a)^{2^l-1} \cdot k$ is at most $2^{r-l} \cdot exp(-a^2(1-a)^{2^l-2}k/2)$. We call this term p_l. Consequently, with probability at least $1-p_1-\ldots-p_l$ all 2^{r-i} sets at iteration $i \leq l$ are of size at least $(1-a)^{2^i-1} \cdot k$, otherwise the unique set at iteration r is declared to have size at most $(1-a)^{2^r-1}k$.

This way for $l = r$ we get Theorem 1 stated in section 2. The bound given in Theorem 1, for large n, is an upper bound of the probability that a set at iteration r is of size at most $(1-a)^{2^r-1} \cdot k$.

For a constant a, the term of the sum in Theorem 1 for $\lambda = r - 1$ becomes equal to $1/e$, for

$$r = \lg\left(\frac{\log\left(a^2\ k/2\right)}{-\log\left(1-a\right)} + 2\right),$$

and thus r can grow as much as $O(\mathrm{lglg}k)$ before this expression gives trivial bounds.

The discussion of the right tail of Y_r is symmetric to the discussion for the left tail except that the bound in [2] for the right tail of the binomial distribution is used. A definition similar to Definition 4 is also needed.

Definition 5 *In Procedure 1 a set formed at iteration r is* declared *to be of size at most* $(1+a)^{2^r-1}k$*, where* $0 < a < 1$*, if either this set is of size at most* $(1+a)^{2^r-1}k$*, or any of the sets formed at some intermediate iteration* $m < r$ *is of size at most* $(1+a)^{2^m-1} \cdot k$.

Similarly to Theorem 1 we then get Theorem 2 stated in section 2. Since a is positive, $(1+a)^{2^r-1} \cdot k$ is polynomial in k as long as $r = \Theta(\lg\lg k)$. If however $a > 1$ then for constant r and a bound $(1+a)^{2^r-1} \cdot k$ that may depend on n one will have to rely on the DeMoivre-Laplace Theorem or its extensions (see [5],pages 13-14). If one claims Theorem 7 of [5] (page 14) that states that for $u \cdot q > 2$, $q = 1 - p$, and $p \cdot n \geq 1$

$$P(S_{n,p} \geq u \cdot p \cdot n) \leq (e/u)^{u \cdot p \cdot n}.$$

and repeating the claims that resulted in the proof of Theorem 2, using u in place of $(1 + a)$ and using the DeMoivre-Laplace bound instead of the Angluin-Valiant bounds, the variant of Theorem 2 expressed by Theorem 3 is derived.

Let in Theorem 3 choose $u^{2^r-1} \cdot k = \sqrt{k \cdot n}$. This requires $u = (\sqrt{n/k})^{1/2^{r-1}}$. For r a constant it is straightforward that the upper bound given by Theorem 3 decreases exponentially fast with n. This in turn shows that a sum of the form $\sum_{i \geq \sqrt{kn}} p_i$ like those calculated in Proposition 2 is at most $1/n^3$ for sufficiently large n. The tighter bound claimed in Proposition 3 follows easily as the contributions of $\mathcal{E}_2[Y_r^2], \mathcal{E}_3[Y_r^2]$ become $o(1)$.

5 Introduction to asymptotic results for large k

In all previous sections we examined the distribution of the random variable Y_r in $G_{n,1/(kn)^{1/2}}$ random graphs for a constant k. In this section we examine the distribution of Y_r for large values of k in addition to having $n \to \infty$. The redundancy parameter k remains in this discussion independent of n. We first introduce some results from Probability Theory.

Let $\{U_m\}$, $\{V_m\}$ be sequences of random variables. We are interested in cases with large m.

Definition 6 (Convergence in Distribution) *The sequence $\{U_m\}$ converges in distribution to the cumulative distribution function $F(u)$ if for all u that are continuity points of $F(u)$*

$$\lim Pr(U_m \le u) = F(u)$$

Equivalently one may require that $\mathcal{E}[h(U_m)]$ tends to $\mathcal{E}[h(U)]$, where U is a random variable having probability distribution function $F(u)$ and $h(u)$ is any bounded continuous function.

For the sake of an example, the Central Limit Theorem states that for a positive integer n, $X_1, \ldots, X_n$ independent and identically distributed random variables with mean μ and variance σ^2, and for each u, $F_{U_n}(u)$, where $U_n = (\overline{X}_n - \mathcal{E}[\overline{X}_n])/(\sigma/\sqrt{n})$, converges in distribution to $\Phi(u)$ as n approaches infinity. $\Phi(\cdot)$ denotes the cumulative distribution function of the standardized normal distribution $N(0,1)$. Therefore $\overline{X}_n$ is approximately distributed $N(\mu, \sigma^2/n)$, and thus asymptotically normal.

The notation $U \simeq F$ will be used in such convergence cases. We shall thus write $X_m \simeq N(\mu, \sigma^2)$ to indicate that the sequence $\{X_m\}$ is asymptotically normal that is, it converges in distribution to the normal distribution with mean μ and variance σ^2. We shall also write $U \sim F$, $U \sim V$ to indicate that U has distribution function F and U and V have the same distribution.

For sequence U_m we shall write $U_m \sim asN(\mu_m, \sigma_m^2)$ to indicate that U_m is asymptotically normal that is, $(U_m - \mu_m)/\sigma_m$ converges in distribution to the standard normal distribution.

In many cases we shall use the "=" sign to indicate a limit but such a usage will be evident by the context of the relevant arguments.

Let X be a random variable that follows, possibly in the limit, a Poisson distribution with mean k. We examine below the limiting distribution of the product aX, where a is a positive real number, for large k. We shall show that for large k, aX has a normal approximation.

The exponential generating function (e.g.f. in short) of a Poisson random variable with expectation k is $\mathcal{E}[exp(t\ X)] = exp(-k(1 - exp(t)))$ [7]. In order to examine the limiting distribution of aX, we first standardize X thus examining the e.g.f. of $(aX - \mathcal{E}[aX])/\sqrt{var(aX)}$.

$$\begin{aligned}
\mathcal{E}[e^{t\ a\ \frac{X-k}{a\ \sqrt{k}}}] &= e^{-t\ \sqrt{k}}\ \mathcal{E}[e^{t\ \frac{X}{\sqrt{k}}}] = e^{-t\ \sqrt{k}}\ e^{-k(1-e^{t/\sqrt{k}})} \\
&= e^{-t\ \sqrt{k}}\ e^{-k(1-(1+\frac{t/\sqrt{k}}{1!}+\frac{(t/\sqrt{k})^2}{2!}+\ldots+\frac{(t/\sqrt{k})^j}{j!}+\ldots))} = e^{1/2t^2+\frac{t^3}{3\sqrt{k}}+O(1/k)}.
\end{aligned}$$

For large k all terms of the exponent of the e.g.f. of $(aX - \mathcal{E}[aX])/\sqrt{var(aX)}$ other than the first one tend to zero. Hence, the standardized random variable has an exponential generating function approaching in the limit for large k that of the standard normal distribution. Therefore, for large k,

$$\frac{aX - ak}{a\sqrt{k}} \simeq N(0,1),$$

and consequently,

$$aX \sim asN(a\,k, a^2\,k).$$

6 Behavior of Y_r for large k

Let X_1, X_2 be two independent random variables which follow, possibly in the limit, the same distribution $N(\mu, \sigma^2)$. We are interested in the behavior of the product $X_1 X_2$.

Fact 1 *Let X_1, X_2 be two independent random variables that follow, possibly in the limit, a normal $N(\mu, \sigma^2)$ distribution. Then,*

$$\mathcal{E}[X_1 X_2] = \mu^2, \quad \text{and} \quad \mathcal{E}[X_1^2 X_2^2] = (\mu^2 + \sigma^2)^2.$$

Proof: The central moments $\mathcal{E}[(X_1 X_2)^r]$ can be derived easily by considering the exponential generating function of a bivariate $X = (X_1, X_2)$ normal distribution, where the correlation between X_1 and X_2 is zero since they are independent, and using the fact (see say, [3], page 131) that $\mathcal{E}[(X_1 X_2)^r]$ is the coefficient of $(t_1^r\ t_2^r)/(r!\ r!)$ in the expansion of $m_X(t_1, t_2) = \mathcal{E}[e^{tX}] = \mathcal{E}[e^{t_1 X_1 + t_2 X_2}]$. An expansion of the exponential generating function of a bivariate normal distribution, remembering that

$$\mathcal{E}[e^{t_1 X_1 + t_2 X_2}] = e^{\mu\, t_1 + 1/2\sigma^2 t_1^2 + \mu t_2 + 1/2\sigma^2 t_2^2},$$

gives the following results after some calculations.

$$\mathcal{E}[X_1 X_2] = \mu^2,$$

$$\mathcal{E}[X_1^2 X_2^2] = (\mu^2 + \sigma^2)^2,$$

and also,

$$\mathcal{E}[X_1^3 X_2^3] = \mu^2(\mu^2 + 3\,\sigma^2)^2.$$

[In case X_1, X_2 are in the limit $N(\mu, \sigma^2)$, the equality signs in the three formulae above should be interpreted as limits.] $\square$

Although this way we can find the moments of the product X_1X_2, we cannot get its exponential generating function easily. We now give a direct method for finding the e.g.f. of $X = X_1X_2$. This is expressed by Theorem 4 of section 2 proven below.

Proof of Theorem 4: For either of the two independent random variables X_1, X_2 we have possibly in the limit for large k that $\mathcal{E}[e^{tX_i}] = e^{\mu t+1/2\sigma^2 t^2}$, with $\mu = \Theta(k)$, $\sigma^2 = \Theta(k)$.

Since X_1, X_2 are also independent, we get the following.

$$\begin{aligned}
m_X(t) &= \mathcal{E}[e^{tX}] \\
&= \mathcal{E}[e^{\mu t X_1 + \frac{1}{2}\sigma^2 t^2 X_1^2}] \\
&= \frac{1}{\sqrt{2\pi}\ \sigma}\int_{-\infty}^{\infty} e^{\mu\ t\ x+\frac{1}{2}\ \sigma^2\ t^2\ x^2}\cdot e^{-\frac{(x-\mu)^2}{2\ \sigma^2}}\ dx \\
&= \frac{1}{\sqrt{1-\sigma^4\ t^2}}\ \frac{1}{\sqrt{2\ \pi}\ \frac{\sigma}{\sqrt{1-\sigma^4\ t^2}}}\int_{-\infty}^{\infty} e^{-\frac{\left(x-\frac{\mu+\sigma^2\ \mu\ t}{1-\sigma^4\ t^2}\right)^2}{2\ \frac{\sigma^2}{1-\sigma^4\ t^2}}}\cdot e^{\frac{\mu^2\ t\ \left(\sigma^2\ t+1\right)}{(1-\sigma^4\ t^2)}}\ dx \\
&= \frac{1}{\sqrt{1-\sigma^4\ t^2}}\ e^{\frac{\mu^2\ t\ \left(\sigma^2\ t+1\right)}{(1-\sigma^4\ t^2)}}\ \frac{1}{\sqrt{2\ \pi}\ \frac{\sigma}{\sqrt{1-\sigma^4\ t^2}}}\int_{-\infty}^{\infty} e^{-\frac{\left(x-\frac{\mu+\sigma^2\ \mu\ t}{1-\sigma^4\ t^2}\right)^2}{2\ \frac{\sigma^2}{1-\sigma^4\ t^2}}}\ dx \\
&= \frac{1}{\sqrt{1-\sigma^4\ t^2}}\ e^{\frac{\mu^2\ t\ \left(\sigma^2\ t+1\right)}{(1-\sigma^4\ t^2)}}.
\end{aligned}$$

Since $\mathcal{E}[X^i] = m_X^{(i)}(0)$, the ith derivative of $m_X(t)$ at $t = 0$, we get after a number of calculations that

$$\mathcal{E}[X] = \mu^2,$$

$$\mathcal{E}[X^2] = \mu^4 + 2\ \mu^2\ \sigma^2 + \sigma^4,$$

$$\mathcal{E}[X^3] = \mu^6 + 6\ \mu^4\ \sigma^2 + 9\ \mu_2\ \sigma^4.$$

Therefore, the variance of X is $var(X) = \mathcal{E}[X^2] - \mathcal{E}^2[X] = \sigma^4 + 2\ \mu^2\ \sigma^2$.

We are interested in examining the limiting behavior of X for large k, for the case where both μ and σ^2 are such that $\mu = \Theta(k)$, $\sigma^2 = \Theta(k)$, where k is some parameter of interest. We standardize X and examine the e.g.f. of $(X - \mathcal{E}[X])/\sqrt{var(X)}$ by taking into consideration the e.g.f. of X derived earlier.

$$\begin{aligned}
\mathcal{E}[e^{t\ \frac{X-\mu^2}{\sqrt{\sigma^4+2\ \mu^2\ \sigma^2}}}] &= e^{-t\frac{\mu^2}{\sqrt{\sigma^4+2\ \mu^2\ \sigma^2}}}\cdot\mathcal{E}[e^{\frac{t\ X_1\ X_2}{\sqrt{\sigma^4+2\ \mu^2\ \sigma^2}}}] \\
&= e^{-t\frac{\mu^2}{\sqrt{\sigma^4+2\ \mu^2\ \sigma^2}}}\cdot\frac{1}{\left(1-\sigma^4\ \frac{t^2}{\sigma^4+2\ \mu^2\ \sigma^2}\right)^{1/2}}\ e^{\frac{1}{\left(1-\sigma^4\ \frac{t^2}{\sigma^4+2\ \mu^2\ \sigma^2}\right)}\ \frac{\mu^2\ \sigma^2\ t^2}{\sigma^4+2\ \mu^2\ \sigma^2}}
\end{aligned}$$

$$\cdot \quad e^{\frac{1}{\left(1-\sigma^4 \frac{t^2}{\sigma^4+2\,\mu^2\,\sigma^2}\right)} \frac{\mu^2\,t}{\sqrt{\sigma^4+2\,\mu^2\,\sigma^2}}}.$$

For $\mu = \Theta(k)$ and $\sigma^2 = \Theta(k)$ and for large k we have the convergences $(\sigma^4)/(2\mu^2\,\sigma^2+\sigma^4) \to 0$, $(\mu^2\,\sigma^2)/(2\mu^2\,\sigma^2+\sigma^4) \to 1/2$, thus getting

$$\mathcal{E}[e^{t\,\frac{X-\mu^2}{\sqrt{\sigma^4+2\,\mu^2\,\sigma^2}}}] \to e^{1/2t^2}.$$

Hence, in the limit for large k, the standardized X converges to the standard Normal distribution $N(0,1)$, which implies $X \sim asN(\mu^2, \sigma^4 + 2\,\mu^2\,\sigma^2)$. $\square$

Theorem 4 also holds if $X_1, X_2 \sim asN(\mu,\sigma^2)$ by interpreting the equalities in the proof of the theorem in the limit for large k. Suppose we have two sets A, B of sizes x_1, x_2 respectively. The sizes of the two sets can be treated as random variables and let X_1 (resp. X_2) be the random variable that assumes value x_1 (respectively x_2) when set A (respectively B) is of size x_1, (respectively x_2) with some probability. We resume the discussion of Problem 1.

Example 1. For sets A, B formed at iteration one, X_i, $i = 1, 2$, are binomial random variables $B(k; n_0, r_{kk})$ identically distributed with random variable Y_1 of section 2. Such a binomial random variable has a normal approximation with expectation and variance k for large k and $n \to \infty$. Theorem 4 then implies that $X_1X_2 \sim asN(k^2, 2k^3 + k^2)$. We have by Equation (1) that $r_{x_1,x_2} \approx (x_1x_2)/(kn)$, if $x_1, x_2 = o(\sqrt{n})$. We thus get that $x_1x_2 = o(n)$. Let $X = X_1X_2$. The probability that set C formed at iteration two and common to A, B has size j is given by the following expression.

$$Pr(Y_r = j) \approx \sum_i Pr(X = i)\ \binom{n_0}{j}(\tfrac{i}{k\,n})^j(1-\tfrac{i}{k\,n})^{n_0-j},$$

where Y_r is the random variable that represents the size of C. We examine the exponential generating function of Y_r by showing Proposition 4 below.

Proposition 4 *Let X be the random variable defined in Theorem 4 that, in the limit for large k, has a normal approximation $X \sim asN(k^2, \sigma^2)$, where σ^2 depends on k only. Then the random variable Y_r such that,*

$$Pr(Y_r = j) \approx \sum_i Pr(X = i)\ \binom{n_0}{j}(\tfrac{i}{k\,n})^j(1-\tfrac{i}{k\,n})^{n_0-j}$$

has an exponential generating function that is given, for large k and $n \to \infty$, by the following expression.

$$\mathcal{E}[e^{t\,Y_r}] \approx e^{(\frac{\frac{\sigma^2}{2}-k^3}{k^2})+\frac{1}{k^2}(k^3-\sigma^2)\,e^t+\frac{\sigma^2}{2\,k^2}\,e^{2t}}.$$

Hence, $\mathcal{E}[Y_r] \approx k$ *and* $var(Y_r) \approx k + \sigma^2/k^2$.

Proof of Proposition 4: The approximation $r_{x_1,x_2} \approx \frac{x_1\, x_2}{k\, n}$ holds when both x_1, x_2 are $o(\sqrt{n})$. For $i = \Omega(\sqrt{n})$, the sum over such i of $Pr(X = i)$ is $O(e^{-n})$. This can be proved as follows. By the limiting behavior of X we have

$$\sum_{i=\Omega(\sqrt{n})} Pr(X = i) \approx \frac{1}{\sqrt{2\pi}} \int_{\frac{\sqrt{n}-k^2}{\sigma}}^{\infty} e^{-x^2/2}\, dx = 1 - \Phi(\frac{\sqrt{n}-k^2}{\sigma}).$$

We now claim (see [5], page 9) that for $x \to \infty$,

$$1 - \Phi(x) \approx \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2},$$

A substitution for $x = (\sqrt{n} - k)/\sigma$ yields that in the limit $n \to \infty$ and for large k the sum $\sum_{i=\Omega(\sqrt{n})} Pr(X = i)$ tends to zero exponentially fast.

In the following discussion we shall examine the e.g.f. of Y_r in the limit for large k and $n \to \infty$. When we consider expectations we shall use the ones of the limit rather than that of the limiting quantity as this is implied by Definition 6 and the discussion following it.

We now calculate the exponential generating function $m_{Y_r}(t) = \mathcal{E}[e^{t\, Y_r}]$ of Y_r. We have that $X \sim asN(k^2, \sigma^2)$. We also approximate the binomial term in the expression for $Pr(Y_r = j)$, by a Poisson distribution with mean i/k. After some calculations similar to the ones that led to the calculation of $\mathcal{E}[e^{t\, X}]$ in the Proof of Theorem 4, we get the following.

$$\mathcal{E}[e^{t\, Y_r}] \approx e^{(\frac{\frac{\sigma^2}{2} - k^3}{k^2}) + \frac{1}{k^2}(k^3 - \sigma^2)\, e^t + \frac{\sigma^2}{2\, k^2}\, e^{2t}}.$$

From this expression we easily get that in the limit for large k

$$\mathcal{E}(Y_r) = m'_{Y_r}(0) \approx k,$$

and

$$var(Y_r) \quad = \quad -m'^2_{Y_r}(0) + m''_{Y_r}(0) \approx -k^2 + k^2 + k + \frac{\sigma^2}{k^2} = k + \frac{\sigma^2}{k^2}.$$

□

Example 2. From the discussion of Example 1 we get that at iteration two we have $var(X) = \sigma^2 \approx 2k^3 + k^2$ thus getting by Proposition 4 that $\mathcal{E}[Y_2] \approx k$ and $var(Y_2) \approx 3\, k + 1$.

We then standardize Y_2 and examine the generating function for the standardized random variable $Y_2' = (Y_2 - k)/(\sqrt{k + \frac{\sigma^2}{k^2}})$. Using the expansion $e^t = \sum_i t^i/i!$, we get the following

$$\mathcal{E}[e^{t\, Y_2'}] \approx e^{\frac{1}{2}t^2 + O(\frac{1}{\sqrt{k}} t^3)}.$$

For large enough k the exponent of e in $\mathcal{E}[e^{t\,Y_2'}]$ is asymptotically $1/2t^2$, that of the standard normal distribution. Thus for large k Y_2' is approximately $N(0,1)$ and consequently, Y_2 is approximately $N(k, 3k+1)$ at iteration 2. This allows us to proceed iteratively to the third iteration. Random variable Y_2 can similarly be shown to be approximately $N(k, 7k+11)$.

We shall then show that random variable Y_r previously defined at a constant iteration r approximates, in the limit for large k, $N(k, (2^r-1)k + O(1))$. The term $O(1)$ describes contributions which are due to the propagation in higher iterations of the additive one in the variance $3k+1$ of iteration two. For large iterations but small k it may be the case that this constant is the overall dominant term in the variance. We must thus bear in mind that the results we claim in this section hold for large k only. These results are summarized in Theorem 5 whose proof follows.

Proof of Theorem 5: The proof is by induction on the iteration. The basis of the induction has been proved in Example 1.

In the inductive step we take two sets A and B formed at iteration r, with sizes represented by random variables, which are identically distributed and approximately $N(k, (2^r-1)k+O(1))$. Let us call $b_r^2 = (2^r-1)k + O(1)$. The product of these random variables, by Theorem 4, can be thus approximated by a normal random variable $N(k^2, b_r^2 + 2\,k^2\,b_r) = N(k^2, \sigma_r^2)$.

Random variable Y_{r+1} that represents the size of set C formed at iteration $r+1$ and common to A and B has, by Proposition 4, expectation $\mathcal{E}[Y_{r+1}] \approx k$ independent of the iteration and variance $var(Y_{r+1}) \approx k + (b_r^4 + 2\,k^2\,b_r^2)/k^2$. Substituting for b_r^2 we get that in the limit for large k and constant r $var(Y_{r+1}) = (2^{r+1}-1)k + O(1)$.

The hidden constants grow exponentially fast with the iteration. That is, for $b_r^2 = (2^r-1)k + A_r$ we get $A_{r+1} = (2^r-1)^2 + A_r + A_r^2/k^2 + (2^{r+1}-2)A_r/k$.

We now standardize Y_{r+1} and for the standardized random variable $(Y_{r+1}-k)/(\sqrt{var(Y_{r+1})})$ we find its exponential generating function in the limit for large k. Let for clarity in the computations that will follow $k + \sigma_r^2/k^2 = var(Y_{r+1}) = (2^{r+1}-1)k + O(1)$. The limiting e.g.f. of Y is given by Proposition 4 as well. We thus get

$$
\begin{aligned}
\mathcal{E}[e^{t\,\frac{Y_{r+1}-k}{\sqrt{k+\sigma_r^2/k^2}}}] &= e^{-tk/\sqrt{k+\sigma_r^2/k^2}}\ \mathcal{E}[e^{t\,Y_{r+1}/\sqrt{k+\sigma_r^2/k^2}}] \\
&= e^{-tk/\sqrt{k+\sigma_r^2/k^2}}\ \exp\left((\frac{\sigma_r^2}{2\,k^2} - k) + \frac{1}{k^2}(k^3-\sigma_r^2)e^{t/\sqrt{k+\sigma_r^2/k^2}} + \frac{\sigma_r^2}{2\,k^2} e^{2\,t/\sqrt{k+\sigma_r^2/k^2}} \right) \\
&= e^{-tk/\sqrt{k+\sigma_r^2/k^2}}\ \exp\left(\frac{\sigma_r^2}{2\,k^2} - k \right)
\end{aligned}
$$

$$
\begin{aligned}
&\cdot\quad \exp\left(\frac{k^3-\sigma_r^2}{k^2}\ \left(1+\frac{t}{1!\ \sqrt{k+\sigma_r^2/k^2}}+\frac{t^2}{2!\ (\sqrt{k+\sigma_r^2/k^2})^2}+\frac{t^3}{3!\ (\sqrt{k+\sigma_r^2/k^2})^3}+\ldots\right)\right)\\
&\cdot\quad \exp\left(\frac{\sigma_r^2}{2\ k^2}\ \left(1+\frac{2\ t}{1!\ \sqrt{k+\sigma_r^2/k^2}}+\frac{4\ t^2}{2!\ (\sqrt{k+\sigma_r^2/k^2})^2}+\frac{8\ t^3}{3!\ (\sqrt{k+\sigma_r^2/k^2})^3}+\ldots\right)\right)\\
=&\quad e^{t^2/2+O\left((k+3\ \sigma_r^2/k^2)/3!\ (\sqrt{k+\sigma_r^2/k^2})^3\right)}.
\end{aligned}
$$

We then substitute the value for σ_r^2 and for large k the O term in the expression above becomes $O(1/\sqrt{2^r\ k})$ thus approaching zero. The standardized variable Y_r has then e.g.f. that approaches in the limit for large k that of the standard normal distribution and the claim of the Theorem follows. Note that at each iteration in order to establish our result we take limits over k. Since our claims hold for constant number of iterations only this eliminates problems caused by having a constant number of nested limits.□

6.1 Tails of the distribution

Since Y_r, the random variable representing the size of a set formed at some iteration r, for large enough k, $n \to \infty$, and constant r, follows a normal distribution $N(k,(2^r-1)k+O(1))$, we get that the probability that Y_r is zero is given by

$$
Pr(0 \le Y_r \le 1) = Pr(\frac{-k}{\sqrt{(2^r-1)k+O(1)}} \le \frac{Y_r-k}{\sqrt{(2^r-1)k+O(1)}} \le \frac{1-k}{\sqrt{(2^r-1)k+O(1)}}).
$$

Then

$$
\begin{aligned}
Pr(0 \le Y_r \le 1) &= \frac{1}{\sqrt{2\pi}}\int_{\frac{-k}{\sqrt{(2^r-1)k+O(1)}}}^{\frac{1-k}{\sqrt{(2^r-1)k+O(1)}}} e^{-x^2/2}\ dx\\
&= \frac{1}{\sqrt{2\pi}}\ \frac{1}{\sqrt{2^r\ k}} O(e^{-k/2^r}).
\end{aligned}
$$

7 Conclusions

We examined in this paper the properties of a class $G_{n,1/(kn)^{1/2}}$ of random graphs for an iterative set forming procedure described by Procedure 1. We got some non-trivial upper bounds for the tails of the distribution of random variable Y_r that represents the size of a set formed after r iterations of this procedure. We concluded a dependence of these bounds on the iteration and the redundancy parameter k. The bounds imply that moderate values of k (say, 25-50) are sufficient for various applications such as those involving the learning of simple boolean expressions and sets formed at the first few iterations of Procedure 1 have rarely sizes greater than $3-5$ times the parameter

k. We also showed some asymptotic results for large k. We proved a normal approximation for random variable Y_r.

The author would like to thank L.G.Valiant for suggesting this problem and for his various comments in an earlier draft of this work.

References

[1] Miklós Ajtai, János Komlós, and Endre Szemerédi. The longest path in a random graph. *Combinatorica*, 1:1–12, 1981.

[2] D. Angluin and L. G. Valiant. Fast probabilistic algorithms for hamiltonian circuits and matchings. *J. Computer and System Sciences*, 18:155–193, 1979.

[3] O. E. Barndorff-Nielsen and D. R. Cox. *Asymptotic Techniques for Use in Statistics*. Chapman and Hall, 1989.

[4] P. Billingsley. *Probability and Measure*. Wiley, New York, 1986.

[5] Béla Bollobás. *Random graphs*. Academic Press, New York 1985.

[6] P. Erdős, A. Rényi, and V. T. Sós. On a problem of graph theory. *Studia Scientarium Mathematicarum Hungarica*, 1, 1966.

[7] William Feller. *An Introduction to Probability Theory and its Applications*, Volume 2. Wiley, New York, 1971.

[8] A. V. Gerbessiotis. A graph-theoretic result for a model of neural computation. Discrete Applied Mathematics 82(1998), pp 257-262, Elsevier Science B.V., 1998.

[9] Alfred Rényi. *Probability Theory*. Akademiai Kiado, Budapest.

[10] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27:1134–1142, 1984.

[11] L. G. Valiant. Functionality in neural nets. *Proceedings of the American Association for Artificial Intelligence*, Volume 2, 1988.

[12] L. G. Valiant. Circuits of the mind. Oxford University Press, 1995.