On Induced Paths, Holes and Trees in Random Graphs

Kunal Dutta* and C.R. Subramanian †

Downloaded 11/20/19 to 58.247.249.188. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Abstract

The concentration of the sizes of largest induced paths and cycles (holes) are studied in the random graph model $\mathcal{G}(n, p)$. A 2-point concentration is proved for the size of the largest induced path and cycle, for all $p = p(n)$ satisfying $p \geq n^{-1/2}(\ln n)^2$ and $p \leq 1 - \epsilon$ where $\epsilon > 0$ is any constant. No such tight concentration (within two consecutive values) was known before for induced paths and cycles. As a corollary, a significant additive improvement is obtained over a 25-year old result of Erdős and Palka [8]concerning the size of the largest induced tree in a dense random graph. The proofs are based on second moment calculations and an explanation as to why more powerful concentration tools cannot be employed is also provided.

1 Introduction

In this paper, we study the problem of finding large induced subgraphs, such as paths, holes and trees in dense or morderately dense random graphs.

The $\mathcal{G}(n, p)$ model of random graphs, introduced by Erdős and Rényi [9] is the most widely studied model of random graphs. The $\mathcal{G}(n, p)$ model denotes the probability space over all labeled graphs having a specific vertex set (assumed without loss of generality to be $V = [n] := \{1, \ldots, n\}$) and obtained by including every potential edge $\{u, v\} \in \binom{V}{2}$ as an edge in the random graph independently and randomly with probability $p = p(n)$.

Notation : Given a natural number $n \in \mathcal{N}$, we indicate the set $\{1, \ldots, n\}$ by $[n]$. Define $q := (1-p)^{-1}$. We ignore floors and ceilings wherever they are not crucial. By the term *asymptotically almost surely* (a.a.s) in reference to a sequence of events $[E_n]_{n\geq 1}$, we mean that the sequence of probabilities $[P_n]_{n\geq 1}$ associated to the events $[E_n]$ tends to 1 as n tends to infinity. We use $\mathcal{B}(n, \mu)$ to denote the sum of n identically and independently distributed indicator variables each having mean μ. For non-negative integers n and b, we use $(n)_b$ to denote the expression $\prod_{0\leq j\leq b-1}(n - j)$. We use $\ln n$ to denote the natural logarithm of n. For a set A and an integer $k \geq 0$, we use $\binom{A}{k}$ to denote the collection $\{B \subseteq A : |B| = k\}$. We use $\mathcal{N}, \mathcal{Z}^+, \mathcal{R}^+$ to denote the sets of natural numbers, positive integers and positive reals respectively.

*DataShape, INRIA Sophia Antipolis – Méditerranée, Sophia Antipolis, France. Email: kdutta@mpi-inf.mpg.de.

†The Institute of Mathematical Sciences, HBNI, Taramani, Chennai - 600113, India. Email: crs@imsc.res.in

1.1 Previous Work

The problem of finding large induced trees in the random graph $\mathcal{G}(n, p)$ was first studied by Erdős and Palka in [8]. Given a graph G, denote by $T(G)$ the size (= number of vertices) of any largest induced tree in G. Erdős and Palka showed that

THEOREM 1.1. *For every $\epsilon > 0$, for every fixed $p : 0 < p < 1$, a.a.s. $G \in \mathcal{G}(n, p)$ has $T(G)$ satisfying*

$$(2 - \epsilon) \log_q n \quad < T(G) \quad < (2 + \epsilon) \log_q n.$$

They also conjectured that for $p = c/n$ ($c > 1$ is any constant), $G \in \mathcal{G}(n, p)$ almost surely has an induced tree of size $\gamma(c)n$ where $\gamma(c)$ depends only on c. This was verified affirmatively and independently by de la Vega [6] and several others including Frieze and Jackson [10], Kučera and Rödl [13], and Łuczak and Palka [16] who showed that when $p = c/n$, $c \in \Re^+$, $T(G) \geq \gamma(c)n$, where $\gamma(c)$ depends only on the constant c. Later de la Vega [7] determined the constant $\gamma(c)$ to be essentially $2 \ln c/c$. For the case of $p = p(n) = \frac{c(\ln n)}{n}$ ($c \geq e$ fixed but arbitrary), Palka and Rucinski [17] established that for any fixed $\epsilon > 0$, $\left(\frac{1}{c} - \epsilon\right) \frac{n(\ln \ln n)}{\ln n} \leq T(G) \leq \left(\frac{2}{c} + \epsilon\right) \frac{n(\ln \ln n)}{\ln n}$ a.a.s.

Given a graph G, let $h(G)$ denote the size of a largest induced cycle (in short *hole*) in G. Large holes in random graphs were first studied by Frieze and Jackson in [11], in which they showed that the random graph $\mathcal{G}(n, p)$, $p = c/n$, a.a.s. has a hole of size $n \cdot \Omega(c^{-3})$. They also proved that for any fixed $d \geq 3$, the random regular graph $G(n, d)$ a.a.s. has a hole of size $n \cdot \Omega(d^{-2})$. Later Suen [19] improved the lower bound for $h(G)$, for $G \in \mathcal{G}(n, c/n)$, for any fixed $c > 1$ and $\epsilon > 0$, to at least $(h(c) - \epsilon)n$ where $h(c)$ is defined below and approaches $(\ln c)/c$ for large enough c. Luczak [15] established that, for every $\epsilon > 0$ there is a $d = d(\epsilon)$ such that for every $p = p(n)$ satisfying $p \geq d/n$ and $p = o(1)$, we have $(1 - \epsilon)p^{-1}(\log np) \leq h(G) \leq 2p^{-1}(\log np + 2)$ a.a.s. Thus, $h(G)$ is determined almost surely within a multiplicative factor of nearly two for this range of p. For dense random graphs (corresponding to $p = p(n)$

Copyright © 2018 by SIAM
Unauthorized reproduction of this article is prohibited

being a constant less than 1), Rucinski [18] established that $h(G) = 2(\log_q n)[1 + o(1)]$ where $q = (1-p)^{-1}$. While this determines (for dense random graphs), the value of $h(G)$ upto asymptotically negligible additive terms, it does not lead to a tight concentration of $h(G)$.

The determination of the size $mip(G)$ of a largest induced *path* in $\mathcal{G}(n,p)$, was first studied by Frieze and Jackson in [11], in the course of their work on holes. Since a hole is just an induced path with an edge joining the endpoints, the existence of a large hole in $\mathcal{G}(n,p)$ is very likely if a large induced path is shown to exist a.a.s., and this was the idea used by Frieze and Jackson. On the other hand, large induced paths in $\mathcal{G}(n,p)$ are interesting in their own right, and Suen [19] studied this problem, showing that when $p = c/n$, for any fixed $c > 1$ and any $\epsilon > 0$, a.a.s. the random graph $\mathcal{G}(n,p)$ has an induced path of size at least $(1-\epsilon)h(c)n$, where

$$h(c) \quad = \quad c^{-1}\int_1^c \frac{(1-y(\zeta))}{\zeta}d\zeta$$

where $y(\zeta)$ is the smallest positive root of $y = e^{\zeta(y-1)}$. As $c \to \infty$, $\frac{c\cdot h(c)}{\ln c} \to 1$ and hence for every fixed $\epsilon > 0$, a.a.s., $mip(G) \geq (1-\epsilon)(n\ln c)/c$ for every sufficiently large c. Almost all of the above mentioned previous results are for sparse random graphs (that is, for $p = c/n$ for constant $c > 1$). No previous work on tight concentration of these invariants is known to have been carried out. Also, the dense case (corresponding to higher values of $p = p(n)$) has not been looked at in such great detail. In this paper, we look at this case and obtain very tight concentration results.

1.2 Improved results on sizes of induced paths, trees and holes Throughout the paper, we *assume* that $G \in \mathcal{G}(n,p)$ for $p = p(n) \leq 1-\epsilon$ where $\epsilon > 0$ is an arbitrary but fixed constant. We study induced subgraphs (paths, holes and trees) for dense random graphs. Our first result is a 2-point concentration for $mip(G)$, for $G \in \mathcal{G}(n,p)$ and $p \geq n^{-1/2}(\ln n)^2$:

DEFINITION 1.1. *Let $b^* = b^*(n,p)$ be the largest positive integer b such that $(n)_b p^{b-1}(1-p)^{\binom{b-1}{2}} \geq np/(\ln\ln n)$.*

In other words, b^* is the largest positive integer b such that the expected number of induced paths of length b (shortly, b-paths) is at least $\frac{np}{2(\ln\ln n)}$. A tight determination of b^* within at most 3 consecutive values is presented in Claim 2.1.

THEOREM 1.2. *If $p \geq n^{-1/2}(\ln n)^2$, then $mip(G)$ lies in the set $\{b^*, b^*+1\}$ a.a.s.*

Since an induced path is also an induced tree, we get a significant additive improvement over Erdős and Palka's long-standing (25 year old) lower bound for $T(G)$ (which was for fixed $0 < p < 1$) as a corollary.

COROLLARY 1.1. *If $p \geq n^{-1/2}(\ln n)^2$, then there exists an induced tree of size b^* in G a.a.s.*

The above corollary, combined with a more careful analysis of Erdős and Palka's first moment bound gives

COROLLARY 1.2. *For $p \geq n^{-1/2}(\ln n)^2$, we have*

$$b^* \quad \leq T(G) \leq \quad 2\log_q np + O(1/\ln q)$$

a.a.s. As a result, $T(G) = 2(\log_q np) + O(1/\ln q)$ a.a.s.

Hence it is seen that the asymptotic upper bound on the range of concentration of $T(G)$ is significantly improved, from the previously known bound of $O(\ln n/\ln q)$, to $O(1/\ln q) = O(1/p)$.

As for induced paths, we also obtain a similar 2-point concentration for the size of a largest hole:

DEFINITION 1.2. *Let $h^* = h^*(n,p)$ be the maximum integer b such that $(n)_b p^b(1-p)^{\binom{b-1}{2}-1}/2b \geq np/(\ln\ln n)$.*

In other words, h^* is the largest positive integer b such that the expected number of induced cycles of length b (shortly, b-cycles) is at least $\frac{np}{(\ln\ln n)}$. A tight determination of h^* within at most 3 consecutive values is presented in Claim 3.1.

THEOREM 1.3. *Let $G \in \mathcal{G}(n,p)$. Then, provided $p \geq n^{-1/2}(\ln n)^2$, a.a.s., $h(G) \in \{h^*, h^*+1\}$.*

The proofs of the above results involve just the well-known first and second moment methods. However, the proof involves a number of cases and a careful analysis of the these cases and also the asymptotic behavior of various functions to obtain the desired two-point concentration result that we present.

The second moment method seems inevitable since more powerful concentration tools like Azuma's inequality or Talagrand inequality are not applicable to each of the three invariants we study. Each of these invariants is not monotone and can significantly vary with the addition or deletion of a single edge. Indeed, if the edges incident on just a single vertex in a graph G are changed to get the graph G', all we can say is that $|mip(G) - mip(G')| \leq mip(G)/2$, since there could be a single largest induced path in G which gets split into two equal sized induced paths, while all other induced paths of G are of size at most $mip(G)/2$, and continue to remain so in G'.

Thus these results provide evidence of scenarios where more powerful and stronger concentration tools

Downloaded 11/20/19 to 58.247.249.188. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © 2018 by SIAM
Unauthorized reproduction of this article is prohibited

like those of Azuma or Talagrand fail but second moment method helps us in obtaining tight concentration results.

Outline : The proof of Theorem 1.2 is presented in Section 2. The proof of Theorem 1.3 is presented in Section 3.

2 Induced paths : Proof of Theorem 1.2

For $b \geq 1$, let $X(b) = X(n, b, p)$ be the number of induced paths on b vertices in $G \in \mathcal{G}(n, p)$. The following claim determines b^* up to constant additive factors.

CLAIM 2.1. *For $p \geq n^{-1/2}(\ln n)^2$, $\lfloor 2(\log_q np) + 2 \rfloor \leq b^* \leq \lceil 2(\log_q np) + 3 \rceil$.*

The proof of this claim can be found in the Appendix.

2.1 Proof of $mip(G) \leq b^* + 1$:

Using Claim 2.1 and also the lower bound on p, we notice that $b^* = O(n^{1/2}/(\ln n))$.

The probability that a given ordered (orderings considered up to reversal) set A of b vertices induces a path is given by

$$\mathbf{Pr}[G[A] \text{ is an induced path }] = p^{b-1}(1-p)^{\binom{b-1}{2}}.$$

By the phrase "up to reversal", we mean that an ordering σ and its reversal σ^R are considered the same. Hence the expected number of induced b-paths ($b \geq 2$) is

$$\begin{aligned} \mathbf{E}[X(b)] &= \binom{n}{b}\frac{b!}{2}p^{b-1}(1-p)^{\binom{b-1}{2}} \\ &= \frac{(n)_b}{2}p^{b-1}(1-p)^{\binom{b-1}{2}}. \qquad (2.1) \end{aligned}$$

Hence, for $b = b^* \pm O(1)$,

$$\begin{aligned} \frac{\mathbf{E}[X(b+1)]}{\mathbf{E}[X(b)]} &= (n-b)p(1-p)^{b-1} \\ &= \frac{np \cdot [1-o(1)] \cdot (1-p)^{\pm O(1)}}{(np)^2} \\ &= \Theta\left(\frac{1}{np}\right). \end{aligned}$$

It follows from the definition of b^* that

$$\begin{aligned} \mathbf{E}[X(b^*+2)] &= O(\mathbf{E}[X(b^*+1)](np)^{-1}) \\ &= O((\ln\ln n)^{-1}) = o(1). \qquad (2.2) \end{aligned}$$

This establishes that $mip(G) \leq b^* + 1$ a.a.s. ∎

In fact, we can prove (see the Appendix) the following upper bound on $mip(G)$ which holds for any value of $p = p(n)$.

CLAIM 2.2. *For any $p = p(n) \geq \frac{2}{n}$,*

$$mip(G) \leq \lceil 2(\log_q np) + 3 \rceil$$

a.a.s.

2.2 Proof of $mip(G) \geq b^*$:

Let $b = b^*$. Consider the variance and the expectation of the random variable $X = X(b)$ defined in the previous sub-subsection. Let X_i be the indicator variable for the i-th ordered b-set A_i to induce a path, for a fixed enumeration of ordered b-sets. As stated before, the orderings are considered up to reversal. Therefore, $X = \sum_i X_i$. Applying Chebyshev's Inequality and using standard simplifications (see e.g. [2], Chapter 4), it follows that

$$Var(X) \leq \mathbf{E}[X] + M \cdot \mathbf{E}[X]^2 \qquad (2.3)$$

$$\mathbf{Pr}(X = 0) \leq \frac{Var(X)}{\mathbf{E}[X]^2} \leq \frac{1}{\mathbf{E}[X]} + M^* \qquad (2.4)$$

where

$$M^* := \sum_i \sum_{j:2 \leq |A_i \cap A_j| \leq b} \frac{\mathbf{E}[X_i X_j] - \mathbf{E}[X_i]\mathcal{E}[X_j]}{\mathbf{E}[X]^2}.$$

$A_i \cap A_j$ denotes intersection of unordered sets represented by A_i and A_j.

For those pairs (i, j) such that A_i and A_j are the same (as unordered sets), we have $X_i X_j = 0$ always. Also, since the random graph model $\mathcal{G}(n, p)$ is homogenous, the above expression for M^* simplifies to $M^* \leq M$ where :

$$M := \sum_{l=2}^{b-1} \sum_{j:|A_1 \cap A_j| = l} \frac{\mathbf{E}[X_j | X_1 = 1] - \mathbf{E}[X_j]}{\mathbf{E}[X]}. \qquad (2.5)$$

By our choice of $b = b^*$, it follows that $\mathbf{E}[X] \to \infty$ and hence it suffices to prove that $M = o(1)$ in order to deduce that $mip(G) \geq b^*$ with probability $1 - o(1)$. That $M = o(1)$ follows from Claim 2.5 (established below) as follows : Using the previously observed fact $b^* = O(n^{1/2}/(\ln n))$, we infer that $M = O(b^4 p/n^2) = o(1)$. This completes the proof of Theorem 1.2.

It remains to show that $M = o(1)$ and the following bound on M will be useful in that direction and it is established in Subsection 2.1.

For the remainder of this section, we use $\alpha = \alpha(n)$ to denote any fixed and "sufficiently slowly" growing function such that α grows slower than any growing function hidden by the notation $\omega(1)$ that we will implicitly use in the proof arguments. We will use α in place of any $\omega(1)$ growth that arises in the proof arguments.

Downloaded 11/20/19 to 58.247.249.188. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © 2018 by SIAM
Unauthorized reproduction of this article is prohibited

LEMMA 2.1.

$$M \leq \sum_{l=2}^{b-1} F_l$$

where

$$\begin{aligned}
F_l &= \frac{(n-b)_{b-l}}{(n)_b} \cdot p^{-l} \cdot (1-p)^{l-\binom{l}{2}} \\
&\cdot \sum_{k=1}^{\min\{l,b-l+1\}} f(k) \; ; \\
f(k) &= \left(\frac{(b-l+1)^{2k}}{(k!)^2}\right) \cdot \left(\frac{l^k}{k!}\right) \cdot 2^k \cdot k! \cdot \\
&\left[\left(\frac{p}{1-p}\right)^k - p^l(1-p)^{\binom{l}{2}-l}\right].
\end{aligned}$$

The proof of the above lemma forms the content of Section 2.3. We continue with the proof of $mip(G) \geq b^*$. Split the right side of (2.6) as

$$\begin{aligned}
M &\leq M_1 + M_2 \\
\text{where } M_1 &:= \sum_{2 \leq l \leq (b+1)/2} F_l \\
M_2 &:= \sum_{(b+1)/2 < l \leq b-1} F_l.
\end{aligned}$$

CLAIM 2.3. *For every $l \geq 2$, $f(k)$ is maximized at $k = k_{max} = \min\{l, b-l+1\}$. Further $\sum_k f(k)$ is $(1+o(1))f(k_{max})$.*

Proof. [of Claim 2.3 :] We prove the claim for all large values of n. For $k < k_{max}$, we have the ratio

$$\begin{aligned}
\frac{f(k+1)}{f(k)} &= \frac{2(b-l+1)^2 l}{(k+1)^2} \cdot \\
&\left(\frac{p^{k+1}(1-p)^{-k-1} - p^l(1-p)^{\binom{l}{2}-l}}{p^k(1-p)^{-k} - p^l(1-p)^{\binom{l}{2}-l}}\right) \\
&= \frac{2(b-l+1)^2 lp}{(k+1)^2(1-p)} \cdot \\
&\left(\frac{1-p^{l-k-1}(1-p)^{\binom{l}{2}-l+k+1}}{1-p^{l-k}(1-p)^{\binom{l}{2}-l+k}}\right) \\
&= \frac{2(b-l+1)^2 lp}{(k+1)^2(1-p)} \cdot S \text{ where} \\
S &= \left(\frac{1-a\left(\frac{1-p}{p}\right)}{1-a}\right) \quad \text{and} \\
a &= \left(\frac{p}{1-p}\right)^{l-k} \cdot (1-p)^{\binom{l}{2}}.
\end{aligned}$$

We use Claim 2.4 stated and proved below. If $p \geq 1/2$, then the ratio is at least $(b+1)p/(1-p) \geq \alpha$ for every large n. Suppose that $p \leq 1/2$. When $l \leq L := \sqrt{\frac{2}{100(\ln q)}}$, $k_{max} = l$, hence, for all $k < k_{max}$, the ratio is at least $\frac{2(b-l+1)^2 l^3 p^2}{8(k+1)^2(1-p)} \geq (b+1)p \geq \alpha$ for all large n, for the assumed range of p. For l with $L < l \leq (b+1)/2$, the ratio is at least $\frac{(b+1)pS}{(1-p)} \geq \alpha$ again as $n \to \infty$. Therefore $f(k)$ achieves its maximum f_{max} at $k = l$ when $l \leq (b+1)/2$. When $l > (b+1)/2$, we have $k_{max} = b-l+1$ and hence, for all $k < k_{max}$, the ratio is at least $\frac{(b+1)pS}{1-p} \geq \alpha$ and hence $f(k)$ achieves its maximum at $k = b-l+1$. It follows that $\sum_k f(k)$ is upper bounded by the sum of a finite and increasing geometric series with a common ratio $\alpha = \omega(1)$. Hence $\sum_k f(k) = [1+o(1)]f(k_{max})$. ∎

CLAIM 2.4. (i) *$S \geq 1$ if $p \geq 1/2$.*
(ii) *When $p \leq 1/2$, then $S \geq \frac{l^2(\ln q)}{8}$ if $l \leq L := \sqrt{\frac{2}{100(\ln q)}}$; $S \geq 1 - e^{-1/200}$ if $l > L$.*

Proof. [of Claim 2.4 :] We have $a < 1$ always. When $p \geq 1/2$, we have $(1-p)/p \leq 1$ and hence $S \geq 1$. This proves (i). To prove (ii), write $l = \beta\sqrt{\frac{2}{\ln q}}$ and let $x = (1-p)^{\binom{l}{2}}$. Since $k < k_{max} \leq l$, we have $\frac{a(1-p)}{p} \leq x$. Hence, for each l, we have $S \geq 1-x$. Now $\binom{l}{2} \geq \frac{l^2}{4}$ for $l \geq 2$ and hence we have $x \leq e^{-\beta^2/2}$.

When $l \leq L$, we have $\beta \leq 1/10$ and hence $x \leq 1 - \beta^2/2 + \frac{\beta^2}{4} = 1 - \frac{\beta^2}{4}$ and $S \geq \beta^2/4 \geq \frac{l^2(\ln q)}{8}$.

For $l > L$, we have $\beta \geq 1/10$ and $x \leq e^{-1/200}$ and hence $S \geq 1 - e^{-1/200}$. ∎

CLAIM 2.5. $M \leq M_1 + M_2 = O(b^4 p/n^2)$.

Proof. We consider two cases.
Case 1: $l \leq (b+1)/2$. By Claim 2.3,

$$\begin{aligned}
\sum_{k=1}^{l} f(k) &= (1+o(1))f(l) \\
&\leq \left[2e^2 \frac{(b-l+1)^2}{l} \cdot \frac{p}{1-p}\right]^l \\
&\cdot \left[1-(1-p)^{\binom{l}{2}}\right] \cdot [1+o(1)].
\end{aligned}$$

Therefore, $F_l \leq G_l$ where

$$\begin{aligned}
G_l &:= \frac{(1+o(1)) \cdot (n-b)_{b-l}}{(1-p)^{\binom{l}{2}} \cdot (n)_b} \cdot \\
&\left(\frac{2e^2(b-l+1)^2}{l}\right)^l \cdot \left[1-(1-p)^{\binom{l}{2}}\right].
\end{aligned}$$

Downloaded 11/20/19 to 58.247.249.188. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © 2018 by SIAM
Unauthorized reproduction of this article is prohibited

By definition of G_l,

$$G_2 = (1+o(1)).\frac{p(n-b)_{b-2}}{(1-p)(n)_b} \cdot \left(\frac{2e^2(b-1)^2}{2}\right)^2 = O\left(\frac{b^4 p}{n^2}\right).$$

Therefore, the ratio G_l/G_2 is given by

$$\begin{aligned}\frac{G_l}{G_2} &= [1+o(1)] \cdot \frac{(n-b)_{b-l}}{(n-b)_{b-2}} \cdot (1-p)^{1-\binom{l}{2}} \cdot \\ &\quad \frac{4(2e^2(b-l+1)^2)^l}{l^l\,(2e^2(b-1)^2)^2} \cdot \frac{[1-(1-p)^{\binom{l}{2}}]}{p} \\ &= \frac{O(1)}{n^{l-2}} \cdot (1-p)^{1-\binom{l}{2}} \cdot \frac{(2e^2(b-l+1)^2)^{l-2}}{l^l} \cdot \\ &\quad \frac{[1-(1-p)^{\binom{l}{2}}]}{p} \\ &= O(1) \cdot \left(\frac{2e^2(b-l+1)^2(1-p)^{-(l+1)/2}}{nl}\right)^{l-2} \cdot \\ &\quad \frac{[1-(1-p)^{\binom{l}{2}}]}{pl^2}.\end{aligned}$$

Consider the base T_1 of the second term of the product, namely, $T_1 := \frac{2e^2(b-l+1)^2(1-p)^{-(l+1)/2}}{nl}$. When $l \le 2(\ln q)^{-1} - 1$, we have $(1-p)^{-(l+1)/2} \le e$. Therefore $T_1 \le \frac{2e^3 \cdot b^2}{n} \le \alpha^{-1}$ for every large n. When l is such that $\lfloor (\ln q)^{-1} \rfloor \le l \le (b+1)/2$, we have $(1-p)^{-(l+1)/2} \le O(1) \cdot \sqrt{np}$ and hence $T_1 \le \frac{O(1)\cdot(\ln np)^2}{\sqrt{np}} \le \alpha^{-1}$, since $b = O(p^{-1}(\ln np))$.

Suppose $p \le 1/2$. Then, $\frac{\ln q}{p} \le 3/2$. Write $l = \beta\sqrt{\frac{2}{\ln q}}$ and $x = (1-p)^{\binom{l}{2}}$. We have $x \ge (1-p)^{l^2/2} = e^{-\beta^2}$. Hence for $\beta \le 1$, we have $1 - x \le \beta^2 = \frac{l^2(\ln q)}{2}$. Hence, we obtain (for $\beta \le 1$) that $T_2 := \frac{1-x}{pl^2} \le \frac{3}{4}$. For $\beta \ge 1$, we have $pl^2 \ge \frac{2p}{\ln q} \ge \frac{4}{3}$. Hence, $T_2 = \frac{1-x}{pl^2} \le \frac{3}{4} < 1$. As a result, $\frac{G_l}{G_2} = \Theta(1) \cdot T_1^{l-2} \cdot T_2 \le (1/\alpha)^{l-2}$ for every $l \ge 3$.

Suppose $p \ge 1/2$. Then, $T_2 \le 2/l^2 < 1$. Hence, $\frac{G_L}{G_2} = \Theta(1) \cdot T_1^{l_1} \cdot T_2 \le (1/\alpha)^{l-2}$ for every $l \ge 3$.

Therefore, the sum $\sum_{l=2}^{l \le (b+1)/2} G_l$ is upper bounded by the sum of an infinite geometric progression, whose first term is G_2 and common ratio is $1/\alpha$. Hence we get,

$$\begin{aligned} M_1 &= \sum_{l=2}^{l \le (b+1)/2} F_l \\ &\le \sum_{l=2}^{l \le (b+1)/2} G_l \\ &\le [1+o(1)] \cdot G_2 \\ &= O\left(\frac{b^4 p}{n^2}\right).\end{aligned}$$

Case 2: $l > (b+1)/2$. Using Stirling's asymptotic estimate of factorials,

$$\begin{aligned}\sum_{k=1}^{b-l+1} f(k) &= [1+o(1)] \cdot f(b-l+1) \\ &\le \left(\frac{p}{1-p} \cdot (2e^2 l)\right)^{b-l+1}.\end{aligned}$$

Therefore, $F_l \le G_l$ where

$$G_l := \frac{(n-b)_{b-l} \cdot (2e^2 l)^{b-l+1}}{(n)_b} \cdot \frac{p^{b-2l+1}}{(1-p)^{b-2l+\binom{l}{2}+1}}.$$

Using the definition of G_l and also that of $b = b^*$,

$$\begin{aligned} G_{b-1} &= \frac{n-b}{(n)_b} \cdot (2e^2(b-1))^2 \cdot \left(\frac{p}{1-p}\right)^{-b+3} \cdot \frac{1}{(1-p)^{\binom{b-1}{2}}} \\ &\le \frac{np^2(4e^4b^2)(1-p)^{b-3}}{(n)_b p^{b-1}(1-p)^{\binom{b-1}{2}}} \\ &= \frac{np^2(1-p)^{b-3}(4e^4b^2)}{2\mathbf{E}[X(b)]} \\ &= O\left(\frac{np^2(\ln\ln n)b^2}{(np)^3}\right) \\ &= O\left(\frac{b^2(\ln\ln n)}{n^2 p}\right) = o\left(\frac{b^4 p}{n^2}\right).\end{aligned}$$

Now, the ratio G_l/G_{b-1} is given by

$$\begin{aligned}\frac{G_l}{G_{b-1}} &= (n-b-1)_{b-l-1} \cdot (2e^2 l)^{b-l-1} \cdot \left(\frac{p}{1-p}\right)^{2b-2l-2} \\ &\quad \cdot (1-p)^{\binom{b-1}{2}-\binom{l}{2}} \cdot \left(\frac{l}{b-1}\right)^2 \\ &\le \left(\frac{np^2(2e^2 l)}{(1-p)^2}\right)^{b-l-1} \cdot (1-p)^{\binom{b-1}{2}-\binom{l}{2}} \\ &= \left(\frac{np^2(2e^2 l)(1-p)^{(b+l)/2}}{(1-p)^3}\right)^{b-l-1} \\ &\le \left(\frac{np^2 l \cdot O(1)}{(np)^{3/2}}\right)^{b-l-1} \\ &\le \left(\frac{O(1) \cdot l\sqrt{p}}{\sqrt{n}}\right)^{b-l-1} \quad \text{since } \frac{b+l}{2} \ge \frac{3b}{4} \text{ for } l > \frac{b+1}{2} \\ &\le \left(\frac{O(1)\cdot(\ln np)}{\sqrt{np}}\right)^{b-l-1} \\ &\le \left(\frac{1}{\alpha}\right)^{b-l-1} \quad \text{for } l < b-1.\end{aligned}$$

Downloaded 11/20/19 to 58.247.249.188. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © 2018 by SIAM
Unauthorized reproduction of this article is prohibited

Therefore,

$$\begin{aligned} M_2 &\leq \sum_{l>(b+1)/2} G_l \\ &= [1+o(1)] \cdot G_{b-1} \; = \; o\left(\frac{b^4 p}{n^2}\right). \end{aligned}$$

Hence, $M \;\leq\; M_1 + M_2 \;=\; O\left(\frac{b^4 p}{n^2}\right)$. ∎

2.3 Proof of Lemma 2.1

Proof. Consider any $j \neq 2$ such that $2 \leq |A_1 \cap A_j| \leq b-1$. Let the number of vertices in the intersection be l.

CLAIM 2.6. *If both $G[A_1]$ and $G[A_j]$ are induced paths, then the induced subgraph $G[A_1 \cap A_j]$ must be a union of vertex disjoint path segments. The order of occurrence, as well as the alignment (i.e. which end of a segment comes first) of these segments in $G[A_1]$ and $G[A_j]$ however, can differ.*

Proof. Notice that $G[A_1 \cap A_j]$ is an induced subgraph of $G[A_1]$ as well as $G[A_j]$. Since both $G[A_1]$ and $G[A_j]$ are paths of length b, the only possible induced subgraphs are unions of vertex disjoint path segments. It is easy to see that there is no constraint on the relative ordering and alignment of these segments in $G[A_j]$. ∎

We next make an observation on the conditional probability $\mathbf{Pr}[X_j = 1 | X_1 = 1]$ when there are l vertices in the intersection $A_1 \cap A_j$. A pair (A_1, A_j) of potential induced paths is said to be a compatible pair if, with positive probability, both of them can be induced paths. It is easy to see that a pair is a compatible pair if and only if there does not exist a pair $u, v \in A_1 \cap A_j$ of distinct and shared vertices such that u and v are consecutive in one potential path and are not consecutive in the other potential path. If a pair (A_1, A_j) is not compatible, then clearly $\mathbf{Pr}[X_j = 1 | X_1 = 1] = 0$. Hence, for the remainder of this section, we focus only on compatible pairs (A_1, A_j) of potential paths. Also, whenever we use the phrase "A_j shares l vertices and k mutually disconnected segments with A_1", we mean the following : $|A_1 \cap A_j| = l$ and $G[A_1 \cap A_j]$ is a forest of k vertex disjoint paths whenever each of $G[A_1]$ and $G[A_j]$ is a path.

CLAIM 2.7. *Let A_j be any ordered set compatible with A_1. Suppose A_j shares l vertices k mutually disconnected segments with A_1. Then, the conditional probability is exactly*

$$\mathbf{Pr}[X_j = 1 | X_1 = 1] = p^{b-1-(l-k)} (1-p)^{\binom{b-1}{2} - \binom{l}{2} + l - k}.$$

Proof. Since both of A_1 and A_j can be induced paths with positive probability, it follows from Claim 2.6 that $G[A_1 \cap A_j]$ is a forest of paths whenever each of A_1 and A_j induces a path in G. The conditional probability is exactly the expression in the right side of the above inequality. The expression follows by counting the number of edges and non-edges lying in the intersection $A_1 \cap A_j$, when $|A_1 \cap A_j| = l$ and $A_1 \cap A_j$ induces k contiguous segments on A_1. ∎

Given the ordered b-set A_1, we use $S(l,k)$ to denote the number of ways of choosing an ordered b-set A_j compatible with A_1, such that A_j shares l vertices and k mutually disconnected segments with A_1. Then from Claim 2.7 and Equations (2.1) and (2.5), it follows that

$$M \;\leq\; \sum_{l=2}^{b-1} \sum_{k=1}^{l} \left(\frac{S(l,k)}{(n)_b} \cdot \left[p^{-(l-k)} (1-p)^{-\binom{l}{2} + l - k} - 1 \right] \right).$$

We next define three sets of tuples which will help determine the value of $S(l,k)$:

DEFINITION 2.1. *Let the set $\mathcal{T}_1$ denote all tuples $(a_0, b_1, a_1, b_2, \ldots, b_k, a_k)$, such that*

(i) $\sum_{i=1}^{k} b_i = l$, $\forall i \,:\, b_i \geq 1$, $b_i \in \mathcal{Z}^+$.

(ii) $\sum_{i=0}^{k} a_i = b - l$, $a_i \in \mathcal{Z}^+$, $a_0, a_k \geq 0$, $\forall i \neq 0, k \,:\, a_i \geq 1$.

DEFINITION 2.2. *Given a tuple $\tau \in \mathcal{T}_1$, let $\mathcal{T}_2(\tau)$ denote the set of all tuples $(\pi, c_1, \ldots, c_k)$, where π is an ordering of $\{1, \ldots, k\}$, and for $i = 1, \ldots, k$,*

$$c_i \in \begin{cases} \{0\} & \text{if } b_i = 1 \\ \{0,1\} & \text{otherwise} \end{cases}$$

DEFINITION 2.3. *Let $\mathcal{T}_{12}$ denote the set of all ordered pairs (τ_1, τ_2) where $\tau_1 \in \mathcal{T}_1$ and $\tau_2 \in \mathcal{T}_2(\tau_1)$.*

DEFINITION 2.4. *Let $\mathcal{T}_3$ denote the set of tuples $(d_0, \ldots, d_k)$ such that $\sum_{i=0}^{k} d_i = b - l$, $d_i \in \mathcal{Z}^+$, $d_0, d_k \geq 0$ and $d_i \geq 1$ for all other i.*

Finally, let $\mathcal{T} = \mathcal{T}_{12} \times \mathcal{T}_3$. To get an upper bound on $S(l,k)$, just observe the following:

PROPOSITION 2.1. *Given an ordered set B (disjoint with A_1) of $b-l$ vertices, each element of $\mathcal{T}$ specifies (in a bijective fashion) a unique ordered set A_j compatible with A_1 and having b vertices, such that $A_j \setminus A_1 = B$, $|A_1 \cap A_j| = l$ and A_j shares k mutually disconnected segments with A_1.*

Proof. Given an element of $\mathcal{T} = \mathcal{T}_{12} \times \mathcal{T}_3$, say $u = ((a_0, b_1, \ldots, a_k), (\pi, c_1, \ldots, c_k), (d_0, \ldots, d_k))$, first,

Downloaded 11/20/19 to 58.247.249.188. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © 2018 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 11/20/19 to 58.247.249.188. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

divide the ordered set A_1 into an ordered partition into paths having $a_0, b_1, a_1, \ldots, a_k$ vertices respectively. The paths corresponding to integers b_i are chosen to lie in the intersection $A_1 \cap A_j$. Let these intersection segments be called $P_1, \ldots, P_k$. By the definition of b_i's, the total number of vertices in the intersection is thus l. Order the P_i's using the ordering π. For each P_i, call the end-point occuring earlier in A_1 as the "head" if $c_i = 0$, otherwise the head is the end-point occuring later in A_1. Next, divide B into ordered parts $D_0, \ldots, D_k$, having sizes $d_0, \ldots, d_k$. Now, insert P_1 between D_0 and D_1, P_2 between D_1 and D_2, and so on, while ensuring that the head of each P_i succeeds the last vertex of D_{i-1}. By the definitions of P_is and D_is, we get a unique ordered b-set A_j such that $A_1 \cap A_j$ has l vertices, divided into k separated path segments. ∎

The value of $S(l,k)$ can now be ascertained from the following claim:

Claim 2.8. $S(l,k) = (n-b)_{b-l} \cdot |\mathcal{T}|$.

Proof. Choose an ordered set B of size $b-l$, in $(n-b)_{b-l}$ ways, and apply Proposition 2.1. ∎

Hence, to upper-bound $S(l,k)$, we need to upper-bound the sizes of the sets $\mathcal{T}_i$, $i = 1,2,3$. Clearly, $|\mathcal{T}_2(\tau)|$ is at most $k!2^k$ for each $\tau \in \mathcal{T}_1$. To estimate the sizes of $\mathcal{T}_i$, $i = 1,3$, we recall a basic combinatorial fact which can proved using the balls and bar model of counting.

Proposition 2.2. (see e.g. [20], Chapter 13) *The number of integral solutions of $\sum_i x_i = a$, with integral constraints $x_i \geq c_i$; $a, c_i \in \mathcal{Z}$; $1 \leq i \leq r$, is* $\binom{a-(\sum_i c_i)+r-1}{r-1}$.

Hence, the set $\mathcal{T}_3$ consisting of all integral solutions of $\sum_i d_i = b-l$, such that $d_0, d_k \geq 0$ and $d_i \geq 1$, for $i = 1, \ldots, k$, has cardinality $\binom{b-l+1}{k}$. The size of $\mathcal{T}_1$ can be determined by counting all solutions, in non-negative integers of the following pairs of equations

$$\sum_{(a_i):a_0,a_k \geq 0, a_1,\ldots a_{k-1} \geq 1} a_i = b-l \tag{2.6}$$

$$\sum_{(b_j):b_j \geq 1} b_j = l \tag{2.7}$$

The number of solutions satisfying both of the above equations, by Proposition, 2.2 comes to $\binom{b-l+1}{k}\binom{l-1}{k-1}$. Therefore, $|\mathcal{T}_1| = \binom{b-l+1}{k}.\binom{l-1}{k-1}$. From the above argument and Claim 2.8 we get

$$S(l,k) \quad \leq \quad (n-b)_{b-l}\binom{b-l+1}{k}^2\binom{l-1}{k-1}2^k k!$$

Plugging the above bound on $S(l,k)$ in Equation (2.6) proves the Lemma. ∎

3 Holes – Proof of Theorem 1.3

Redefine, for this section, $X := X(n,b,p)$ to be the number of holes of size b in $G \in G(n,p)$. The following claim determines h^* upto constant additive factors.

Claim 3.1. *(i) For $p \geq n^{-1/2}(\ln n)^2$, $\lfloor 2\log_q np + 1 \rfloor \leq h^* \leq \lceil 2\log_q np + 2 \rceil$.*
(ii) For any $p = p(n) \geq \frac{2}{n}$, $h(G) \leq \lceil 2(\log_q np) + 2 \rceil$ a.a.s.

The proof of this claim can be found in the Appendix.

Proof. [Proof of Theorem 1.3] The proof of this theorem is along similar lines as the proof of Theorem 1.2. Consider the ratio $r(b)$ of the expected number of holes of size $b+1$ to the expected number of holes of size b, where $b = h^* \pm O(1)$.

$$\begin{aligned} r(b) &= \frac{\mathbf{E}[X(b+1)]}{\mathbf{E}[X(b)]} = (n-b)p(1-p)^{b-1} \cdot \frac{b}{b+1} \\ &= \Theta\left(np\frac{1}{(np)^2}\right) = \Theta\left(\frac{1}{np}\right). \end{aligned}$$

Hence, it follows from the definition of h^* that

$$\begin{aligned} \mathbf{E}[X(h^*+2)] &= O(\mathbf{E}[X(h^*+1)](np)^{-1}) \\ &= O\left(\frac{1}{\ln\ln n}\right) = o(1). \end{aligned}$$

This establishes that $h(G) \leq h^* + 1$ almost surely.

For the lower bound, set $b = h^*$ and using a fixed but arbitrary enumeration of all cyclically ordered subsets of size b, write $X = X(b) = \sum_{1 \leq i \leq m} X_i$, where $m = (n)_b/(2b)$, note that $E[X] \to \infty$. Here, X_i denotes the indicator random variable for the i-th cyclically ordered set inducing a hole.

We again use Chebyshev's inequality and arguments similar to those used for induced paths. As before, using second moment calculations (with analagous definition of M_h as for paths defined by Equation (2.5)), we can deduce that

$$\begin{aligned} \mathbf{Pr}[h(G) < h^*] &= \mathbf{Pr}[X = 0] \\ &\leq \frac{Var(X)}{\mathbf{E}[X]^2} \leq (\mathbf{E}[X])^{-1} + M_h^* \\ &\leq \frac{\ln\ln n}{np} + O\left(\frac{b^5 p}{n^2}\right) = o(1), \end{aligned}$$

provided we can show that $M_h^* \leq M_h = O\left(\frac{b^5 p}{n^2}\right)$. This follows from the upper bound for M_h established by Lemma 3.1 below. A sketch of its proof is provided in the appendix.

Copyright © 2018 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 11/20/19 to 58.247.249.188. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

LEMMA 3.1.

$$M_h \leq \sum_{l=2}^{b-1} \sum_{k=1}^{\min\{l,b-l\}} t_{l,k} = O\left(\frac{b^5 p}{n^2}\right) \text{ where}$$

$$t_{l,k} := \frac{(2b)(n-b)_{b-l}}{(n)_b} \cdot \binom{b-l+1}{k}^2 \binom{l}{k} 2^k k! \cdot \left(p^{-(l-k)}(1-p)^{-\binom{l}{2}+l-k} - 1\right).$$

Thus, we have shown that

$$\mathbf{Pr}[h(G) < h^*] = o(1).$$
$$\mathbf{Pr}[h(G) \geq h^* + 2] = o(1).$$

Hence, it follows that a.a.s., $h(G) \in \{h^*, h^* + 1\}$. ∎

4 Open Problems

We investigated certain non-monotone functions on graphs, such as $mip(G)$, $h(G)$ and $T(G)$, in the random graph model $\mathcal{G}(n,p)$ and obtained a 2-point concentration of $mip(G)$ and $h(G)$, while improving the lower bound for $T(G)$ for a certain range of p. The concentration results for $mip(G)$ and $h(G)$ motivate us to pose the following open problem for further study.

Question 1 : Given an arbitrary but fixed p : $0 < p < 1$, do there exist 2 consecutive values $b(n,p), b(n,p)+1$, such that a.a.s. $T(G)$ (for $G \in \mathcal{G}(n,p)$) is either $b(n,p)$ or $b(n,p)+1$?.

Question 2 Our results work for moderately-dense to dense random graphs: e.g. $mip(G)$ is 2-point concentrated for $p = p(n)$ satisfying $n^{-1/2}(\ln n)^2 \leq p \leq 1 - \epsilon$. Can we establish similar or slightly weaker concentration of these invariants when the edge probability is smaller ?.

de la Vega [7] posed whether, for $p = \frac{c}{n}$, $c > 1$ is any fixed but sufficiently large constant, $G \in \mathcal{G}(n,p)$ has (a.a.s.) an induced path of size at least $\frac{2n}{c} \cdot (\ln c - \ln\ln c - 1)$. This when combined with an upper bound of $\frac{(2+\epsilon)n(\ln c)}{c}$ established by Erdos and Palka [8] nearly determines (a.a.s.) the size of $mip(G)$ within a multiplicative factor which can be made arbitrarily close to 1 by choosing c suitably large and ϵ suitably small. This problem has remained open for more than two decades. We propose a strengthening of this problem as the following question. We conjecture that this question admits an affirmative answer.

Question 3 : Is the following true ? : For every fixed and sufficiently large $c > 1$ and for $G \in \mathcal{G}(n,p)$ with $p = \frac{c}{n}$, $mip(G) = 2(\log_q np) \cdot [1 + o(1)]$ a.a.s.

References

[1] N. Alon, J.H. Kim, and J. Spencer, Nearly perfect matchings in regular simple hypergraphs, *Israel J. Math.* **100** (1997), 171-187.

[2] N. Alon and J.H. Spencer, *The Probabilistic Method*, Wiley International, 2001.

[3] B. Bollobás, Random Graphs (2nd Edition) Camb. Univ. Press (2001).

[4] B. Bollobás, The chromatic number of random graphs. *Combinatorica* **8(1)**: 49-55 (1988).

[5] B. Bollobás and A. Thomason, The structure of hereditary properties and colourings of random graphs. *Combinatorica* **20(2)**: 173-202 (2000).

[6] W. Fernandez de la Vega, Induced trees in sparse random graphs. *Graphs and Combinatorics*, **2** 1 (1986), pp. 227-231.

[7] W. Fernandez de la Vega, The largest induced tree in a sparse random graph. *Random Struct. Alg.* **9** 1-2, (1996), pp. 93-97.

[8] P. Erdős, Z. Palka, Trees in Random Graphs. *Discrete Mathematics* **46** (1983) pp. 145-150.

[9] P. Erdős, A. Rényi, On random graphs I. *Publicationes Mathematicae* **6** (1959) pp. 290-297.

[10] A. M. Frieze, B. Jackson, Large induced trees in sparse random graphs. *Journal of Combinatorial Theory, Series B* **42** pp. (1987) 181-195.

[11] A. M. Frieze, B. Jackson, Large holes in sparse random graphs. *Combinatorica* **7** pp. (1987) 265-274.

[12] S. Janson, T. Łuczak and A. Ruciński, An exponential bound for the probability of nonexistence of a specified subgraph in a random graph. *Random Graphs '87 (Proceedings, Poznan 1987)*, John Wiley & Sons (1990), pp. 73-87.

[13] L. Kučera and V. Rödl, Large trees in random graphs, *Comment. Math. Univ. Carolina* **28**, 7-14, (1987).

[14] M. Krivelevich, B. Sudakov, V. Vu and N. Wormald, On the probability of indpendent sets in random graphs, *Rand. Struct. Alg.* **22** (2003) 1-14.

[15] T. Łuczak, The size of the largest hole in a random graph, *Discrete Mathematics* **112**, 151-163 (1993).

[16] T. Łuczak and Z. Palka, Maximal induced trees in sparse random graphs, *Discrete Mathematics* **72** (1988), 257-265.

[17] Z. Palka and A. Ruciński, On the order of the largest induced tree in a random graph, *Discrete Applied Mathematics* **15** (1986), 75-83.

[18] A. Ruciński, Induced subgraphs in a random graph, *Annals of Discrete Mathematics* **33** (1987), 275-296.

[19] W. C. Suen, On large induced trees and long induced paths in sparse random graphs, *Journal of Combinatorial Theory, Series B*, **56**(2) (1992) pp. 250-262.

[20] J. H. Van Lint, R. M. Wilson, A Course in Combinatorics, *Cambridge University Press*, (1991).

[21] V. Vu, Concentration of non-Lipschitz functions and applications, *Random Structures and Algorithms*, **20** (3) (2002), pp. 262-316.

Copyright © 2018 by SIAM
Unauthorized reproduction of this article is prohibited

Appendix

4.1 Proof of Claim 2.1

Proof. The upper bound follows from the proof of Claim 2.2 given below. Hence we establish only the lower bound. Suppose $b = \lfloor 2(\log_q np) + 2 \rfloor$. Let $0 \leq \delta < 1$ be such that $b = 2(\log_q np) + 2 - \delta$. Let X denote $X(b)$.

$$(b-2)/2 = \log_q np - (\delta/2)$$

Now, $(n)_b \geq (n-b)^b = n^b(1-b/n)^b$. From the assumed lower bound on p, it follows that $b \leq \frac{3n^{1/2}}{\ln n}$. Hence $(1-b/n)^b = e^{-o(1)} = 1 - o(1)$. Hence for all $p \geq n^{-1/2}(\ln n)^2$, we get

$$\begin{aligned}
2\mathbf{E}[X] &\geq (0.5)n\left(np(1-p)^{(b-2)/2}\right)^{b-1} \\
&= (0.5)n\left(np(1-p)^{\log_q np}(1-p)^{-\delta/2}\right)^{b-1} \\
&= (0.5)n(1-p)^{-\delta(b-1)/2} \\
&\geq (0.5)n \text{ for all large } n.
\end{aligned}$$

This establishes that $b^* \geq \lfloor 2(\log_q np) + 2 \rfloor$. ∎

4.2 Proof of Claim 2.2

Proof. Suppose $b = \lceil 2(\log_q np) + 4 \rceil$. Let X denote $X(b)$. Let $0 \leq \delta < 1$ be such that $b = 2\log_q np + 4 + \delta$.

$$(b-2)/2 = \log_q np + 1 + (\delta/2)$$

Now, $(n)_b = n(n-1)\ldots(n-b+1) \leq n^b e^{-b(b-1)/2n}$. Hence for all $p \geq 2/n$, we get

$$\begin{aligned}
2\mathbf{E}[X] &\leq ne^{-\binom{b}{2}/n}(np(1-p)^{(b-2)/2})^{b-1} \\
&= ne^{-b(b-1)/2n}(np(1-p)^{\log_q np}(1-p)^{1+\delta/2})^{b-1} \\
&= ne^{-\binom{b}{2}/n}(1-p)^{(2+\delta)\log_q np + (3+\delta)(2+\delta)/2} \\
&= e^{-b(b-1)/2n}\left(\frac{n(1-p)^{\Theta(1)}}{(np)^{2+\delta}}\right) \\
&= A \cdot B
\end{aligned}$$

where $A \leq 1$ and $B \leq n$ always. Let $\omega = \omega(n)$ be a sufficiently slowly growing function. For $p \geq w/\sqrt{n}$, we have $B \to 0$. For p such that $p \leq \omega/\sqrt{n}$, we have $A = o(n^{-1})$. Hence, for $p \geq 2/n$, $\mathbf{E}[X] \to 0$. ∎

4.3 Proof of Claim 3.1

Proof. Suppose $p \geq n^{-1/2}(\ln n)^2$. Write $b = 2(\log_q np) + 2 + \delta$ where δ is defined by the value we assign to b. Let X denote $X(b)$. We have, after employing simplifications similar to the ones employed before for inducd paths,

$$\begin{aligned}
\mathbf{E}[X] &= \frac{(n)_b}{2b} \cdot p^b \cdot (1-p)^{\binom{b-1}{2}-1} \\
&= \frac{(np)^{1-\delta} \cdot [1-o(1)] \cdot (1-p)^{\delta(1+\delta)/2}}{2b(1-p)} \\
&\to 0 \quad \text{for} \quad b = \lceil 2(\log_q np) + 3 \rceil. \\
&\geq \frac{(np)^2 \cdot \Theta(1)}{b} \\
&\geq np \quad \text{for} \quad b = \lfloor 2(\log_q np) + 1 \rfloor.
\end{aligned}$$

This establishes Part (i) of the claim. For Part (ii) of the claim, the '=' in the second equation will be replaced by '$\leq$' and also the $[1-o(1)]$ term is ignored, thereby establishing the claim. ∎

4.4 Proof of Lemma 3.1

Proof. **(sketch :)** An induced cycle can be considered to be an induced path with the endpoints joined. The argument for holes therefore, follows along lines very similar to those of paths, with the following differences:
(i) : $A_1 \ldots, A_m$ now denote a fixed but arbitrary enumeration of all cyclically ordered sets of size b. Here, $m = \frac{(n)_b}{2b}$ denotes the number of such sets.
(ii) : We define $S(l,k)$ in an analogous way for holes. To upper bound $S(l,k)$, we note that in choosing the intersection $A_1 \cap A_j$ and the difference set $A_j \backslash A_1$, the parts a_0 and a_k are now considered as one part - say a_0 - since the last and first vertex of A_1 will be joined. Since a_0 must now differentiate between b_1 and b_k, there must be at least one vertex in the segment corresponding to a_0. This changes the number of solutions of equation (2.6) to $\binom{b-l-1}{k-1}$, for both choices of $A_1 \backslash A_j$ and $A_j \backslash A_1$. Also, we only need to consider only the $\frac{k!}{k}$ cyclic orderings of the k path segments forming $A_1 \cap A_j$. For a given splitting $(d_1, \ldots, d_k) : \sum_i d_i = b - l$ and a given linear ordering of $b - l$ vertices making up $A_j \setminus A_1$, we have k ways of placing the k segments of $A_j \setminus A_1$ in a given cyclic ordering of the segments of $A_j \cap A_1$ to get various other A_j's. So, we introduce a multiplicative factor of k here. Hence, $S(l,k)$ (for holes) now gets upper bounded as

$$S(l,k) \quad \leq \quad (n-b)_{(b-l)} \cdot \binom{b-l-1}{k-1}^2 \binom{l-1}{k-1} \cdot 2^k \cdot k!$$

(iii) : Another multiplicative factor of $2b$ is introduced to account for the fact that number of potential holes is $(n)_b/(2b)$.

Thus, the term M_h (corresponding to M for paths) becomes bounded as

$$M_h \quad \leq \quad \sum_{l=2}^{b-1} F_h(l) \text{ where}$$

Downloaded 11/20/19 to 58.247.249.188. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © 2018 by SIAM
Unauthorized reproduction of this article is prohibited

$$
\begin{aligned}
F_h(l) &= \frac{2b\cdot(n-b)_{b-l}}{(n)_b}\cdot p^{-l}\cdot(1-p)^{l-\binom{l}{2}} \\
&\quad \cdot \left(\sum_{k}^{\min\{l,b-l\}} f_h(k)\right).
\end{aligned}
$$

$$
\begin{aligned}
f_h(k) &= \binom{b-l-1}{k-1}^2\binom{l-1}{k-1}\cdot 2^k\cdot k! \\
&\quad \cdot \left[\left(\frac{p}{1-p}\right)^k - p^l(1-p)^{\binom{l}{2}-l}\right] \\
&= \binom{b-l}{k}^2\left(\frac{k}{b-l}\right)^2\binom{l-1}{k-1}\cdot 2^k\cdot k! \\
&\quad \cdot \left[\left(\frac{p}{1-p}\right)^k - p^l(1-p)^{\binom{l}{2}-l}\right] \\
&\le \binom{b-l+1}{k}^2\left(\frac{k}{b-l}\right)^2\binom{l-1}{k-1}\cdot 2^k\cdot k! \\
&\quad \cdot \left[\left(\frac{p}{1-p}\right)^k - p^l(1-p)^{\binom{l}{2}-l}\right] \\
&\le \left(\frac{(b-l+1)^2}{(k!)^2}\right)^k\left(\frac{k}{b-l}\right)^2\binom{l-1}{k-1}\cdot 2^k\cdot k! \\
&\quad \cdot \left[\left(\frac{p}{1-p}\right)^k - p^l(1-p)^{\binom{l}{2}-l}\right] \\
&\le f(k) \text{ using } k \le b-l.
\end{aligned}
$$

To prove the bound in the lemma, we shall use the results of Section 2. Notice that expressions for M, M_h are both very similar. However, note that for $k > b-l$, $\binom{b-l}{k} = 0$. Hence, for purposes of easy comparison with the derivation in Section 2, redefine $f_h(k)$ as follows:

$$
f_h(k) = \begin{cases} 0 \text{ if either } k > b-l \text{ or } k > l. \\ f(k) \text{ otherwise.} \end{cases}
$$

Define

$$
\begin{aligned}
F_h(l) &:= \frac{(2b)(n-b)_{b-l}}{(n)_b}p^{-l}(1-p)^{l-\binom{l}{2}}\sum_{k}^{\min\{l,b-l\}} f(k) \\
&\le (2b)\cdot F_l.
\end{aligned}
$$

Also, defining $M_1(h), M_2(h)$ analogously, we get that for each $i = 1,2$, $M_i(h) \le (2b)\cdot M_i = O\left(\frac{b^5 p}{n^2}\right) = o(1)$ provided $p \ge n^{-1/2}(\ln n)^2$. Thus, $M_h^* \le M_h = o(1)$ thereby completing the proof. ∎

Copyright © 2018 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 11/20/19 to 58.247.249.188. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php