Large induced distance matchings in certain sparse random graphs *

Fang Tian\textsuperscript{1}†  Yu-Qin Sun\textsuperscript{2}  Zi-Long Liu\textsuperscript{3}

\textsuperscript{1}School of Mathematics
Shanghai University of Finance and Economics, Shanghai, 200433, China
tianf@mail.shufe.edu.cn

\textsuperscript{2}School of Mathematics and Physics
Shanghai University of Electric Power, Shanghai, 200090, China
2008000011@shiep.edu.cn

\textsuperscript{3}School of Optical-Electrical and Computer Engineering
University of Shanghai for Science and Technology, Shanghai, 200093, China
liuzl@usst.edu.cn

\textbf{Abstract}  For a fixed integer $k \geq 2$, let $G \in \mathcal{G}(n, p)$ be a simple connected graph on $n \to \infty$ vertices with the expected degree $d = np$ satisfying $d \geq c$ and $d^{k-1} = o(n)$ for some large enough constant $c$. We show that the asymptotical size of any maximal collection of edges $M$ in $G$ such that no two edges in $M$ are within distance $k$, which is called a distance $k$-matching, is between $\left(\frac{k-1}{d^{k-1}}\right) n \log d$ and $\frac{kn \log d}{2d^{k-1}}$. We also design a randomized greedy algorithm to generate one large distance $k$-matching in $G$ with asymptotical size $\frac{kn \log d}{4d^{k-1}}$. Our results partially generalize the results on the size of the largest distance $k$-matchings from the case $k = 2$ or $d = c$ for some large constant $c$.

\textbf{Keywords}  random graphs, induced matchings, distance matchings, expected degree.

\textbf{Mathematics Subject Classification}  05C35, 05C70, 05C80.

\section{Introduction}

Random graphs are classical mathematical models to capture the main behaviors of real-world graphs in massive data sets. Recall that the Erdős-Rényi random graph model $\mathcal{G}(n, p)$ is on

\textsuperscript{*}The work was supported by NSFC.
\textsuperscript{†}Corresponding Author: tianf@mail.shufe.edu.cn(Email Address)
the vertex set $[n] = \{1, \cdots, n\}$ and every edge appears independently with probability $p$. A classic question is to investigate whether $G \in \mathcal{G}(n, p)$ contains a copy of given graph as an induced subgraph with high probability (abbreviated to w.h.p., meaning with probability tending to 1 as $n$ tends to infinity). Some large induced subgraphs to look for in $G \in \mathcal{G}(n, p)$ are independent sets, trees, cycles and matchings.

For any fixed positive integer $k$, a set of edges $M_k$ of a graph $G = (V, E)$ defined on the vertex set $[n]$, with the additional constraint that no two edges are within distance $k$, is called a distance $k$-matching. The distance between two vertices is the number of edges in a shortest path between them, while the distance between two edges is the number of vertices in a shortest path between them. We often refer to a path by the natural sequence of its vertices, writing $P = x_0 x_1 \cdots x_k$ with distinct vertices $x_i$ for $0 \leq i \leq k$ and calling $P$ a path between $x_0$ and $x_k$ of length $k$. The qualifier "distance" is normally omitted in the remainder of this paper. A 1-matching is a matching and a 2-matching is also known as an induced matching or a strong matching $[2, 4, 6]$. A $k$-matching in $G$ is said to be maximal if no other edges can be added to $M_k$ while keeping the property of being a distance $k$ away. Let $e(M_k)$ denote the size of the $k$-matching $M_k$. The $k$-matching number, denoted by $um_k(G)$, is $e(M_k)$ of the largest $k$-matching $M_k$ in $G$.

Little is known about the random variable $um_k(G)$ when $G \in \mathcal{G}(n, p)$ and $k \geq 2$. Maftouhi and Gordon $[6]$ firstly considered $um_2(G)$ for $G \in \mathcal{G}(n, p)$ when $0 < p < 1$ is a constant. Czygrinow and Nagle $[3]$ generalized this result. They showed that w.h.p. $um_2(G) \leq \frac{n \log d}{d}$ when the expected degree $d = np = \Omega(1)$ and $d = o(n)$, further $um_2(G) \leq \log_h n$ with $h = \frac{1}{1-p}$ when $p > n^{-\epsilon}$ for any given $\epsilon > 0$. In a specially sparse case when $d$ is a sufficiently large constant, for any fixed integer $k \geq 2$ and $G \in \mathcal{G}(n, p)$, Kang and Manggala $[4]$ improved the upper bound of $um_k(G)$ to $um_k(G) \leq (1 + o(1))^{\frac{kn \log d}{2q^{k-1}}}$. It is speculated in $[5]$ that the upper bound here is close to the correct value of $um_k(G)$ when $k \geq 2$.

Recently, Cooley et al. in $[2]$ confirmed this conclusion when $k = 2$, $d \geq c$ and $d = o(n)$ for some large enough constant $c$. They showed that $um_2(G) = (1 + o(1))^{\frac{n \log d}{d}}$ relying on two main ingredients: the second moment method and Talagrand’s inequality. Talagrand’s inequality is a useful tool to show that a random variable is tightly concentrated under certain conditions. In fact, they showed that $um_2(G) = (1 + o(1)) \log_h d$ with $h = \frac{1}{1-p}$ in a broader range of $p$. Unfortunately, the conditions of Talagrand’s inequality only hold on $um_2(G)$. It is interesting to consider the lower bounds of $um_k(G)$ when $k \geq 3$, while there are less discussions. Tian in $[7]$ showed a trivial lower bound of $um_k(G) = \Omega(\log n)$ when $d$ is a sufficiently large constant.

In this paper, for any fixed integer $k \geq 2$ and $G \in \mathcal{G}(n, p)$, as a further step of $[2, 3, 5, 7]$, we consider the bounds of $um_k(G)$. In fact, we mainly investigate $e(M_k)$ of any maximal $k$-matching $M_k$ in $G \in \mathcal{G}(n, p)$. Let $d(n) = np$ be the expected degree of $G$, and we usually write $d$ instead of $d(n)$.

**Theorem 1.1.** Let $k \geq 2$ be a fixed integer, $G \in \mathcal{G}(n, p)$ with $d \geq c$ and $d^{k-1} = o(n)$ for some large enough constant $c$. For any maximal $k$-matching $M_k$ in $G$, with high probability,
Theorem 1.3. Let $\ell(n) \leq e(M_k) \leq (1 + o(1)) \frac{kn\log d}{2d^{k-1}}$.

As a corollary of Theorem 1.1, we actually obtain the bounds of $um_k(G)$ when $k \geq 3$.

Corollary 1.2. Let $k \geq 3$ be a fixed integer, $G \in \mathcal{G}(n, p)$ with $d \geq c$ and $d^{k-1} = o(n)$ for some large enough constant $c$. Then, with high probability, $(1 + o(1)) \frac{(k-1)n\log d}{4d^{k-1}} \leq um_k(G) \leq (1 + o(1)) \frac{kn\log d}{2d^{k-1}}$.

In fact, we can run a randomized greedy algorithm to generate a specially large $k$-matching with asymptotic size $\frac{kn\log d}{2d^{k-1}}$ for an input graph $G \in \mathcal{G}(n, p)$. It improves the lower bound of $um_k(G)$ in Corollary 1.2.

Theorem 1.3. Let $k \geq 2$ be a fixed integer, $G \in \mathcal{G}(n, p)$ with $d \geq c$ and $d^{k-1} = o(n)$ for some large enough constant $c$. With high probability, there exists a $k$-matching with size $(1 + o(1)) \frac{kn\log d}{4d^{k-1}}$.

We remark that $d^{k-1} = o(n)$ is a necessary condition in the proof of Theorem 1.1 and Theorem 1.3. The way to obtain the upper bound of the largest $k$-matching in Theorem 1.1 is similar with the one when $d$ is a large constant in [5]. For $k = 2$, it coincides with the upper bound $um_2(G) \leq (1 + o(1)) \frac{n\log d}{d}$ when $d = \Omega(1)$ and $d = o(n)$ in [2, 3]. The way to obtain the lower bound of $um_k(G)$ for $k \geq 3$ must be distinct with the one in [2]. We firstly show the lower bound of $e(M_k)$ for any maximal $k$-matching $M_k$ in Theorem 1.1, and we further investigate an interesting property for some given number of vertices of $G$ to show the other one in Theorem 1.3. We believe it is hard to improve the results here.

The organization of our paper is as follows. Notation and auxiliary results used throughout the paper are presented in Section 2. For $G \in \mathcal{G}(n, p)$, the upper and lower bounds of $e(M_k)$ for any maximal $k$-matching $M_k$ in Theorem 1.1 are discussed in Section 3 and Section 4, respectively. In Section 5, we design a simple randomized greedy algorithm to generate a large $k$-matching $M_k$ in $G$ and prove its efficiency. The last section concludes the works.

2 Preliminaries

In this section, we introduce a few useful lemmas that help us prove the main results and fix our notations. We will use the standard Landau notations $o(\cdot), O(\cdot), \Omega(\cdot)$ and $\Theta(\cdot)$. Throughout the following sections we assume that $n \to \infty$, $c$ is a large enough constant and $k \geq 2$ is a fixed integer. When not otherwise explicitly stated, the asymptotics in this notation are with respect to $n$ or with respect to $c$. For example, for two positive-valued functions $f, g$, we write $f \sim g$ to denote $\lim_{n \to \infty} f/g = 1$ or $\lim_{c \to \infty} f/g = 1$, and $f = o(g)$ to denote $\lim_{n \to \infty} f/g = 0$ or $\lim_{c \to \infty} f/g = 0$. For an event $\mathcal{A}$ and a random variable $Z$ in an arbitrary probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $\overline{\mathcal{A}}$ denotes the complement of the event $\mathcal{A}$, $\mathbb{P}[\mathcal{A}], \mathbb{E}[Z]$ and $\mathbb{V}[Z]$ denote the probability of $\mathcal{A}$, the expectation and the variance of $Z$. An event is said to occur with high probability (w.h.p. for short), if the probability that it holds tends to 1 when $n \to \infty$. Recall
that $\text{Bin}(n, p)$ denotes a Binomially distributed random variable with parameters $n$ and $p$, 
that is, it is the sum of $n$ independent variables, each equal to 1 with probability $p$ and 0 
otherwise. All logarithms are natural. The floor and ceiling signs are omitted whenever they 
are not crucial.

We use the following standard Chernoff inequality to estimate the tail probability of the 
binomial distribution, and Janson’s inequality to estimate the probability that none of a set 
of bad events occur when these events are mostly independent, see [1]. The statements have 
been adapted slightly from [1] in terms of our settings.

**Lemma 2.1.** (Chernoff Inequality) For any $\delta \in (0, 1)$,
\[
\mathbb{P}[|\text{Bin}(n, p) - np| > \delta np] < 2 \exp[-\delta^2 np/2].
\]

Let $G \subseteq \mathcal{G}(n, p)$ be a random edge subset of the complete graph $K_n$ and $u, v$ be any two 
vertices in $[n]$. Let $P_1, P_2, \cdots, P_n$ denote all paths of length at most $k - 1$ in $K_n$ to connect 
the vertex $u$ with the vertex $v$. We have only one edge in $K_n$ to connect $u$ with $v$. While 
there are $(n - 2) \cdots (n - i)$ paths of length $i$ to connect $u$ with $v$ when $2 \leq i \leq k - 1$ and 
k \geq 3. In summary, we have

\[
l_1 = \begin{cases} 
1, & k = 2, \\
1 + \sum_{i=2}^{k-1} (n - 2) \cdots (n - i), & k \geq 3.
\end{cases}
\]

Let $A_i$ be the event that the path $P_i$ exists in $G$ for $i \in [l_1]$. For $i, j \in [l_1]$, we write $i \sim j$ if 
i \neq j and $P_i \cap P_j \neq \emptyset$. Let $\Delta = \sum_{i\sim j} \mathbb{P}[A_i \cap A_j]$ and $U = \prod_{i=1}^{l_1} (1 - \mathbb{P}[A_i])$.

**Lemma 2.2.** (Janson’s Inequality) If $\mathbb{P}(A_i) \leq \frac{1}{2}$ for any $i \in [l_1]$, then $U \leq \mathbb{P}[\cap_{i=1}^{l_1} \overline{A_i}] \leq U \exp[\Delta]$.

Similarly, let $M$ be a matching with size $m$ in $K_n$. Assume that $uu'$ and $vv'$ are two 
distinct edges in $M$, which have $\binom{m}{2}$ ways to choose them. To count the number of paths of 
length at most $k - 1$ in $K_n$ to connect the edge $uu'$ with the edge $vv'$, there are four edges $uv,$ 
$uw$, $uv'$ or $uw'$ in $K_n$ between the edges $uu'$ and $vv'$; $4(n - 4) \cdots (n - i - 2)$ counts the number 
of paths of length $i$ in $K_n$ to connect the edge $uu'$ with the edge $vv'$ when $2 \leq i \leq k - 1$ and 
k \geq 3. Let $P_{1i}', P_{2i}', \cdots, P_{l_2i}'$ be the paths of length at most $k - 1$ in $K_n$ between any two edges 
of $M$. We have

\[
l_2 = \begin{cases} 
\binom{m}{2}, & k = 2, \\
\binom{m}{2} + 4 \binom{m}{2} \sum_{i=2}^{k-1} (n - 4) \cdots (n - i - 2), & k \geq 3.
\end{cases}
\]

Let $B_i$ be the event that the path $P_{1i}'$ exists in $G$ for $i \in [l_2]$. For $i, j \in [l_2]$, we write $i \sim j$ if 
i \neq j and $P_{1i}' \cap P_{1j}' \neq \emptyset$. Let $\Delta' = \sum_{i\sim j} \mathbb{P}[B_i \cap B_j]$ and $U' = \prod_{i=1}^{l_2} (1 - \mathbb{P}[B_i])$.

**Lemma 2.3.** (Janson’s Inequality) If $\mathbb{P}(B_i) \leq \frac{1}{2}$ for any $i \in [l_2]$, then $U' \leq \mathbb{P}[\cap_{i=1}^{l_2} \overline{B_i}] \leq U' \exp[\Delta']$.  

4
We make use of these inequalities to show some probability inequalities of distances in $G \in \mathcal{G}(n,p)$.

**Lemma 2.4.** Let $k \geq 2$ be a fixed integer, $G \in \mathcal{G}(n,p)$, $d = np$ satisfying $d \geq c$ and $d^{k-1} = o(n)$ for some large enough constant $c$. Define $p_d$ to satisfy the equation $np_d = d^{k-1}$.

(a) Given two vertices $u$ and $v$ in $[n]$, let $d_G(u,v)$ denote the distance between $u$ and $v$ in $G$. Then,

$$P[d_G(u,v) \geq k] = \begin{cases} 1 - p, & k = 2; \\ (1 - p_d) \exp[O(n^{k-3}p^{k-2} + n^{2k-4}p^{2k-2})], & k \geq 3. \end{cases}$$

(b) Let $M$ be a matching with size $m$ in $K_n$ and $m = O\left(\frac{n \log d}{1 + k} \right)$. Then,

$$P[M \text{ is a } k\text{-matching in } G] = \begin{cases} p^m(1-p)^{\binom{n}{2}}, & k = 2; \\ p^m(1-p_d)^{\binom{m}{2}} \exp[O(m^2n^{k-3}p^{k-2} + m^2n^{2k-4}p^{2k-2} + m^3n^{2k-5}p^{2k-3})], & k \geq 3. \end{cases}$$

**Proof.** (a) For $k = 2$, it is obviously true because $P[d_G(u,v) \geq 2] = 1 - p$. For $k \geq 3$, we consider $\Delta$ and $U$ in Lemma 2.2 because $P[d_G(u,v) \geq k] = P[r_i^{\Delta_i}\mathcal{A}_i]$. Note that $p = d/n = o(1)$ and $k \geq 3$ a fixed integer, then we have

$$U = (1 - p) \prod_{i=2}^{k-1} (1 - p^i)^{(n-2)\cdots(n-i)}$$

$$= (1 - p) \prod_{i=2}^{k-1} \left[1 - (n - 2) \cdots (n - i)p^i + O\left(n^{2i-2}p^{2i}\right)\right]$$

$$= \left[1 - p - \sum_{i=2}^{k-1} (n - 2) \cdots (n - i)p^i\right] \cdot \left[1 + O\left(n^{2k-4}p^{2k-2}\right)\right], \quad (2.1)$$

where

$$(1 - p^i)^{(n-2)\cdots(n-i)} = 1 - (n - 2) \cdots (n - i)p^i + O\left(n^{2i-2}p^{2i}\right)$$

is true because Taylor’s expansion and $n^{i-1}p^i = o(1)$ when $d^{k-1} = o(n)$ and $2 \leq i \leq k - 1$; the last equality is true because $n^{2k-4}p^{2k-2}$ is the largest term in $n^{2i-2}p^{2i}$ when $2 \leq i \leq k - 1$ and $d \geq c$ for some large enough constant $c$.

Since

$$p + \sum_{i=2}^{k-1} (n - 2) \cdots (n - i)p^i = \sum_{i=1}^{k-1} n^{i-1}p^i + O\left(n^{k-3}p^{k-1}\right)$$

and $n^{k-3}p^{k-1} = o(n^{2k-4}p^{2k-2})$, by the equation in (2.1), it follows that

$$U = (1 - \sum_{i=1}^{k-1} n^{i-1}p^i) \cdot \left[1 + O\left(n^{2k-4}p^{2k-2}\right)\right]$$

$$= (1 - \sum_{i=1}^{k-1} n^{i-1}p^i) \exp\left[O\left(n^{2k-4}p^{2k-2}\right)\right], \quad (2.2)$$
where
\[
\exp\left[O\left(n^{2k-4}p^{2k-2}\right)\right] = 1 + O\left(n^{2k-4}p^{2k-2}\right)
\]
is true because Taylor’s expansion and \(n^{2k-4}p^{2k-2} = d^{2k-2}/n^2 = o(1)\) when \(d^{k-1} = o(n)\).

By \(d = np\) and \(d^{k-1} = npd\), we have
\[
1 - \sum_{i=1}^{k-1} n^{i-1}p^i = 1 - \frac{p(1 - (np)^{k-1})}{1 - np} = 1 - \frac{p(1 - d^{k-1})}{1 - d} = 1 + \frac{1}{n} - \frac{pd}{1 - \frac{1}{d}}
\]
in (2.2). Note that
\[
1 + \frac{1}{n} - \frac{pd}{1 - \frac{1}{d}} = 1 + \left(\frac{1}{n} - pd\right)\left(1 + O\left(\frac{1}{d}\right)\right) = 1 - pd + O(n^{-3}p^{k-2})
\]
because \(d \geq c\) for some large enough constant \(c\) and \(O(pd/d) = O(n^{-3}p^{k-2})\). Thus, we obtain
\[
1 - \sum_{i=1}^{k-1} n^{i-1}p^i = (1 - pd)\exp\left[O\left(n^{k-3}p^{k-2}\right)\right]. \tag{2.3}
\]
By the equations in (2.2) and (2.3), we finally have
\[
U = (1 - pd)\exp\left[O\left(n^{k-3}p^{k-2} + n^{2k-4}p^{2k-2}\right)\right]. \tag{2.4}
\]

We need to bound \(\Delta = \sum_{i, j} \mathbb{P}[A_i \cup A_j]\) in Lemma 2.2 below. Assume that the paths \(P_i\) and \(P_j\) between \(u\) and \(v\) are of lengths \(\ell_i\) and \(\ell_j\) satisfying \(2 \leq \ell_i \leq \ell_j \leq k - 1\) and the size of common edges of them is denoted by \(t \geq 1\). Note that \(\Delta = \sum_{i, j} \mathbb{P}[A_i \cup A_j] = \sum_{i, j} \mathbb{P}[A_j|A_i] \cdot \mathbb{P}[A_i]\). The ways to choose \(P_i\) in \(K_n\) are at most \(n^{\ell_i-1}\) and \(\mathbb{P}[A_i] = p^{\ell_i}\). Fix the path \(P_i\), for a given set of \(t\) edges in \(P_i\), then the ways to choose \(P_j\) in \(K_n\) are at most \(n^{\ell_j-t-1}\) and \(\mathbb{P}[A_j|A_i] = p^{\ell_j-t}\). Thus, we have
\[
\Delta = \sum_{i, j} \mathbb{P}[A_j|A_i] \cdot \mathbb{P}[A_i] \leq \sum_{i, j} \sum_{\ell_i=2}^{k-1} n^{\ell_i-1}p^{\ell_i} \sum_{\ell_j=\ell_i}^{k-1} \sum_{t=1}^{\ell_i-1} \binom{\ell_i}{t} n^{\ell_j-t-1}p^{\ell_j-t}. \tag{2.5}
\]
Firstly, for fixed \(t\), the geometric series \(\sum_{\ell_j=\ell_i}^{k-1} \binom{\ell_i}{t} n^{\ell_j-t-1}p^{\ell_j-t} = O(n^{k-t-2}p^{k-t-1})\) because the sum of this expression over \(\ell_j \geq \ell_i\) is bounded by an increasing geometric series with common ratio \(d = np\) dominated by the term when \(\ell_j = k - 1\). Secondly, the geometric series \(\sum_{t=1}^{\ell_i-1} n^{k-t-2}p^{k-t-1} = O(n^{k-3}p^{k-2})\) because the sum of this expression over \(t \geq 1\) is bounded by a decreasing geometric series with common ratio \(1/d\) dominated by the term when \(t = 1\). Thus, for any fixed integer \(k \geq 3\) and \(d \geq c\) for some large enough constant \(c\), we have
\[
\sum_{\ell_j=\ell_i}^{k-1} \sum_{t=1}^{\ell_i-1} \binom{\ell_i}{t} n^{\ell_j-t-1}p^{\ell_j-t} = O\left(n^{k-3}p^{k-2}\right).
\]
At last, the sum of the expression $\sum_{\ell_i=2}^{k-1} n^{\ell_i-1}p^{\ell_i}$ over $\ell_i \geq 2$ is bounded by an increasing geometric series with common ratio $d = np$ dominated by the term when $\ell_i = k - 1$, thus $\sum_{\ell_i=2}^{k-1} n^{\ell_i-1}p^{\ell_i} = O(n^{k-2}p^{k-1})$. By the equation in (2.5), we have

$$\Delta = O(n^{2k-5}p^{2k-3}).$$

(2.6)

Since $n^{2k-5}p^{2k-3} = o(n^{k-3}p^{k-2})$ when $k \geq 3$ and $d^{k-1} = o(n)$, combining the equations in (2.4), (2.6) and Lemma 2.2, we have

$$\mathbb{P}[d_G(u,v) \geq k] = (1 - p_d) \exp \left[ O \left( n^{k-3}p^{k-2} + n^{2k-4}p^{2k-2} \right) \right].$$

(b) Firstly, there exists a penalty factor $p^n$ in the final count because the edges of $M$ are present in $G \in \mathcal{G}(n,p)$. For $k = 2$, $\mathbb{P}[M]$ is a $k$-matching in $G] = p^m(1 - p)^4(m)$. For $k \geq 3$, we consider $\Delta'$ and $U'$ in Lemma 2.3 because $\mathbb{P}[M]$ is a $k$-matching in $G] = p^m \mathbb{P}[\bigcap_{i=1}^m \mathcal{B}_i]$. Following the same discussions between the equation in (2.2) and the equation in (2.4), we have

$$U' = \left[ (1 - p) \prod_{i=2}^{k-1} (1 - p^i)^{(n-4)\cdots(n-i-2)} \right]^{4(m)} = (1 - \sum_{i=1}^{k-1} n^{i-1}p^i)^{4(m)} \exp \left[ O \left( m^2n^{2k-4}p^{2k-2} \right) \right] = (1 - p_d)^{4(m)} \exp \left[ O \left( m^2n^{k-3}p^{k-2} + m^2n^{2k-4}p^{2k-2} \right) \right].$$

(2.7)

Now we bound $\Delta' = \sum_{i\sim j} \mathbb{P}[\mathcal{B}_i \cap \mathcal{B}_j]$ in Lemma 2.3 below. Similarly, assume that one path between two edges in $M$, denoted by $P'_i$, is of length $\ell_i$ satisfying $2 \leq \ell_i \leq k - 1$. Let $P'_j$ be another path of length $\ell_j$, between two edges in $M$ with $2 \leq \ell_j \leq k - 1$. Let the size of common edges of the paths $P'_i$ and $P'_j$ be denoted by $t$ with $t \geq 1$. The ways to choose $P'_i$ in $K_n$ are at most $4(m) n^{\ell_i-1}$ and $\mathbb{P}[\mathcal{B}_i] = p^\ell_i$. Fix the path $P'_i$, for a given set of $t$ edges in $P'_i$, then the ways to choose $P'_j$ in $K_n$ are at most $4(m) n^{\ell_j-t-2}$ if both end-vertices of the path $P'_j$ do not belong to the edges where the ones of the path $P'_i$ are, or at most $2mn^{\ell_j-t-1}$ if one end-vertex of the path $P'_j$ belongs to the edges where one of the end-vertices of the path $P'_i$ is. By $\mathbb{P}[\mathcal{B}_j|\mathcal{B}_i] = p^{\ell_j-t}$ and $4(m) n^{\ell_j-t-2} p^{\ell_j-t} < 2mn^{\ell_j-t-1} p^{\ell_j-t}$ when $m = O(\frac{n \log d}{d^2})$, we have

$$\Delta' = \sum_{i\sim j} \mathbb{P}[\mathcal{B}_j|\mathcal{B}_i] \cdot \mathbb{P}[\mathcal{B}_i] \leq \sum_{\ell_i=2}^{k-1} \sum_{\ell_j=\ell_i}^{k-1} 4 \left( \frac{m}{2} \right) n^{\ell_i-1}p^{\ell_i} \sum_{\ell_j=\ell_i}^{k-1} \sum_{t=1}^{\ell_i} \sum_{t=1}^{\ell_j} \left( \frac{\ell_i}{t} \right) \left( \frac{\ell_j}{t} \right) 2mn^{\ell_j-t-1} p^{\ell_j-t}. $$

(2.8)

Likewise, following the same discussions between the equation in (2.5) and the equation in (2.6), for any fixed integer $k \geq 3$, $d \geq c$ for some large enough constant $c$ and $p = o(1)$, we have

$$\sum_{\ell_j=\ell_i}^{k-1} \sum_{t=1}^{\ell_i} \sum_{t=1}^{\ell_j} \left( \frac{\ell_i}{t} \right) \left( \frac{\ell_j}{t} \right) 2mn^{\ell_j-t-1} p^{\ell_j-t} = O \left( mn^{k-3}p^{k-2} \right) $$
because the main term in this series is when \( \ell_j = k - 1 \) and \( t = 1 \). By the equation in (2.8), we further have

\[
\Delta' = O(m^3n^{2k-5}p^{2k-3}) \tag{2.9}
\]

because the main term in \( \sum_{i=2}^{k-1} 4(m) \ell_i^{-1} p_i \) is \( 4(m) n^{k-2} p^{k-1} \) when \( \ell_i = k - 1 \). By the equations in (2.7), (2.9) and Lemma 2.3, the proof of (b) is complete. \( \square \)

**Lemma 2.5.** Let \( k \geq 2 \) a fixed integer, \( d = np \geq c \) and \( d^{k-1} = o(n) \) for some large enough constant \( c \). Define \( p_d \) to satisfy the equation \( np_d = d^{k-1} \) and

\[
f(x) = \frac{1}{\sqrt{2\pi x}} \left( \frac{edn}{2x} \right)^x (1 - p_d)^{2x(x-1)}.
\]

Then \( f(x) \) is increasing in \( x \) when \( 1 \leq x \leq \frac{(k-1)n \log d}{4d^{k-1}} \).

**Proof.** Given \( n \) and \( d \) satisfying the assumption, take the logarithm to \( f(x) \). Differentiate \( \log(f(x)) \) on \( x \). Thus, we have

\[
f'(x) = f(x) \left[ -\frac{1}{2x} + \log \left( \frac{edn}{2x} \right) - 1 + (4x - 2) \log(1 - p_d) \right]. \tag{2.10}
\]

Let

\[
g(x) = -\frac{1}{2x} + \log \left( \frac{edn}{2x} \right) - 1 + (4x - 2) \log(1 - p_d). \tag{2.11}
\]

Note that

\[
g'(x) = \frac{1}{2x^2} - \frac{1}{x} + 4 \log(1 - p_d) < 0
\]

because \( x \geq 1 \) and \( \log(1 - p_d) \sim -p_d = -d^{k-1}/n \) when \( p_d = o(1) \), then we have \( g(x) \) is a decreasing function in \( x \). The function \( g(x) \) takes its minimum when \( x = \frac{(k-1)n \log d}{4d^{k-1}} \), that is

\[
g(x) \geq \frac{(k-1)n \log d}{4d^{k-1}}
\]

\[
= -\frac{2d^{k-1}}{(k-1)n \log d} + \log \left( \frac{2ed^k}{(k-1) \log d} \right) - 1 + \left( \frac{(k-1)n \log d}{d^{k-1}} - 2 \right) \log(1 - p_d)
\]

\[
= -\frac{2d^{k-1}}{(k-1)n \log d} + \frac{2}{(k-1) \log d} + k \log d - \log \log d
\]

\[
+ \left( \frac{(k-1)n \log d}{d^{k-1}} - 2 \right) \log(1 - p_d), \tag{2.12}
\]

where \( \frac{d^{k-1}}{(k-1)n \log d} = o(1) \) by \( d^{k-1} = o(n) \), and

\[
\left( \frac{(k-1)n \log d}{d^{k-1}} - 2 \right) \log(1 - p_d)
\]

\[
\sim -\frac{(k-1)n \log d}{d^{k-1}} p_d
\]

\[
= -(k-1) \log d \tag{2.13}
\]
by \( \log(1 - p_d) \sim -p_d \) and \( np_d = d^{k-1} \). Then we further have

\[
g(x) \geq \log \frac{2}{k-1} + \log d - \log \log d + o(1) > 0
\]

because \( d \geq c \) for some large enough constant \( c \). It follows that \( f'(x) > 0 \) from the equations in (2.10) and (2.11).

The proof of Lemma 2.5 is complete. \( \square \)

**Remark 2.6.** It is necessary to mention that we cannot show the monotonicity of \( f(x) \) when \( x \leq \frac{kn \log d}{4dk} \) based on the proof of Lemma 2.5 because \( g(x) \sim \log \frac{2}{k} - \log \log d < 0 \) when \( x = \frac{kn \log d}{4dk} \) and \( d \geq c \) for some large enough constant \( c \).

### 3 Upper bound in Theorem 1.1

Kang and Manggala [5] showed an upper bound of \( um_k(G) \) when the expected degree \( d = np \) is a large enough constant \( c \), which improves the result in [3]. In fact, we generalize the range of the expected degree \( d = np \) based on their proof from \( d = c \) to \( d \geq c \) and \( d^{k-1} = o(n) \) for some large enough constant \( c \) with better analysis.

Let \( M \) be the set of matchings with size \( m \) in \( K_n \), \( M_i \in M \) for \( 1 \leq i \leq t \) be all matchings in \( M \), where

\[
t = \binom{n}{2m} \binom{2m}{2, \ldots, 2} \frac{1}{m!}.
\]

Let \( I_i \) be the indicator random variable of the event that \( M_i \) is a \( k \)-matching in \( G \). Clearly, \( X_m = \sum_{i=1}^t I_i \). Thus, we have

\[
\mathbb{E}[X_m] = \binom{n}{2m} \binom{2m}{2, \ldots, 2} \frac{1}{m!} \mathbb{P}[M_i \text{ is a } k\text{-matching}]. 
\]  

(3.1)

**Proof of Upper Bound in Theorem 1.1.** Let

\[
m = (1 + o(1)) \frac{kn \log d}{2d^{k-1}}.
\]  

(3.2)

It is easy to verify that \( m^2 n^{k-3} p^{k-2} = o(m^3 n^{2k-5} p^{2k-3}) \) when \( d \geq c \) for some large enough constant \( c \).

By the equation in (3.1) and Lemma 2.4 (b), we have

\[
\mathbb{E}[X_m] = \binom{n}{2m} \binom{2m}{2, \ldots, 2} \frac{1}{m!} p^m (1 - p_d)^{m(1 - m)} \exp\left[O(m^2 n^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3})\right] \\
= \frac{n!}{m!(n - 2m)!} (\frac{n}{2})^m (1 - p_d)^{m(m - 1)} \exp\left[O(m^2 n^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3})\right].
\]
Note that \(\frac{n!}{(n-2m)!} \sim n^{2m}\) when \(m\) is in (3.2), \(m! \sim \sqrt{2\pi m (m/e)^m}\) when \(m \to \infty\) by the Stirling formula, we further have

\[
\mathbb{E}[X_m] \sim \frac{n^{2m}}{\sqrt{2\pi m}} \left(\frac{ep}{2m}\right)^m (1-p_d)^{2m(m-1)} \exp\left[O(m^2 n^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3})\right]
\]

\[
= \frac{1}{\sqrt{2\pi m}} \left(\frac{epn^2}{2m}\right)^m (1-p_d)^{2m(m-1)} \exp\left[O(m^2 n^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3})\right]
\]

\[
= \frac{1}{\sqrt{2\pi m}} \left(\frac{edn}{2m}\right)^m (1-p_d)^{2m(m-1)} \exp\left[O(m^2 n^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3})\right],
\]

where the last equality is true because \(d = np\). Note that \(1 - p_d \sim \exp[-p_d]\) for \(p_d = o(1)\),

\[
\mathbb{E}[X_m] \sim \frac{1}{\sqrt{2\pi m}} \exp\left[ m \left(\log\left(\frac{edn}{2m}\right) - 2(m-1)p_d + O(mn^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3})\right)\right].
\]

By the equation in (3.2), we have

\[
O(mn^{2k-4} p^{2k-2}) = O\left(\frac{d^{k-1} \log d}{n}\right), \quad O(m^2 n^{2k-5} p^{2k-3}) = O\left(\frac{\log^2 d}{d}\right).
\]

Note that \(d^{k-1} \log d = o(n)\) because

\[
(k-1) \log d + \log \log d = (k-1) \log d \cdot \left[1 + O\left(\frac{\log \log d}{\log d}\right)\right] = o(\log n)
\]

is true when \(d \gg c\) and \(d^{k-1} = o(n)\) for some large enough constant \(c\), then we have \(O(mn^{2k-4} p^{2k-2}) = o(1)\) and \(O(m^2 n^{2k-5} p^{2k-3}) = o(1)\). It follows that

\[
\mathbb{E}[X_m] \sim \frac{1}{\sqrt{2\pi m}} \exp\left[ m \left(\log\left(\frac{edn}{2m}\right) - 2(m-1)p_d + o(1)\right)\right]. \tag{3.3}
\]

Likewise, by the equation in (3.2), we also have

\[
\log\left(\frac{edn}{2m}\right) \sim k \log d - \log \log d,
\]

\[
2(m-1)p_d \sim k \log d;
\]

and then

\[
\log\left(\frac{edn}{2m}\right) - 2(m-1)p_d + o(1) \sim -\log \log d < 0
\]

in (3.3) such that \(\mathbb{E}[X_m] \to 0\). Thus, \(\mathbb{P}[X_m > 0] \to 0\), which implies w.h.p., for any \(k\)-matching \(M_k\),

\[
e(M_k) \leq (1 + o(1)) \frac{kn \log d}{2d^{k-1}}.
\]

The proof of the upper bound in Theorem 1.1 is complete. \(\square\)
4 Lower bound in Theorem 1.1

In this section, we show a lower bound of the size of any maximal $k$-matching in Theorem 1.1. Let $\mathcal{M}$ be the set of matchings with size $m$ in $K_n$, $M \in \mathcal{M}$ and $G \in \mathcal{G}(n, p)$. Let $Y_m$ denote the number of maximal $k$-matchings with size $m$ contained in $G$. Thus, we have

$$\mathbb{E}[Y_m] = \binom{n}{2m} \left( \frac{2m}{2, \ldots, 2} \right) \frac{1}{m!} \mathbb{P}[M \text{ is a maximal } k\text{-matching in } G]. \quad (4.1)$$

For any given real number $\epsilon > 0$, define

$$m^* = (k - 1 - \epsilon) \frac{n \log d}{4d^{k-1}}. \quad (4.2)$$

In order to show the lower bound $e(M_k) \geq (1 + o(1)) \frac{(k-1)n \log d}{4d^{k-1}}$ for any maximal $k$-matching $M_k$ in Theorem 1.1 is true, we only need to show $\mathbb{E}[\sum_{m \leq m^*} Y_m] \rightarrow 0$ when $n \rightarrow \infty$ because

$$\mathbb{P} \left[ \sum_{m \leq m^*} Y_m > 0 \right] \leq \mathbb{E} \left[ \sum_{m \leq m^*} Y_m \right]$$

by Markov’s inequality. Then, for any maximal $k$-matching $M_k$ in $G$, w.h.p., we have $e(M_k) > m^*$.

**Proof of Lower bound in Theorem 1.1.** Fix $m$ to be any positive integer satisfying $m \leq m^*$. Let $M$ be a matching with size $m$ in $K_n$. In fact, the theorem follows from the claim below.

**Claim 4.1** With the assumption in Theorem 1.1, define $p_d$ to satisfy the equation $np_d = d^{k-1}$, then we have

$$\mathbb{P}[M \text{ is a maximal } k\text{-matching in } G] \leq 2p^m(1 - p_d)^{\binom{n}{m}} \times \exp \left[ -\frac{n}{8} d^{-(k-1-\epsilon)} + O(m^2 n^{k-3} p^{k-2} + m^2 n^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3}) \right].$$

We leave the proof of the above claim later. In fact, if the claim is true, by the equation in (4.1), then we have

$$\mathbb{E}[Y_m] < 2 \left( \binom{n}{2m} \left( \frac{2m}{2, \ldots, 2} \right) \frac{1}{m!} \mathbb{P}^{m}(1 - p_d)^{\binom{n}{m}} \times \exp \left[ -\frac{n}{8} d^{-(k-1-\epsilon)} + O(m^2 n^{k-3} p^{k-2} + m^2 n^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3}) \right] \right)$$

$$= \frac{2n!}{m!(n-2m)!} (\frac{p}{2})^{m} (1 - p_d)^{2m(m-1)} \times \exp \left[ -\frac{n}{8} d^{-(k-1-\epsilon)} + O(m^2 n^{k-3} p^{k-2} + m^2 n^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3}) \right]$$

$$< \frac{2n^{2m}}{m!} (\frac{p}{2})^{m} (1 - p_d)^{2m(m-1)} \times \exp \left[ -\frac{n}{8} d^{-(k-1-\epsilon)} + O(m^2 n^{k-3} p^{k-2} + m^2 n^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3}) \right]. \quad (4.3)$$
By the Stirling formula, $m! \geq \sqrt{2\pi m}(m/e)^m$ for all $m \geq 1$, the equation in (4.3) and $d = np$, we further have

$$
\mathbb{E}[Y_m] < \frac{2n^{2m}}{\sqrt{2\pi m}} \left( \frac{ep}{2m} \right)^m (1 - p_d)^{2m(m-1)}
\times \exp\left[ -\frac{n}{8} d^{-(k-1-\epsilon)} + O(m^2 n^{k-3} p^{k-2} + m^2 n^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3}) \right]
= \frac{2}{\sqrt{2\pi m}} \left( \frac{ep}{2m} \right)^m (1 - p_d)^{2m(m-1)}
\times \exp\left[ -\frac{n}{8} d^{-(k-1-\epsilon)} + O(m^2 n^{k-3} p^{k-2} + m^2 n^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3}) \right]
= \frac{2}{\sqrt{2\pi m}} \left( \frac{en}{2m} \right)^m (1 - p_d)^{2m(m-1)}
\times \exp\left[ -\frac{n}{8} d^{-(k-1-\epsilon)} + O(m^2 n^{k-3} p^{k-2} + m^2 n^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3}) \right].
$$

Since $m^* \leq \frac{(k-1)n \log d}{4d^{k-1}}$ in (4.2), by Lemma 2.5 and for all $1 \leq m \leq m^*$, it follows that

$$
\frac{1}{\sqrt{2\pi m}} \left( \frac{en}{2m} \right)^m (1 - p_d)^{2m(m-1)} \leq \frac{1}{\sqrt{2\pi m^*}} \left( \frac{en}{2m^*} \right)^{m^*} (1 - p_d)^{2m^*(m^*-1)},
$$

and

$$
\mathbb{E}[Y_m] < \frac{2}{\sqrt{2\pi m^*}} \left( \frac{en}{2m^*} \right)^{m^*} (1 - p_d)^{2m^*(m^*-1)}
\times \exp\left[ -\frac{n}{8} d^{-(k-1-\epsilon)} + O(m^2 n^{k-3} p^{k-2} + m^2 n^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3}) \right]
= \frac{2}{\sqrt{2\pi m^*}} \left( \frac{en}{2m^*} \right)^{m^*} (1 - p_d)^{2m^*(m^*-1)}
\times \exp\left[ -\frac{n}{8} d^{-(k-1-\epsilon)} + O(m^2 n^{k-3} p^{k-2} + m^2 n^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3}) \right],
$$

(4.4)

where the last equality is true because $m^2 n^{k-3} p^{k-2} = o(m^3 n^{2k-5} p^{2k-3})$ when $m^* = (k - 1 - \epsilon)n \log d$ in (4.2) and $d \geq c$ for some large enough constant $c$. For all $1 \leq m \leq m^*$, adding the corresponding inequalities in (4.4) together,

$$
\mathbb{E}\left[ \sum_{m \leq m^*} Y_m \right] < \frac{2m^*}{\sqrt{2\pi m^*}} \left( \frac{en}{2m^*} \right)^{m^*} (1 - p_d)^{2m^*(m^*-1)}
\times \exp\left[ -\frac{n}{8} d^{-(k-1-\epsilon)} + O\left( m^2 n^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3} \right) \right]
= \sqrt{2m^*} \left( \frac{en}{2m^*} \right)^{m^*} (1 - p_d)^{2m^*(m^*-1)}
\times \exp\left[ -\frac{m^* d^\varepsilon}{2(k - 1 - \epsilon) \log d} + O\left( m^2 n^{2k-4} p^{2k-2} + m^3 n^{2k-5} p^{2k-3} \right) \right],
$$

where the last equality is true because

$$
\exp\left[ -\frac{n}{8} d^{-(k-1-\epsilon)} \right] = \exp\left[ -\frac{m^* d^\varepsilon}{2(k - 1 - \epsilon) \log d} \right].
$$
when \( m^* = (k - 1 - \epsilon) \frac{n \log d}{4d^k} \) in (4.2). By \( 1 - p_d \sim \exp[-p_d] \) for \( p_d = o(1) \),

\[
\mathbb{E} \left[ \sum_{m \leq m^*} Y_m \right] < \sqrt{2m^*} \exp \left[ m^* \left( \log \left( \frac{edn}{2m^*} \right) - 2(m^* - 1)p_d \right) - \frac{d^k}{2(k - 1 - \epsilon) \log d} + O \left( m^* n^{2k-4}p^{2k-2} + m^* n^{2k-5}p^{2k-3} \right) \right]. \tag{4.5}
\]

Again by \( m^* = (k - 1 - \epsilon) \frac{n \log d}{4d^k} \to \infty \) in (4.2), \( d^{k-1} = np_d \) and \( d \geq c \) for some large enough constant \( c \), it follows that

\[
\log \left( \frac{edn}{2m^*} \right) = \log \left( \frac{2ed^k}{(k - 1 - \epsilon) \log d} \right) \sim k \log d, \tag{4.6}
\]

\[
2(m^* - 1)p_d \sim \frac{1}{2} (k - 1 - \epsilon) \log d. \tag{4.7}
\]

Similarly, we also have

\[
m^* n^{2k-4}p^{2k-2} = \frac{(k - 1 - \epsilon)^2 (\log d)^2}{16d} = o(1), \tag{4.8}
\]

\[
m^* n^{2k-5}p^{2k-3} = \frac{(k - 1 - \epsilon) d^{k-1} \log d}{4n} = o(1). \tag{4.9}
\]

Putting the equations in (4.6)-(4.9) into the equation in (4.5), we have

\[
\log \left( \frac{edn}{2m^*} \right) - 2(m^* - 1)p_d - \frac{d^k}{2(k - 1 - \epsilon) \log d} + O \left( m^* n^{2k-4}p^{2k-2} + m^* n^{2k-5}p^{2k-3} \right) < 0
\]

because \( \frac{d^k}{2(k - 1 - \epsilon) \log d} \) dominates all these terms when \( d \geq c \) for some large enough constant \( c \). Finally, using the equation in (4.5), we have

\[
\mathbb{E} \left[ \sum_{m \leq m^*} Y_m \right] \to 0,
\]

and then

\[
\mathbb{P} \left[ \sum_{m \leq m^*} Y_m > 0 \right] \to 0
\]

by Markov’s inequality. W.h.p. we have \( e(M_k) > m^* \) for any maximal \( k \)-matching \( M_k \) in \( G \).

The proof of the lower bound in Theorem 1.1 is complete. \( \square \)

In order to formally finish the proof of Theorem 1.1, it is necessary to prove Claim 4.1.

**Proof of Claim 4.1.** Recall that \( m \) is a fixed positive integer satisfying \( m \leq m^* \) and \( M \) is a matching with size \( m \) in \( K_n \). Let \( \Gamma_{\geq k}(M) \) denote the set of vertices in \( G \in \mathcal{G}(n, p) \) whose distances are at least \( k \) to each vertex in \( M \). If \( M \) is a maximal \( k \)-matching in \( G \), then \( \Gamma_{\geq k}(M) \)
is either empty or an independent set, otherwise it contradicts with the maximal property of $M$. Define two events $E$ and $F$ as
\[
E = \{ \Gamma_{\geq k}(M) \text{ is an independent set} \} \quad \text{and} \quad F = \left\{ \left| \Gamma_{\geq k}(M) \right| > \frac{n}{2} d^{-(k-1-\epsilon)} \right\}. \quad (4.10)
\]
By the total probability formula, we have
\[
\mathbb{P}[M \text{ is a maximal } k\text{-matching in } G] = \mathbb{P}[M \text{ is a maximal } k\text{-matching in } G|F] \cdot P[F] + \mathbb{P}[M \text{ is a maximal } k\text{-matching in } G|F^c] \cdot P[F^c],
\]
in which $\mathbb{P}[M \text{ is a maximal } k\text{-matching in } G|F] \cdot P[F] = \mathbb{P}[\{ M \text{ is a } k\text{-matching in } G \} \cap E \cap F] \cap F$ and $\mathbb{P}[\{ M \text{ is a } k\text{-matching in } G \} \cap F^c] \leq \mathbb{P}[\{ M \text{ is a } k\text{-matching in } G \} \cap F^c]$. It follows that
\[
\mathbb{P}[M \text{ is a maximal } k\text{-matching in } G] \leq \mathbb{P}[\{ M \text{ is a } k\text{-matching in } G \} \cap E \cap F] + \mathbb{P}[\{ M \text{ is a } k\text{-matching in } G \} \cap F^c]
= \mathbb{P}[M \text{ is a } k\text{-matching in } G] (\mathbb{P}[E \cap F] + \mathbb{P}[F^c]),
\]
where the last equality is true because the event $\{ M \text{ is a } k\text{-matching in } G \}$ is independent to the events $E \cap F$ and $F^c$. Combining with Lemma 2.4 (b), we further have
\[
\mathbb{P}[M \text{ is a maximal } k\text{-matching in } G] \leq p^m (1 - p_d) \binom{n}{2} \exp \left[ O(m^2 n^{-3} p^{-2} + m^2 n^{2k-4} p^{-2k} + m^2 n^{2k-5} p^{-2k-3}) \right]
\times \left( \mathbb{P}[E \cap F] + \mathbb{P}[F^c] \right). \quad (4.11)
\]
Hence, in order to finish the proof of Claim 4.1, it is necessary to show
\[
\mathbb{P}[F^c] < \exp \left[ -\frac{n}{8} d^{-(k-1-\epsilon)} \right], \quad (4.12)
\]
\[
\mathbb{P}[E \cap F] < \exp \left[ -\frac{n}{8} d^{-(k-1-\epsilon)} \right]. \quad (4.13)
\]
Firstly, we prove the inequality in (4.12). Fix a vertex $u$ not in $M$. By Lemma 2.4 (a), for any vertex $v$ in $M$, $\mathbb{P}[d_G(u, v) \geq k] = 1 - p$ for $k = 2$ and $\mathbb{P}[d_G(u, v) \geq k] = (1 - p_d) \exp[O(n^{k-3} p^{-2} + n^{2k-4} p^{-2k})]$ for $k \geq 3$. While for two different vertices $v_1$ and $v_2$ in $M$, the events $\{d_G(u, v_1) \geq k\}$ and $\{d_G(u, v_2) \geq k\}$ are not independent. The property of distance is a decreasing property, which means that if the event that $\{d_G(u, v_1) \geq k\}$ is true, then the probability of the event $\{d_G(u, v_2) \geq k\}$ is at least the corresponding value in the case of independence. Since there are exactly $2m$ vertices in $M$, we have the event $\{ u \in \Gamma_{\geq k}(M) \}$ occurs with probability at least $P$, where
\[
P = \begin{cases} (1 - p)^{2m}, & k = 2; \\
(1 - p_d)^{2m} \exp[O(mn^{k-3} p^{-2} + mn^{2k-4} p^{-2k})], & k \geq 3. \end{cases} \quad (4.14)
\]
Hence, the probability of the event that the variable $|\Gamma_{\geq k}(M)|$ is no greater than some value is dominated by the event that the variable $\text{Bin}(n-2m, P)$ is no greater than the same value, which means

$$
P[\mathcal{F}^c] = P\left[|\Gamma_{\geq k}(M)| \leq \frac{n}{2} d^{-\frac{(k-1-\epsilon)}{2}}\right] \leq P\left[\text{Bin}(n-2m, P) \leq \frac{n}{2} d^{-\frac{(k-1-\epsilon)}{2}}\right].$$  \hspace{1cm} (4.15)

For any $1 \leq m \leq m^*$ in (4.2), $d \geq c$ for some large enough constant $c$, we have $n-2m \sim n$ and

$$mn^{k-3}p^{k-2} = O\left(\frac{\log d}{d}\right) = o(1),$$

$$mn^{2k-4}p^{2k-2} = O\left(\frac{d^{k-1}\log d}{n}\right) = o(1).$$

Thus, we have $\exp[O(nm^{k-3}p^{k-2} + mn^{2k-4}p^{2k-2})] \sim 1$ and $P \sim (1-p_d)^{2m}$ in (4.14). By the equation in (4.15),

$$\P[\mathcal{F}^c] \leq \P\left[\text{Bin}(n-2m, P) \leq \frac{n}{2} d^{-\frac{(k-1-\epsilon)}{2}}\right] \sim \P\left[\text{Bin}(n, (1-p_d)^{2m}) \leq \frac{n}{2} d^{-\frac{(k-1-\epsilon)}{2}}\right].$$  \hspace{1cm} (4.16)

The expectation of $\text{Bin}(n, (1-p_d)^{2m})$ is

$$\E\left[\text{Bin}(n, (1-p_d)^{2m})\right] \geq n \exp\left[-\frac{2mp_d}{1-p_d}\right] \geq n \exp\left[-\frac{2m^*pd}{1-p_d}\right] = n \exp\left[-\frac{(k-1-\epsilon)\log d}{2\log d}\right] \sim nd^{-\frac{(k-1-\epsilon)}{2}},$$  \hspace{1cm} (4.17)

where the first inequality is true because $1-x \geq \exp\left[-\frac{x}{1-x}\right]$ for any $0 < x < 1$; the second inequality is true because $m \leq m^*$ in (4.2) and $p_d = d^{k-1}/n$.

Putting $\delta = \frac{1}{2}$ in Lemma 2.1, by the equations in (4.16) and (4.17), we have

$$\P[\mathcal{F}^c] < \exp\left[-\frac{n}{8} d^{-\frac{(k-1-\epsilon)}{2}}\right] < \exp\left[-\frac{n}{8} d^{-k-\epsilon}\right],$$  \hspace{1cm} (4.18)

where the last inequality is obviously true when $k \geq 2$ and any real number $\epsilon > 0$. The proof of the equation in (4.12) is complete.
Secondly, we will prove the equation in (4.13) to complete the proof of Claim 4.1. Let $S$ be the collection of vertex subsets $S$ in $[n]$ satisfying $|S| > \frac{n}{2}d^{-\frac{(k-1-\epsilon)}{2}}$, which implies

$$S = \left\{ S \subset [n] \mid |S| > \frac{n}{2}d^{-\frac{(k-1-\epsilon)}{2}} \right\}.$$ 

Note that

$$\mathbb{P}[S \text{ is independent}] = (1 - p)^{|S|(|S|-1)/2}.$$ 

Since $p = d/n = o(1)$, $1 - p \sim \exp[-p]$ and $|S| > \frac{n}{2}d^{-\frac{(k-1-\epsilon)}{2}} \rightarrow \infty$ when $d^{k-1} = o(n)$, we have

$$\mathbb{P}[S \text{ is independent}] \sim \exp\left[-\frac{|S|^2 - |S|}{2}p\right] < \exp\left[-\frac{n}{8}d^{-(k-1-\epsilon)}\right], \quad (4.19)$$

where the last inequality is true because

$$\frac{|S|^2 - |S|}{2}p \sim \frac{|S|^2}{2} \cdot \frac{d}{n} > \frac{n}{8}d^{-(k-2-\epsilon)} > \frac{n}{8}d^{-(k-1-\epsilon)}$$

when $|S| \rightarrow \infty$. At last, by the equation in (4.19), we have

$$\mathbb{P}[\mathcal{E} \cap \mathcal{F}] = \mathbb{P}[\mathcal{E} | \mathcal{F}]\mathbb{P}[\mathcal{F}] = \sum_{S \in S} \mathbb{P}[S \text{ is independent}] \cdot \mathbb{P}[\Gamma_{\geq k}(M) = S],$$

then

$$\mathbb{P}[\mathcal{E} \cap \mathcal{F}] < \exp\left[-\frac{n}{8}d^{-(k-1-\epsilon)}\right] \cdot \sum_{S \in S} \mathbb{P}[\Gamma_{\geq k}(M) = S] < \exp\left[-\frac{n}{8}d^{-(k-1-\epsilon)}\right].$$

The proof of the equation in (4.13) is complete. \qed

**Remark 4.1.** Why do we define the event $\mathcal{F}$ to be the equation in (4.10)? In fact, it is not necessary to find the optimal coefficient. The form of $\mathcal{F}$ mainly helps us obtain the same term $\exp[-\frac{n}{8}d^{-(k-1-\epsilon)}]$ in (4.18) and (4.19) to finally prove the equations in (4.12) and (4.13).

### 5 Generation of one large $k$-matching

In this section, by investigating the property for some given number of vertices in $G \in \mathcal{G}(n, p)$, we design a randomized greedy algorithm to generate a $k$-matching with size $(1 + o(1))\frac{kn \log d}{4dp^{k-1}}$ to finish the proof of Theorem 1.3.

Consider “The large $k$-matching Generator” randomized greedy algorithm below.

**Algorithm:** The large $k$-matching Generator

**Input:** A graph $G$ on the vertex $[n]$ with $n \rightarrow \infty$, $k \geq 2$, $d \geq c$ and $d^{k-1} = o(n)$ for some large enough constant $c$. 

16
Output: A large $k$-matching of $G$.

Step 1: Uniformly at random choose $2s$ vertices $S$ from $[n]$. Pair them into $s$ pairs randomly and independently, where $s = \frac{n}{4d^{k-1}}[k \log d - 3\log(k \log d)]$.

Step 2: If $S$ is not a $k$-matching in $G$, replace one pair by an edge in $V(G) \setminus S$ at distance at least $k$ to $S$.

Step 3: Output $S$ if $S$ is a $k$-matching in $G$. Otherwise, go to step 2.

We consider the efficiency of the above algorithm by investigating the property for any given $2s$ vertices in $G \in \mathcal{G}(n, p)$, which is guaranteed by the following Theorem 5.1.

**Theorem 5.1.** Let $k \geq 2$ be a fixed integer, $G \in \mathcal{G}(n, p)$ with the expected degree $d = np$ and $n \to \infty$, where $d \geq c$ and $d^{k-1} = o(n)$ for some large enough constant $c$. W.h.p., there are $\frac{n}{4d^{k-1}}(k \log d)^{3/2}$ vertices at distance at least $k$ to every set of $2s$ vertices, where

$$s = \frac{n}{4d^{k-1}}[k \log d - 3\log(k \log d)]. \quad (5.1)$$

Moreover, these $\frac{n}{4d^{k-1}}(k \log d)^{3/2}$ vertices induce at least one edge in $G$.

**Proof of Theorem 5.1.** We prove Theorem 5.1 by two claims in the following. Let $S \subseteq [n]$ be a vertex subset with $|S| = 2s$. Let $\Gamma_i(S)$ be the set of vertices satisfying

$$\Gamma_i(S) = \{w \in V | d(w, S) = i\}$$

for $0 \leq i \leq k - 2$. Define the event $\mathcal{F}_i$ to be $\mathcal{F}_i = \{|\Gamma_i(S)| \sim 2sd^i\}$ for $0 \leq i \leq k - 2$. Let $\overline{\mathcal{F}_i}$ be the complement of the event $\mathcal{F}_i$, which means that there exists some $\epsilon_i \in (0, 1)$ such that

$$|\Gamma_i(S)| - 2sd^i > \epsilon_i sd^i.$$

**Claim 5.1** Fix a vertex subset $S \subseteq [n]$ with $|S| = 2s$. W.h.p. the events $\mathcal{F}_i$ for $0 \leq i \leq k - 2$ are all true.

**Proof of Claim 5.1.** Note that $\mathcal{F}_0$ is true because $|\Gamma_0(S)| = 2s$. If $k = 2$, we are done. Let $k \geq 3$ below. Assume the events $\mathcal{F}_0, \ldots, \mathcal{F}_{i-1}$ are all true when $1 \leq i \leq k - 2$. Thus, $|\Gamma_i(S)|$ is a binomially distributed random variable with parameters $n - \sum_{j=0}^{i-1} |\Gamma_j(S)|$ and $1 - (1 - p)^{|\Gamma_{i-1}(S)|}$ because each vertex in $[n] - \cup_{j=0}^{i-1}\Gamma_j(S)$ is selected into $\Gamma_i(S)$ if and only if it is connected with at least one vertex in $\Gamma_{i-1}(S)$, which occurs with probability $1 - (1 - p)^{|\Gamma_{i-1}(S)|}$. We have

$$\mathbb{P}[\overline{\mathcal{F}_i} | \mathcal{F}_0, \ldots, \mathcal{F}_{i-1}]$$

$$= \mathbb{P}\left[\left|\text{Bin}\left(n - \sum_{j=0}^{i-1} |\Gamma_j(S)|, 1 - (1 - p)^{|\Gamma_{i-1}(S)|}\right) - 2sd^i\right| > \epsilon_i sd^i\right]. \quad (5.2)$$

Since the events $\mathcal{F}_0, \ldots, \mathcal{F}_{i-1}$ are all true by induction, we have obtained $|\Gamma_j(S)| \sim 2sd^j$ for $0 \leq j \leq i - 1$ and

$$\sum_{j=0}^{i-1} |\Gamma_j(S)| \sim \sum_{j=0}^{i-1} 2sd^j \sim 2sd^{i-1}$$
where $\sum_{j=0}^{i-1} 2sd^j \sim 2sd^{i-1}$ is true because $k$ is fixed and the sum $\sum_{j=0}^{i-1} 2sd^j$ is bounded by an increasing geometric series with common ratio $d$ dominated by the term when $j = i - 1$. By $s = O\left(\frac{n \log d}{d^{k-1}}\right)$ in (5.1), we further have

$$\sum_{j=0}^{i-1} |\Gamma_j(S)| \sim 2sd^{i-1} = o(n),$$

and then

$$n - \sum_{j=0}^{i-1} |\Gamma_j(S)| \sim n. \quad (5.3)$$

On the other hand, since $|\Gamma_{i-1}(S)| \sim 2sd^{i-1}$ by induction, and $s = O\left(\frac{n \log d}{d^{k-1}}\right)$ in (5.1), we also have

$$|\Gamma_{i-1}(S)|p = O\left(\frac{\log d}{d^{k-1-1}}\right) = o(1) \quad (5.4)$$

when $d \geq c$ for some large enough constant $c$ and $0 \leq i \leq k - 2$. By Taylor’s expansion,

$$(1 - p)^{|\Gamma_{i-1}(S)|} = \sum_{j=0}^{\infty} \binom{|\Gamma_{i-1}(S)|}{j} (-p)^j = 1 - |\Gamma_{i-1}(S)|p + O(|\Gamma_{i-1}(S)|^2 p^2),$$

and by the equation in (5.4),

$$1 - (1 - p)^{|\Gamma_{i-1}(S)|} = |\Gamma_{i-1}(S)|p \cdot (1 + O(|\Gamma_{i-1}(S)|p)) \sim 2sd^{i-1}p. \quad (5.5)$$

Combining the equations in (5.2), (5.3) and (5.5), we have

$$\mathbb{P}[\mathcal{F}_i|\mathcal{F}_0, \ldots, \mathcal{F}_{i-1}] \sim \mathbb{P}[\text{Bin}(n, 2sd^{i-1}p) - 2sd^i > \epsilon_i sd^i].$$

Note that $\mathbb{E}[\text{Bin}(n, 2sd^{i-1}p)] = 2sd^i$ by $d = np$. By Lemma 2.1 with $\delta = \epsilon_i$, we have

$$\mathbb{P}[\mathcal{F}_i|\mathcal{F}_0, \ldots, \mathcal{F}_{i-1}] < 2 \exp\left[-\epsilon_i^2 sd^i\right] \to 0 \quad (5.6)$$

because $s = \frac{n}{d^{k-1}}[k \log d - 3 \log(k \log d)]$ in (5.1), $d^{k-1} = o(n)$ and $0 \leq i \leq k - 2$. At last, we have

$$\mathbb{P}[\mathcal{F}_0, \ldots, \mathcal{F}_{k-2}] = \mathbb{P}[\mathcal{F}_0] \prod_{i=1}^{k-2} \mathbb{P}[\mathcal{F}_i|\mathcal{F}_0, \ldots, \mathcal{F}_{i-1}] = \prod_{i=1}^{k-2} \mathbb{P}[\mathcal{F}_i|\mathcal{F}_0, \ldots, \mathcal{F}_{i-1}] \to 1$$

because $\mathbb{P}[\mathcal{F}_0] = 1$ and $k \geq 3$ is a fixed integer.
Claim 5.2 For any $S \subseteq [n]$ such that $|S| = 2s$, w.h.p. there are $\frac{n}{d^{k/2}}(k \log d)^{3/2}$ vertices at distance at least $k$ to $S$. Moreover, these $\frac{n}{d^{k/2}}(k \log d)^{3/2}$ vertices induce at least one edge in $G$.

Proof of Claim 5.2. Firstly, fix a vertex subset $S \subseteq [n]$ with $|S| = 2s$. By Claim 5.1, we have $\Pr[F_0, \ldots, F_{k-2}] \rightarrow 1$. Assume that the events $F_0, \ldots, F_{k-2}$ hold below. Let $A$ be the value satisfying

$$A = \frac{n}{d^{k/2}}(k \log d)^{3/2}. \quad (5.7)$$

Note that $A \rightarrow \infty$ when $n \rightarrow \infty$ because $k \geq 2$ is a fixed integer and $d^{k-1} = o(n)$. Define the event $E_k = \{|\Gamma_{\geq k}(S)| \sim A\}$, where $\Gamma_{\geq k}(S)$ denotes the set of vertices in $G$ whose distances are at least $k$ to each vertex in $S$. Then, the complement $\overline{E}_k$ of $E_k$ means that there exists some $\epsilon \in (0, 1)$ such that

$$||\Gamma_{\geq k}(S)| - A| > \epsilon A.$$ 

Since the events $F_0, \ldots, F_{k-2}$ hold, $|\Gamma_{\geq k}(S)|$ is a binomially distributed random variable with parameters $n - \sum_{i=0}^{k-2} |\Gamma_i(S)|$ and $(1 - p)^{|\Gamma_{k-2}(S)|}$ because each vertex in $[n] - \cup_{i=0}^{k-2} \Gamma_i(S)$ is selected into $\Gamma_{\geq k}(S)$ independently if and only if it is not connected with any vertex in $\Gamma_{k-2}(S)$, which occurs with probability $(1 - p)^{|\Gamma_{k-2}(S)|}$. Thus, we have

$$\Pr[E_k|F_0, \ldots, F_{k-2}] = \Pr\left[\left|\left|\text{Bin}\left(n - \sum_{i=0}^{k-2} |\Gamma_i(S)|, (1 - p)^{|\Gamma_{k-2}(S)|}\right) - A\right| > \epsilon A\right]. \quad (5.8)$$

Under the condition that the events $F_0, \ldots, F_{k-2}$ hold and $k \geq 2$ is a fixed integer, we similarly have $\sum_{i=0}^{k-2} |\Gamma_i(S)| \sim \sum_{i=0}^{k-2} 2sd^i \sim 2sd^{k-2}$ because the sum is dominated by the term when $i = k - 2$. By $s = O(\frac{n \log d}{d^{k/2}})$ in (5.1) and $d \geq c$ for some large enough constant $c$, we also have $\sum_{i=0}^{k-2} |\Gamma_i(S)| = O(\frac{n \log d}{d^{k/2}}) = o(n)$, and then

$$n - \sum_{i=0}^{k-2} |\Gamma_i(S)| \sim n. \quad (5.9)$$

By $p = d/n = o(1), 1 - p \sim \exp[-p]$ and $k \geq 2$ a fixed integer,

$$(1 - p)^{|\Gamma_{k-2}(S)|} \sim \exp\left[-|\Gamma_{k-2}(S)| \cdot p\right] \sim \exp\left[-2sd^{k-2}p\right] = \exp\left[-\frac{1}{2}k \log d + \frac{3}{2} \log(k \log d)\right] = \frac{A}{n}, \quad (5.10)$$


where \(|\Gamma_{k-2}(S)| \sim 2sd^{k-2} = \frac{n}{d^{k-2}}[k \log d - 3 \log(k \log d)]\) is true because the event \(\mathcal{F}_{k-2}\) holds;
and the last equation is true because \(A = \frac{n}{d^{k-2}}(k \log d)^{3/2}\) in (5.7). By Lemma 2.1 with \(\delta = \epsilon\),
the equations in (5.8), (5.9) and (5.10), we have
\[
\mathbb{P}[\mathcal{E}_k | F_0, \cdots, F_{k-2}]
\sim \mathbb{P}[|\text{Bin}(n, A/n) - A| > \epsilon A]
< 2 \exp\left[ -\frac{\epsilon^2 A}{2} \right]
= 2 \exp\left[ -\frac{\epsilon^2 n}{2d^{k/2}}(k \log d)^{3/2} \right]
\rightarrow 0.
\]

It follows that \(\mathbb{P}[\mathcal{E}_k | F_0, \cdots, F_{k-2}] \rightarrow 1\). By Claim 5.1, for a fixed vertex subset \(S \subseteq [n]\) with
\(|S| = 2s\),
\[
\mathbb{P}[\mathcal{F}_0, \cdots, F_{k-2}, \mathcal{E}_k] = \mathbb{P}[\mathcal{E}_k | F_0, \cdots, F_{k-2}] \cdot \mathbb{P}[F_0, \cdots, F_{k-2}] \rightarrow 1. \tag{5.11}
\]

Furthermore, under the condition that the events \(\mathcal{F}_0, \cdots, F_{k-2}, \mathcal{E}_k\) hold, define the event \(K\) to be
\[
K = \{\text{The vertex subset } \Gamma_{\geq k}(S) \text{ induces at least one edge in } G\}.
\]

Since \(|\Gamma_{\geq k}(S)| \sim A, p = o(1)\) and \(1 - p \sim \exp[-p]\), we further have
\[
\mathbb{P}[K | F_0, \cdots, F_{k-2}, \mathcal{E}_k]
\sim (1 - p)^{\binom{A}{2}}
\sim \exp\left[ -\frac{A^2}{2} p \right]
= \exp\left[ -\frac{n}{2d^{k-1}}(k \log d)^3 \right], \tag{5.12}
\]
where the last equality is true by the equation \(A = \frac{n}{d^{k-2}}(k \log d)^{3/2}\) in (5.7).

On the other hand, by the equation in (5.1), the number of the vertex subsets \(S \subseteq [n]\)
with \(|S| = 2s\) is
\[
\binom{n}{2s} \leq \left( \frac{en}{2s} \right)^{2s} = \exp\left[ 2s \log\left( \frac{en}{2s} \right) \right] < \exp\left[ -\frac{n}{2d^{k-1}}(k \log d)^2 \right]. \tag{5.13}
\]

By the equations in (5.12) and (5.13), the union bound of the probability of bad events is
\[
\binom{n}{2s} \mathbb{P}[K^c | F_0, \cdots, F_{k-2}, \mathcal{E}_k]
\leq \exp\left[ -\frac{n}{2d^{k-1}}(k \log d)^2 - \frac{n}{2d^{k-1}}(k \log d)^3 \right]
\rightarrow 0 \tag{5.14}
\]
when \( d \geq c \) for some large enough constant \( c \).

Finally, by the equations in (5.11) and (5.14), for any vertex subset \( S \subseteq [n] \) of size \(|S| = 2s\), w.h.p. there are \( \frac{n}{d^{c/2}}(k \log d)^{3/2} \) vertices at distance at least \( k \) to \( S \), and these \( \frac{n}{d^{c/2}}(k \log d)^{3/2} \) vertices induce at least one edge in \( G \).

We complete the proof of Claim 5.2. \( \square \)

According to Claim 5.1 and Claim 5.2, we complete the proof of Theorem 5.1.

**Remark 5.2.** Using the exactly same argument in Theorem 5.1 and taking \( s = \frac{n}{2d^{c/2}}((k - 1) \log d - 3\log((k - 1) \log d)) \), it also can be shown that every vertex subset \( S \) with \( 2s \) vertices has \( \frac{n}{2d^{c/2}}((k - 1) \log d)^3 \) vertices at distance at least \( k \) to \( S \), instead it is not enough to show these vertices induce at least one edge. We can’t improve Theorem 1.3 through this approach.

**Remark 5.3.** Recently, Cooley et al. in [2] showed that \( um_2(G) = (1 + o(1)) \frac{n \log d}{d} \) by the second moment method and Talagrand’s inequality when \( d \geq c \) and \( d = o(n) \) for some large enough constant \( c \). For \( um_k(G) \) when \( k \geq 3 \), can we apply the similar approach to obtain better results than the one in Theorem 1.1? Unfortunately, it only holds on \( um_2(G) \).

Regard \( G(n, p) = \prod_{i=1}^{n-1} Z_i \) as a product of \( n - 1 \) probability spaces \( Z_i \) for \( i \in [n-1] \), where each \( Z_i \) picks uniformly at random a subset of \([i]\) of size \( \text{Bin}(i, p) \) as the neighbours of the vertex \( i + 1 \) within \([i]\). In order to apply Talagrand’s inequality on the random variable \( um_k(G) \), it should be Lipschitz and \( \theta \)-certifiable for some function \( \theta : \mathbb{N} \to \mathbb{N} \). We say that \( um_k(G) \) is Lipschitz if \(|um_k(G) - um_k(G')| \leq 1 \) for every \( G \) and \( G' \) which differ in at most one coordinate. We say that \( um_k(G) \) is \( \theta \)-certifiable if for any \( G \in G(n, p) \) and \( \xi \in \mathbb{N} \) such that \( um_k(G) \geq \xi \), these exists a set of coordinates \( I \subset [n] \) with \(|I| \leq \theta(\xi) \) such that each \( G' \in G(n, p) \) which agrees with \( G \) on \( I \) also satisfies \( um_k(G') \geq \xi \).

For \( k \geq 3 \), \( um_k(G) \) is not Lipschitz and certifiable. If we change one coordinate of \( G \) such that the chosen vertex adjacent to every vertex of a largest \( k \)-matching in \( G \), then the \( k \)-matching number in \( G' \) possibly is one. Hence, the effect of changing one coordinate of \( G \) is that the size of a considered \( k \)-matching may drop from \( \Theta(\frac{kn \log d}{d^2}) \) to 1, which means that \( um_k(G) \) is not Lipschitz. Similarly, \( um_k(G) \) is also not \( \theta \)-certifiable for \( \theta(x) = 2x \).

Without Talagrand’s inequality, we only obtain a weak lower bound of \( um_k(G) = \Omega(\log n) \) by the second moment method. We show the proof in the Appendix for readers’ reference. It’s an improved version of [7].

### 6 Conclusions

The main results of this paper characterize the value of \( um_k(G) \) for some \( G \in G(n, p) \) and any fixed integer \( k \geq 2 \). The lower bounds here are within a factor two of the upper bound, where the first one is appropriate for any maximal \( k \)-matching and the second one is obtained by a randomized greedy algorithm. The results partially generalize some of the known results from the case when \( k = 2 \) or \( d = c \) for some large enough constant \( c \). It is interesting to consider
the lower bounds of $um_k(G)$ for $G \in \mathcal{G}(n, p)$ when $k \geq 3$ and $p$ lies in broader ranges. We also speculate the upper bound of $um_k(G)$ in Theorem 1.1 for some $G \in \mathcal{G}(n, p)$ is the asymptotic value of $um_k(G)$. The approach of Cooley et al. [2] only holds on $um_2(G)$. We believe it is hard to improve our results for $k \geq 3$ by similar discussions. Note that $d^{k-1} = o(n)$ is a necessary condition in the proof of Lemma 2.4 to finally help us prove Theorem 1.1 and Theorem 1.3. This is the limitation of our approaches. It will take some new ideas and these new problems will be more complicated than the one here. We leave these problems for future work.

References

[1] N. Alon and J. H. Spencer, The Probabilistic Method. Wiley-Interscience Ser. Discrete Math. Optim., John Wiley Sons Inc., Hoboken, NJ, 2008.

[2] O. Cooley, N. Draganić, M. Kang and B. Sudakov, Large induced matchings in random graphs. SIAM J. Discrete Math., 35(1) (2021), 267-280.

[3] A. Czygrinow and B. Nagle, Bounding the strong chromatic index of dense random graphs. Discrete Math., 281(1) (2004), 129-136.

[4] F. Joos, D. Rautenbach and T. Sasse, Induced matchings in subcubic graphs. SIAM J. Discrete Math., 28(1) (2014), 468-473.

[5] R. J. Kang and P. Manggala, Distance edge-colorings and matchings. Discrete Appl. Math., 160 (2012), 2435-2439.

[6] A. El Maftouhi and L. M. Gordones, The maximum order of a strong matching in a random graph. Australas. J. Combin., 10 (1994), 97-104.

[7] F. Tian, On distance matching number in a class of sparse random graphs. Advances in Mathematics(China), 47(2) (2018), 175-181.

Appendix

Without Talagrand’s inequality, we show a weak lower bound of $um_k(G)$ by the second moment method. For a fixed integer $k \geq 3$, define $m \cdot 2^{m+1} = n$, which implies $m = \Theta(\log n)$. Let $\mathcal{M}$ be the set of matchings of size $m$ in $K_n$ and $M_i \in \mathcal{M}$ for $1 \leq i \leq t$ be the matchings in $\mathcal{M}$, where

$$t = \binom{n}{2m} \left( \frac{2m}{2, \ldots, 2} \right) \frac{1}{m!}.$$
Let $I_i$ be the indicator random variable of the event that $M_i$ is a $k$-matching in $G$ and $X_m = \sum_{i=1}^t I_i$. We also have

$$
\mathbb{E}[X_m] = \left( \frac{n}{2m} \right) \left( \frac{2m}{2, \ldots, 2} \right) \frac{1}{m!} \mathbb{P}[M_i \text{ is a } k\text{-matching}]
$$

$$
\sim \frac{n^{2m}}{2^{m!} m!} \mathbb{P}[M_i \text{ is a } k\text{-matching}].
$$

(1)

In order to prove the lower bound of $um_k(G)$, by Chebyshev’s inequality, we will show

$$
\mathbb{P}[um_k(G) < m] \leq \mathbb{P}[X_m = 0] \leq \mathbb{P}[|X_m - \mathbb{E}(X_m)| \geq \mathbb{E}(X_m)] \leq \frac{\mathbb{V}[X_m]}{\mathbb{E}^2(X_m)} \to 0,
$$

then w.h.p. $um_k(G) \geq m$.

Let $M_i, M_j \in \mathcal{M}$. It suffices to consider the case where $M_i$ contains no edges with end-vertices lying in different edges of $M_j$ and vice versa. Define

$$
C_e = \{e \in M_i \cap M_j\} \text{ with } c_e = |C_e|;
$$

$$
C_v = \{v \in M_i \cap M_j \mid v \in e_i \in M_i, v \in e_j \in M_j, e_i \neq e_j\} \text{ with } c_v = |C_v|.
$$

Namely, $C_e$ is the common edge set of $M_i$ and $M_j$, instead $C_v$ is the set of common vertices $v$ in $M_i$ and $M_j$ such that $v$ is incident to two distinct edges, one in $M_i$ and the other in $M_j$. Let $RM_i$ and $RM_j$ denote the rest edges in $M_i - C_e$ and $M_j - C_e$ that are not incident with $C_v$, respectively. Denote $|RM_i| = |RM_j| = m - c_e - c_v = r$, that is, $c_e + c_v + r = m$ (see Fig.1). The two matchings $M_i$ and $M_j$ characterized by the above parameters are called an $(r, c_v, c_e)$-pair. Let $w_{r,c_v,c_e}$ be the number of $(r, c_v, c_e)$-pair $(M_i, M_j)$ for any given nonnegative vector $(r, c_v, c_e)$ satisfying $r + c_e + c_v = m$. Define

$$
\mathcal{D} = \{(r, c_v, c_e) \mid r, c_v, c_e \text{ are nonnegative integers such that } r + c_e + c_v = m\}.
$$

For any given $(r, c_v, c_e) \in \mathcal{D}$ and an $(r, c_v, c_e)$-pair $(M_i, M_j)$, let

$$
\mathbb{E}_{r,c_v,c_e} = w_{r,c_v,c_e} \mathbb{P}[M_i \text{ and } M_j \text{ are } k\text{-matchings in } G].
$$

(2)

Thus,

$$
\mathbb{E}[X_m^2] = \sum_{(r,c_v,c_e) \in \mathcal{D}} \mathbb{E}_{r,c_v,c_e}.
$$

(3)

Claim A.1 For any given $(r, c_v, c_e) \in \mathcal{D}$, we have

$$
w_{r,c_v,c_e} \sim \frac{n^{4m-(2c_e+c_v)}}{(r!)^2 c_e! c_v!} 2^{-2r-c_e}.
$$

23
Proof of Claim A.1. Since $w_{r,c_v,c_e}$ represents the number of $(r,c_v,c_e)$-pair $(M_i,M_j)$, thus

$$w_{r,c_v,c_e} = \left( \binom{n}{2r,2r,2c_e,c_v,c_v,c_v,n-4r-3c_v-2c_e} \cdot \left( \frac{2r}{r!} \right)^2 \right)^2 \left( \frac{2c_e}{c_e!} \right)^2 (c_v!)^2$$

$$= \frac{n!}{(r!)^2 c_v! c_e!} \cdot \frac{(n-4r-3c_v-2c_e)!}{n^{4m-(2c_e+c_v)}} \cdot 2^{-2r-c_e},$$

where the last approximate equality comes from the fact that $4r+3c_v+2c_e = 4m-(2c_e+c_v) \leq 4m = o(n)$ because $m \cdot 2^{m+1} = n$. □

In particular, as $r = m$, then $M_i$ and $M_j$ are independent. By the equation in (1) and Claim A.1, we have

$$\mathbb{E}_{m,0,0} = w_{m,0,0} \mathbb{P}[M_i \text{ and } M_j \text{ are } k\text{-matchings}]$$

$$= w_{m,0,0} \mathbb{P}[M_i \text{ is a } k\text{-matching}]^2$$

$$\sim \mathbb{E}^2[X_m]$$

and by the equation in (3),

$$\mathbb{V}[X_m] = \mathbb{E}[X_m^2] - \mathbb{E}^2[X_m] = \sum_{(r,c_v,c_e) \in \mathcal{D}, 0 \leq r < m} \mathbb{E}_{r,c_v,c_e}.$$  \hspace{1cm} (4)

![Fig.1 The $(r,c_v,c_e)$-pair $M_i$ and $M_j$.]

Claim A.2  For any given $(r,c_v,c_e) \in \mathcal{D}$ with $0 \leq r < m$, let $M_i$ and $M_j$ be an $(r,c_v,c_e)$-pair. Then,

$$\frac{\mathbb{P}[M_i, M_j \text{ are both } k\text{-matchings}]}{\mathbb{P}[M_i \text{ is a } k\text{-matching}]^2} \sim p^{-c_e} (1 - p_d)^{1/2} \left[ 4c_e + c_v - (2c_e + c_v)^2 \right].$$

Proof of Claim A.2. Let $M_i$ and $M_j$ be an $(r,c_v,c_e)$-pair. We have known

$$\mathbb{P}[M_i \text{ is a } k\text{-matching}] = p^m \mathbb{P}[d_G(u,v) \geq k]^{2m(m-1)},$$

thus it suffices to consider $\mathbb{P}[M_i, M_j \text{ are both } k\text{-matchings}]$. First, there exists a penalty factor $p^{2m-c_e}$ to guarantee the edges of $M_i \cup M_j$ present in $G$, and the total number of the paths
which join vertices on any two edges in $M_i$ or any two ones in $M_j$ is exactly

$$2 \left( \binom{2m}{2} \right) - \left( \frac{(2c_e + c_v)}{2} \right)^2 - (2m - c_e)$$

$$= 8 \left( \frac{m}{2} \right) - \left( \frac{(2c_e + c_v)}{2} \right) + c_e$$

$$= 4m^2 - 4m + \frac{1}{2} [4c_e + c_v - (2c_e + c_v)^2],$$

where $2 \left( \binom{2m}{2} \right) - \left( \frac{(2c_e + c_v)}{2} \right)^2$ is the sum of the number of pairs of vertices in $M_i$ and the number of pairs of vertices in $M_j$, and $(2m - c_e)$ is the number of edges in $M_i \cup M_j$. Thus,

$$P[M_i, M_j \text{ are both } k\text{-matchings}] = p^{2m-c_e}P[d_G(u, v) \geq k]\left[4m^2 - 4m + \frac{1}{2} [4c_e + c_v - (2c_e + c_v)^2]\right]. \quad (6)$$

Using the equations in (5) and (6),

$$
\frac{P[M_i, M_j \text{ are both } k\text{-matchings}]}{P[M_i \text{ is a } k\text{-matching}]^2} = p^{-c_e}P[d_G(u, v) \geq k]\left[\frac{1}{2} [4c_e + c_v - (2c_e + c_v)^2]\right]
$$

$$\sim p^{-c_e}(1 - pd)\left[\frac{1}{2} [4c_e + c_v - (2c_e + c_v)^2]\right],$$

where the last approximate equality comes from $P[d_G(u, v) \geq k] \sim 1 - pd$ when $n \to \infty$ because

$$\exp\left[O(n^{k-3}p^{k-2} + n^{2k-4}p^{2k-2})\right] \to 1$$

in Lemma 2.4 (a) when $d^{k-1} = o(n)$. Hence, Claim A.2 holds.

Using the equations in (1), (2), Claim A.1 and Claim A.2, for $0 \leq r < m$, we also have

$$
\frac{\mathbb{E}[c_r c_{-r} c_0]}{\mathbb{E}^2[X_m]} \sim \frac{(m!)^2 n^{-2c_e - c_v} - 2c_e + c_v}{(r!)^2 c_r! c_{-r}!} p^{-c_e}(1 - pd)\left[\frac{1}{2} [4c_e + c_v - (2c_e + c_v)^2]\right]. \quad (7)
$$

**Claim A.3** Let $f(x) = (1 - \frac{2n}{d^{k-1}} \log \frac{2m}{n})x - 4x^2$, then $f(x) \geq f(m) \sim \frac{n(4m+2)}{c^m} \log 2$ for $1 \leq x \leq m$.

**Proof of Claim A.3.** Since $f(x)$ is a symmetric concave downward quadratic function, it is easy to see the maximum point of $f(x)$ occurs at

$$x = \frac{1 - \frac{2n}{d^{k-1}} \log \frac{2m}{n}}{8} = \frac{mn \log 2}{4d^{k-1}} > m$$

when $m \cdot 2^{m+1} = n$. Therefore,

$$f(x) \geq f(1) = \frac{2mn \log 2}{d^{k-1}} - 3 \sim \frac{2mn \log 2}{d^{k-1}}$$

when $1 \leq x \leq m$. 

\[\square\]
Claim A.4 $\mathbb{V}[X_m] = o(\mathbb{E}^2[X_m]).$

Proof of Claim A.4. By the equation in (4), in order to show $\mathbb{V}[X_m] = o(\mathbb{E}^2[X_m])$, it remains to prove that

$$\sum_{(r,c_e,c_v),0 \leq r < m} \frac{\mathbb{E}_{r,c_e,c_v}}{\mathbb{E}^2[X_m]} \to 0.$$  

By the equation in (7) and $m = r + c_e + c_v$, we have

$$\frac{\mathbb{E}_{r,c_e,c_v}}{\mathbb{E}^2(X_m)} \sim \frac{m!}{(r!)^2c_e!c_v!} 2^{2c_e+c_v} p^{-c_e} (1 - p_d)^{\frac{1}{2}[c_e+c_v-(2c_e+c_v)^2]}$$

$$< \frac{m!}{r!c_e!c_v!} 2^{c_v} \left(\frac{(1 - p_d)^{1.5}}{d} \right)^{c_e} \left(\frac{2m}{n} \right)^{c_e+c_v} (1 - p_d)^{\frac{1}{2}[c_e+c_v-4(c_e+c_v)^2]},$$

(8)

where the inequality comes from $\frac{m!}{r!} \leq m^{m-r} = m^{c_e+c_v}$, $d = np$ and

$$(1 - p_d)^{\frac{1}{2}[c_e+c_v-(2c_e+c_v)^2]} < (1 - p_d)^{\frac{1}{2}[c_e+c_v-4(c_e+c_v)^2]}.$$

Note that

$$\left(\frac{2m}{n}\right)^{c_e+c_v} (1 - p_d)^{\frac{1}{2}[c_e+c_v-4(c_e+c_v)^2]}$$

$$= (1 - p_d)^{\frac{1}{2}[c_e+c_v] \left[1+2 \log_2 \frac{2m}{n} - 4(c_e+c_v)\right]}$$

$$= (1 - p_d)^{\frac{1}{2}[c_e+c_v] \left[1+2 \frac{\log 2n}{\log (1-p_d)} - 4(c_e+c_v)\right]}$$

$$\sim (1 - p_d)^{\frac{1}{2}[c_e+c_v] \left[1-\frac{2n}{d} \log \frac{2m}{n} - 4(c_e+c_v)\right]}$$

where the last approximate equality is true because $\log (1 - p_d) \sim -p_d$ when $p_d = o(1)$. By Claim A.3 and $1 \leq c_e + c_v = m - r \leq m$ when $0 \leq r < m$, we further have

$$\frac{\mathbb{E}_{r,c_e,c_v}}{\mathbb{E}^2(X_m)} \leq \frac{m!}{r!c_e!c_v!} 2^{c_v} \left(\frac{(1 - p_d)^{1.5}}{d} \right)^{c_e} (1 - p_d)^{\frac{2m}{dk} \log 2}.$$

At last, it follows that

$$\sum_{(r,c_e,c_v),0 \leq r < m} \frac{\mathbb{E}_{r,c_e,c_v}}{\mathbb{E}^2(X_m)} < (1 - p_d)^{\frac{2m}{dk} \log 2} \sum_{(r,c_e,c_v),0 \leq r < m} \frac{m!}{r!c_e!c_v!} 2^{c_v} \left(\frac{(1 - p_d)^{1.5}}{d} \right)^{c_e}$$

$$< (1 - p_d)^{\frac{2m}{dk} \log 2} \left(3 + \frac{(1 - p_d)^{1.5}}{d}\right)^m$$

$$= \left[(1 - p_d)^{\frac{2m}{dk} \log 2} \left(3 + \frac{(1 - p_d)^{1.5}}{d}\right)\right]^m,$$

(9)

where the second inequality uses the fact that for real number $x_0, \ldots, x_l,$

$$(x_0 + x_1 + \cdots + x_l)^m = \sum_{k_0,\ldots,k_l \geq 0, k_0 + \cdots + k_l = m} \frac{m!}{k_0! \cdots k_l!} x_0^{k_0} \cdots x_l^{k_l}.$$
In (9), since $p_d = d^{k-1}/n = o(1)$, we have

$$\left(1 - p_d\right)^{2n} \log^2 \left(3 + \frac{(1 - p_d)^{1.5}}{d}\right) \sim 3(1 - p_d)^{2n} \log^2 \rightarrow \frac{3}{4},$$

thus $\sum_{(r,c_v,c_e),0 \leq r < m} \frac{E_{r,c_v,c_e}}{E(X_m)} \rightarrow 0$ and Claim A.4 holds. \qed