Simple probabilistic analysis to generalize bottleneck graph multi-partitioning

Fang Tian,*, Zi-Long Liu

* Department of Applied Mathematics, Shanghai University of Finance and Economics, Shanghai, 200433, China
† School of Optical–Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, 200093, China

ABSTRACT

What is the smallest $\Phi(h, k, m)$ such that for any graph $G = (V, E)$ involving $m$ edges and any integer $k \geq 2h$ for $h \geq 1$, there is a partition $V = \bigcup_{i=1}^{h} V_i$ such that the number of edges induced by the union of any $h$ parts is at most $\Phi(h, k, m)$? For $h = 1$ and 2, this coincides with the judicious partitioning problems proposed by Porter (1992) in [1] and by Bollobás and Scott in [8]. Bollobás, A. D. Scott, Problems and results on judicious partitions, Random Structure Algorithms, 21 (2002), 414–430. We show that $(h-1)^m \leq \Phi(h, k, m) \leq (h - \frac{1}{2h-2})^m + O(m^2)$ for general $k \geq 2h$ for $h \geq 2$, and for certain cases $\Phi(2, k, m) \leq 1.5m/k + O(m^2)$ improves on previous results for $h = 2$.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Classical graph partitioning refers to a family of computational problems in which the vertices of a graph have to be partitioned into two (or more) large pieces while minimizing the number of the edges under certain conditions. The ability of finding groups of similar objects (customers, products, cells, words, documents and so on) in large data sets and image segmentation is a useful primitive in the first step of “divide and conquer” algorithms for a wide selection of combinatorial optimization problems, for reducing communication cost and achieving maximal performance, such as in laying out large circuits on silicon chips [2], distributing computation among processors [3] and designing catalog segmentation in microeconomic views of data mining [4]. Unfortunately, most graph-partitioning problems often optimize a single quantity and are NP-hard, which imply that we should not expect efficient methods or algorithms for finding optimal solutions. Therefore, researchers have resorted to various approaches in order to address such problems from the theoretical viewpoint and as regards approximate algorithms [5,6]. Problems of another new type, the judicious partitioning problems, proposed by Bollobás and Scott in [7,8], ask for a partition of a given graph that optimizes several quantities simultaneously; they belong to the multi-objective optimization problems and have been widely investigated in simple graphs and hypergraphs recently [9–15]. Here, we first introduce some notation used in this work.

Let $G = (V, E)$ be a finite simple graph with vertex set $V = V(G)$ and edge set $E = E(G)$. For $S, T \subseteq V(G)$, $G[S]$ denotes the subgraph induced by $S$, and $e(S)$ and $e(S, T)$ denote the numbers of edges with both ends in $S$ and with one end in $S$ and the other in $T$, respectively. For any $v \in V(G)$ and $S \subseteq V(G)$, let $\deg_S(v)$ denote the number of vertices that are adjacent to $v$ in $S$. Thus, $\deg(v)$ denotes the degree of $v$ in $G$, abbreviated as $deg(v)$.

The bottleneck bipartitioning problem, which was also shown to be NP-hard, by Shahrrokhi and Székely in [16], is a special case of a judicious partitioning problem raised by Entringer for finding a bipartition $V_1, V_2$ of $V(G)$ that minimizes...
max\{e(V_1), e(V_2)\}. Bollobás and Scott in [10] obtained the best upper bound for this problem and further extended it to bottleneck \(k\)-partitioning problems. Bollobás and Scott also proposed the paired bottleneck \(k\)-partitioning problem in [8]: that of finding a \(k\)-partition of the vertex set \(V(G) = \bigcup_{j=1}^{k} V_j\) that minimizes \(\max(e(V_i \cup V_j), 1 \leq i < j \leq k)\). For \(k \geq 4\), the paired bottleneck \(k\)-partitioning problem is much more difficult than the bottleneck \(k\)-partitioning problem since in the former case one needs to bound \(\binom{k}{2}\) quantities, while in the latter case one only needs to bound \(k\) quantities. Ma and Yu [11] used a probabilistic method to show that for any graph \(G\) with \(m\) edges, there exists a \(k\)-partition such that \(e(V_i \cup V_j) < 1.6m/k + o(m)\) for \(k \geq 4\), where \(1 \leq i < j \leq k\). Other kinds of judicious partitioning problems are also considered in [9,12,13,17]. In fact, the bottleneck \(k\)-partitioning problem and the paired bottleneck \(k\)-partitioning problem are both subcases of the following problem: what is the smallest \(\Phi(h, k, m)\) such that for any graph \(G\) with \(m\) edges and any integer \(k \geq 2h\) for \(h \geq 1\), there is a \(k\)-partition \(V(G) = \bigcup_{j=1}^{k} V_j\) satisfying \(e(V_i \cup \cdots \cup V_{ih}) \leq \Phi(h, k, m)\) for any \(1 \leq i_1 < \cdots < i_h \leq k\). Likewise, it is more difficult than the above two problems, since \(\binom{k}{h} = \binom{k}{2} \geq k\) for \(k \geq 2h\) for \(h \geq 2\).

In this work, we will prove that \(\frac{(h-1)m}{k-1} \leq \Phi(h, k, m) \leq \frac{2h^2 - 2h - 1}{2h - 2} \frac{m}{k} + O(m^{\frac{2}{k}})\) for any graph \(G\) with \(m\) edges and any integer \(k \geq 2h\) for \(h \geq 2\). The lower bound depends on a constructive structure. The approach for getting the upper bound is a refinement of those of Bollobás and Scott [9] and Ma and Yu [11–13], and proceeds by first partitioning a set of large degree vertices, then establishing a random process in order to partition the remaining vertices, and finally applying a concentration inequality to bound the deviations. The key to our improvement is to pick the tougher probabilities and better iterative process so that the expectation of the process will be in a range that we prefer. Finally, for certain cases we improve on the results of Ma and Yu [11] for \(h = 2\).

The rest of this work is organized as follows. The lemmas and main results are presented in Sections 2 and 3, respectively. Finally, we summarize our conclusions and future research directions in Section 4.

2. Some lemmas

One of the first (and key) methods is known as the Azuma–Hoeffding inequality method. We use the version given in [9,11,13]. For an event \(D\) and a random variable \(Z\) of an arbitrary probability space \((\Omega, \mathcal{F}, \mathbb{P})\), \(\mathbb{P}(D)\) and \(\mathbb{E}(Z)\) denote the probability of \(D\) and expectation of \(Z\), respectively.

**Lemma 2.1.** Let \(g\) be a random variable determined by \(n\) independent trials \(T_i\) for \(1 \leq i \leq n\) taking outcomes from the set \([1, \ldots, k]\). Let \(L: \{1, \ldots, k\}^n \to \mathbb{N}\) be a function satisfying the Lipschitz condition, that is, \(|L(g) - L(g')| \leq l_i\) for any \(g\) and \(g'\) that differ only in the \(i\)th coordinate. Let \(\mu = \mathbb{E}(L(g))\) and for all \(\lambda > 0\),

\[
\mathbb{P}(L(g) \geq \mu + \lambda) \leq e^{-\lambda^2 / \left(2 \sum_{i=1}^{n} \frac{l_i^2}{f_i}\right)} \quad \text{and} \quad \mathbb{P}(L(g) \leq \mu - \lambda) \leq e^{-\lambda^2 / \left(2 \sum_{i=1}^{n} \frac{l_i^2}{f_i}\right)}.
\]

**Lemma 2.2.** Let \(k = 2h\) for \(h \geq 2\) and \(\phi(x_1, \ldots, x_h) = \left(\Sigma_{j=1}^{h} c_j\right) \cdot \left(\Sigma_{j=1}^{h} x_j\right)\), where \(1 \leq i_1 < \cdots < i_h \leq k\). For \(1 \leq i \leq k\), if \(c_i \geq 0\) and \(C = \sum_{j=1}^{k} c_j > 0\), then there exists \(p_i \in [0, \frac{1}{k-1}]\) such that \(\sum_{i=1}^{k} p_i = 1\) and \(\phi(p_1, \ldots, p_h) \leq \frac{k-1}{2k-2} C\) for any \(1 \leq i_1 < \cdots < i_h \leq k\).

**Proof.** Let \(p_i = \frac{1}{k-1} - \frac{c_i}{(k-1)C}\) for \(1 \leq i \leq k\), so we have

\[
\phi(p_1, \ldots, p_h) = \left(\sum_{j=1}^{h} c_j\right) \left(\frac{k}{2k-2} - \frac{\sum_{j=1}^{h} c_j}{(k-1)C}\right) = -\frac{1}{(k-1)C} \left(\sum_{j=1}^{h} c_j\right) \left(\sum_{j=1}^{h} c_j - \frac{k}{2}\right) + \frac{k^2 C}{16(k-1)}.
\]

For \(h = 2\) and \(k = 4\), we have that the maximum value of \(\phi(p_1, \ldots, p_h)\) is \(\frac{k^2 C}{16(k-1)}\) as \(\sum_{j=1}^{h} c_j = \frac{kC}{4} = C\). Likewise, for \(k > 4\), the maximum value of \(\phi(p_1, \ldots, p_h)\) occurs at \(\sum_{j=1}^{h} c_j = C\), that is \(-\frac{1}{(k-1)C} \left(C - \frac{kC}{4}\right) + \frac{k^2 C}{16(k-1)} = \frac{k-2}{2k-2} C\). □

**Lemma 2.3.** Let \(f(x) = \frac{h x - h x^2}{k \cdot (k-1)}\) for \(k \geq 2h\), for \(h \geq 2\), where \(x \in [\frac{k(h-1)}{h(k-1)}, 1]\). Then \(f(x) \leq \frac{h-1}{h(k-1)}\).

**Proof.** By differentiating \(f(x)\), we obtain

\[
f'(x) = \frac{h^2 x^2 - 2hx + kh}{(k-1)x^2} < 0.
\]
if and only if \( x \in \left(\frac{k - \sqrt{k(k-1)}}{h}, \frac{k + \sqrt{k(k-1)}}{h}\right) \). Clearly, \( \frac{k + \sqrt{k(k-1)}}{h} > 1 \); thus we only need to verify that \( \frac{k - \sqrt{k(k-1)}}{h} < \frac{k(h-1)}{h(k-1)} \) by rationalizing the numerator, and this is further reduced to \( \sqrt{k} - \sqrt{h} < \sqrt{k(h-1)} \), which is always right for \( h \geq 2 \). Hence, \( f(x) \leq f\left(\frac{k(h-1)}{h(k-1)}\right) = \frac{h-1}{h(k-1)} \) for \( x \in \left[\frac{k(h-1)}{h(k-1)}, 1\right) \). □

Lemma 2.4. Let \( \chi_1(h) = 4h^6 - 2h^5 - 16h^4 + 10h^3 + 5h^2 \), \( \chi_2(h) = 40h^7 - 16h^6 - 188h^5 + 132h^4 + 48h^3 - 4h^2 \), \( \chi_3(h) = 32h^9 - 12h^8 - 164h^7 + 104h^6 + 76h^5 - 24h^4 - 3h^2 \) and \( \chi_4(h) = 16h^9 - 8h^8 - 80h^7 + 28h^6 + 84h^5 - 30h^4 - 2h^3 - 2h^2 \); thus \( \chi_1(h) > 0, \chi_2(h) > 0 \) and \( \chi_3(h) > 0 \) for \( h \geq 2 \), while \( \chi_4(h) > 0 \) for \( h \geq 3 \).

Proof. Since \( \chi_1^{(5)}(h) = 2880h - 240 > 0 \) for \( h \geq 1 \), it follows that \( \chi_1^{(4)}(h) = 1440h^2 - 240h - 384 > \chi_1^{(4)}(1) > 0 \) for \( h \geq 1 \); then \( \chi_1^{(3)}(h) = 480h^3 - 120h^2 - 384h + 60 > \chi_1^{(3)}(1) > 0 \) for \( h \geq 1 \), and thus \( \chi_1^{(1)}(h) = 120h^4 - 40h^3 - 192h^2 + 60h + 10 \geq \chi_1'(2) > 0 \) for \( h \geq 2 \), and similarly, \( \chi_1'(h) = 24h^5 - 10h^4 - 64h^3 + 30h^2 + 10h \geq \chi_1'(2) > 0 \) for \( h \geq 2 \). Finally, \( \chi_1(h) \geq \chi_1(2) > 0 \) for \( h \geq 2 \). For \( \chi_2(h), \chi_3(h) \) and \( \chi_4(h) \), the approach adopted is the same as for \( \chi_1(h) \). □

Lemma 2.5. Given fixed \( h \geq 2 \) and \( k \geq 2h + 1 \), let \( \Delta_h = h - \frac{h - 1}{2h-1} \); then there exist \( s_{h,k} \in [0, \frac{1}{h}] \) and \( \Delta_{h,k} \in \left[\frac{h-1}{k-1}, h\right] \) such that

\[
\Delta_{h,k} - \frac{h(hk-k-4)}{k(k^2-4)} s_{h,k}.
\]

Furthermore, \( \Delta_{h,k} < \frac{2h^2-2h+1}{2h-1-k} \) for \( k \geq 2h \).

Proof. Fix \( h \geq 2 \). Let \( \psi_1(s) = \frac{h-1}{k-s} \) and \( \psi_2(s) = \frac{k-1}{k-s} + \left[\frac{h(hk-k-4)}{k(k^2-4)} + s\right] \) for some \( k \geq 2h + 1 \), which can be guaranteed by \( \Delta_h = \frac{h-1}{2h-1} \).

Firstly, \( \psi_2(0) = \frac{k-1}{k+2} - \Delta_{h,k} \leq \frac{k-1}{k+2} < \frac{h}{k} = \psi_1(0) \) and

\[
\psi_2\left(\frac{1}{h}\right) = \frac{k-h}{k+2} \Delta_{h,k} - \frac{h(hk-k-4)}{k(k^2-4)} \geq \frac{k-h}{k+2} \Delta_{h,k} - \frac{h(hk-k-4)}{k(k^2-4)} \geq \frac{k^2(h-1) + 2hk(k-2)}{k^2(k^2-4)}\]

\[
The second step is clear that \( \psi_1(s) \) and \( \psi_2(s) \) are decreasing and increasing on \( s \), respectively. Thus there must exist \( s_{h,k} \in [0, \frac{1}{h}] \) such that \( \psi_1(s_{h,k}) = \psi_2(s_{h,k}) \) since both \( \psi_1(s) \) and \( \psi_2(s) \) are continuous over \( [0, \frac{1}{h}] \).

Let \( \Delta_{h,k} = \psi_1(s_{h,k}) = \psi_2(s_{h,k}) \); thus we have \( \Delta_{h,k} = \frac{h(hk-k-4)}{k^2(k^2-4)} s_{h,k} \) for all \( k \geq 2h + 1 \) since \( s_{h,k} \in [0, \frac{1}{h}] \). Let \( \Delta_{h,k} = M_{h,k} h/k \); thus \( M_{h,k} \in \left[\frac{(h-1)(h-1)}{h(k-1)}, 1\right] \). By the first part of the lemma, we have

\[
M_{h,k} = \frac{k-k_{h,k}}{h} \Delta_{h,k} = \frac{k(h-1)}{(k-1)(k+2)} h_{h,k,k} + \frac{h(hk-k-4)}{k^2(k^2-4)} s_{h,k}.
\]

Since \( M_{h,k} = \frac{k-k_{h,k}}{h} \), we thus have \( s_{h,k} = \frac{k-k_{h,k}}{h} M_{h,k} \) and

\[
M_{h,k} = \frac{k(k-h-1)}{(k-1)(k+2)} h_{h,k,k} + \frac{h(hk-k-4)}{k^2(k^2-4)} s_{h,k}.
\]

Now, we will prove \( M_{h,k} < \frac{2h^2-2h+1}{2h-1-k} \) for all \( k \geq 2h \). Since \( \Delta_{h,2h} = \frac{h-1}{2h-1} \), we have \( M_{h,2h} = \frac{h-2}{2h-1} = \frac{2h^2-2h+1}{2h-1-k} \) for \( h \geq 2 \).

Case 1. If \( h = 2 \) and \( k = 2h + 1 = 5 \), then \( M_{2,5} = \frac{5}{21} + \frac{45-45M_{2,5}}{35-14M_{2,5}} \) by \( M_{2,4} = \frac{2}{3} \) and (2), which is reduced to \( M_{2,5} < \frac{125-1805}{42} \approx 0.7294 < 0.75 \).

Case 2. For \( k \geq 6 \) for \( h = 2 \) or \( k \geq 2h + 1 \) for \( h \geq 3 \), we will prove \( M_{h,k} < \frac{2h^2-2h+1}{2h-1-k} \) by induction on \( k \). Suppose that \( M_{h,k-1} < \frac{2h^2-2h+1}{2h-1-k} \), which can be guaranteed by \( M_{2,5} < 0.75 \) in Case 1 for \( h = 2 \) and \( M_{h,2h} = \frac{h-2}{2h-1} < \frac{2h^2-2h+1}{2h-1-k} \) for \( h \geq 3 \);
then (2) is changed into
\begin{equation}
M_{h,k} < \frac{k(k-h-1)2h^2 - 2h + 1 + h(k-2)2h^2 - 2h + 1 + h(k-1)(h(k + k - 4))}{(k-1)(k-2) - 4} \cdot \frac{k - kM_{h,k}}{k - hM_{h,k}}.
\end{equation}

Let $A = 2h^2 - 2h$ and $B = (A - 1)hk(k-2) + Ah(k-1)(hk + k - 4)$; thus the above inequality is transformed into
\begin{equation}
A(k-1)(k^2 - 4)M_{h,k} < (A - 1)k(k-h-1)(k-2) + B(1 - M_{h,k}) + B \frac{hM_{h,k}(1 - M_{h,k})}{k - hM_{h,k}}
\end{equation}

by multiplying by $A(k-1)(k^2 - 4)$ on both sides, and the last inequality in (4) is derived by using Lemma 2.3. Therefore, $M_{h,k} < [(A - 1)k(k-h-1)(k-2) + B + B \frac{h-1}{h(k-1)}]/[A(k-1)(k^2 - 4) + B]$.

The last step in our proof is to show that
\begin{equation}
\frac{(A - 1)k(k-h-1)(k-2) + B + B \frac{h-1}{h(k-1)}}{A(k-1)(k^2 - 4) + B} \leq \frac{A-1}{A}
\end{equation}

for $k \geq 6$ for $h = 2$ or $k \geq 2h+1$ for $h \geq 3$, which is reduced by Matlab 7.0 to $a(h)k^3 - b(h)k^2 + c(h)k - d(h) \geq 0$, where
\begin{align*}
a(h) &= 4h^2 - 2h^2 - 16h^3 + 10h^5 + 5h^6, \\
b(h) &= 4h^2 + 8h^5 - 8h^3 - 54h^4 + 36h^5 + 17h^6, \\
c(h) &= 4h^2 + 24h^5 - 50h^3 - 20h^4 + 26h^5 + 18h^6, \\
d(h) &= 16h^5 - 32h^5 - 8h^6.
\end{align*}

Let $\sigma(k) = a(h)k^3 - b(h)k^2 + c(h)k - d(h)$. On the basis of Lemma 2.4, since $\sigma^{(3)}(k) = 6a(h) = 6 \chi_1(h) > 0$ for $h \geq 2$, it follows that $\sigma^{(k)}(k) = 6a(h)k^3 - 2b(h)k^2 \geq \sigma'^{(2)}(2h+1) = \chi_3(h) > 0$ for $k \geq 2h+1$, for $h \geq 2$; likewise, $\sigma'(k) = 3a(h)k^2 - 2b(h)k + c(h) \geq \sigma'(2h+1) = \chi_3(h) > 0$ for $k \geq 2h+1$, for $h \geq 2$; finally, $\sigma(k) \geq \sigma(6) = 216a(2) - 36b(2) + 6c(2) - d(2) = 704 > 0$ for $k \geq 6$, for $h = 2$, and $\sigma(k) \geq \sigma(2h+1) = \chi_4(h) > 0$ for $k \geq 2h+1$, for $h \geq 3$.

**Lemma 2.6.** Let $k \geq 2h$ for $h \geq 2$ and $\varphi(x_1, \ldots, x_h) = (\Sigma_{j=1}^h c_j \cdot (\Sigma_{j=1}^h x_j)$, where $1 \leq i_1 < \cdots < i_h \leq k$. For $1 \leq i \leq k$, if $c_i \geq 0$ and $C = \Sigma_{i=1}^k c_i > 0$, then there exists $p_i \in [0, \frac{1}{k+1}]$ such that $\Sigma_{i=1}^k p_i = 1$ and $\varphi(p_{i_1}, \ldots, p_{i_h}) \leq \Delta_{h,k} C$ for any $1 \leq i_1 < \cdots < i_h \leq k$, where $\Delta_{h,2h} = \frac{h-1}{2h-1}$, $\Delta_{h,k}$ is defined in (1) and $\Delta_{h,k} < \frac{2h^2 - 2h - 1}{(2h-1)^2}$ for $k \geq 2h+1$.

**Proof.** Fix $h \geq 2$, then we prove this lemma by induction on $k$. If $k = 2h$, the lemma is proved by Lemma 2.2. Now, assume that $k \geq 2h+1$.

**Case 1.** Assume that $c_i \leq s C$ for all $1 \leq i \leq k$, where $s$ will be determined later. For any $r \in \{1, \ldots, k\}$, by the induction hypothesis, there must exist $p_i^r \in [0, \frac{1}{k+1}]$ such that $\Sigma_{i=1}^k p_i^r = 1$ and $\varphi(p_{i_1}^r, \ldots, p_{i_h}^r) \leq (\Sigma_{i=1}^k c_i) \cdot (\Sigma_{i=1}^k p_i^r) \leq \Delta_{h,k-1} (C - c_r)$ for any $\{i_1, \ldots, i_h\} \subseteq \{1, \ldots, k\} \setminus \{r\}$.

Let $p_i = \frac{1}{k+1} \left(\Sigma_{r=1}^k (\Sigma_{i=1}^k (\Sigma_{j=1}^h p_{i_j}^r + \frac{2h}{k+1} \right)$. We have $0 \leq p_i \leq \frac{1}{k+1} \cdot \left(\frac{k+1}{k+2} + \frac{2h}{k+1}\right) < \frac{1}{k+1}$ for $k \geq 5$ and $\Sigma_{i=1}^k p_i = \frac{1}{k+1} \Sigma_{i=1}^k (\Sigma_{r=1}^k (\Sigma_{j=1}^h p_{i_j}^r + \frac{2h}{k+1} \right) = \frac{1}{k+1} (\Sigma_{j=1}^k (1 + 2h)) = 1$. Now let $S = \{i_1, \ldots, i_h\}$ for any $1 \leq i_1 < \cdots < i_h \leq k$; thus
\begin{align*}
\varphi(p_{i_1}, \ldots, p_{i_h}) &= \left(\Sigma_{j=1}^h c_j \right) \cdot \left(\Sigma_{j=1}^h p_{i_j} \right) = \frac{1}{k+1} \left(\Sigma_{j=1}^h c_j \right) \left(\Sigma_{j=1}^h p_{i_j} \right) \left(\frac{k+1}{k+2} + \frac{2h}{k+1}\right) \\
&\leq \frac{\Delta_{h,k-1}}{k+2} \sum_{r=1}^k (C - c_r) + \frac{1}{k+2} \left(\sum_{j=1}^h c_j \right) \left(\frac{h(h-1)}{k-2} + \frac{2h}{k}\right) \\
&\leq \frac{\Delta_{h,k-1}}{k+2} \left(\frac{h(h-1)}{k-2} + \frac{2h}{k}\right) (\Sigma_{j=1}^h c_j) \\
&\leq \left[\frac{k-h-1}{k+2} - \Delta_{h,k-1} + \left(\frac{\Delta_{h,k-1}}{k+2} \left(\frac{h(h-1)}{k-2} + \frac{2h}{k}\right) \right) C\right].
\end{align*}
Case 2. Assume that $c_i > sc$ for some $1 \leq i \leq k$. Without loss of generality, let $c_1 > sc$. Let $p_i = x$ for $2 \leq i \leq k$ for $0 \leq x < \frac{1}{k-1}$ and $p_1 = 1 - (k - 1)x$, where $x$ will be determined later. For $2 \leq i_1 < \cdots < i_h \leq k$, $\varphi(p_1, \ldots, p_h) = \left( \sum_{j=1}^{h} c_j \right) \cdot \left( \sum_{j=1}^{h} p_j \right) < (c - sc)hx = (1 - s)hx$. For $i_1 = 1$ and $2 \leq i_2 < \cdots < i_h \leq k$, $\varphi(p_1, p_{i_2}, \ldots, p_{i_h}) = (c_1 + \sum_{j=2}^{i_h} c_j) (p_1 + \sum_{j=2}^{i_h} p_j) \leq \left[ 1 - (k - 1)x \right] x + (h - 1)x \] \leq \left[ 1 - (k - h)x \right] x$. 

Now, we wish to minimize $\max \left\{ \left[ 1 - (k - h)x \right], (1 - s)hx \right\}$. Since $\left[ 1 - (k - h)x \right]$ is decreasing on $x$ and $(1 - s)hx$ is increasing on $x$, the minimum of $\max \left\{ \left[ 1 - (k - h)x \right], (1 - s)hx \right\}$ occurs as $1 - (k - h)x = (1 - s)hx$, that is, $x = \frac{1}{k-h}$. Thus, for any $1 \leq i_1 < \cdots < i_h \leq k$,

$$\varphi(p_1, p_{i_2}, \ldots, p_{i_h}) \leq \frac{h - hs}{k - hs} c.$$ 

Here, since $0 \leq x < \frac{1}{k-1}$ and $x = \frac{1}{k-1}$, we have $0 \leq s \leq \frac{1}{k}$ and $p_1 = 1 - (k - 1) \frac{1}{k-h} = \frac{1-hs}{k-h} \leq \frac{1}{k-1}$ because $\frac{1-hs}{k-h}$ is a decreasing function on $s$ and $p_1 = x = \frac{1}{k-1}$ for $2 \leq i \leq k$.

By Lemma 2.5, there exist $s_{h,k} \in [0, \frac{1}{k})$ and $\Delta_{h,k} \in \left[ \frac{1-hs}{k-h}, \frac{h}{k} \right]$ for $k \geq 2h + 1$ as $h \geq 2$, such that

$$\Delta_{h,k} = \frac{h(1 - s_{h,k})}{k - hs_{h,k}} = \frac{k - h - 1}{k + 2} \Delta_{h,k-1} + \left[ \frac{h\Delta_{h,k-1}}{k + 2} + \left( k\Delta_{h,k-1} + (kh + k - 4)h^2 \right) \right] s_{h,k}.$$ 

Furthermore, $\Delta_{h,k} < 2\frac{k^2 - 2h - 1}{2(h-1)k}$ for $k \geq 2h + 1$. $\square$

3. The main results

**Theorem 3.1.** For any integer $k \geq 2h$ for $h \geq 2$ and any graph $G$ with $m$ edges, let $\Phi(h, k, m)$ be the smallest integer such that there is a $k$-partition $V(G) = \bigcup_{i=1}^{k} V_i$ satisfying $\deg(V_1) \cup \cdots \cup V_h) \leq \Phi(h, k, m)$ for any $1 \leq i_1 < \cdots < i_h \leq k$. Then $\Phi(h, k, m) \leq \Delta_{h,k}m + O(m^{\frac{3}{2}})$, where $\Delta_{h,k}$ is defined in (1) for $k \geq 2h + 1$, and $\Delta_{h,k} < 2\frac{k^2 - 2h - 1}{2(h-1)k}$ for $k \geq 2h$.

**Proof.** Suppose $G$ is a connected graph with $n$ vertices; otherwise, consider any one of its components. Arrange the vertex set of $G$ as $V(G) = \{v_1, \ldots, v_n\}$ such that $\deg(v_1) \geq \deg(v_2) \geq \cdots \geq \deg(v_n)$. Let $V_1 = \{v_1, \ldots, v_a\}$, where $a = \lceil m^n \rceil$ and $0 < b < \frac{1}{2}$ will be determined later. Clearly, we have $a < n$ for $m < n^{\frac{3}{2}}$. Since $(a + 1)\deg(v_{a+1}) \leq \sum_{i=1}^{a+1} \deg(v_i) \leq 2m$, we have $\deg(v_{a+1}) < 2m^{\frac{3}{2}}$. Let $V_2 = V(G) - V_1$ and schedule the vertices of $V_2$ in sequence as $[v_1, \ldots, v_{n-a}]$ such that $\deg(v_{a+1}) < 2m^{1/2}$. Let $p_i$ be the color of $v_i$ for $1 \leq i \leq n - t$. The process is guaranteed by $G$ being connected.

Let $P_1, P_2, \ldots, P_h$ be any $k$-partition of $V_1$ such that $V_1 = \bigcup_{i=1}^{h} P_i$ and each vertex in $P_i$ is colored with $t$ for $1 \leq i \leq k$. For each vertex $u_i \in V_2$ with $1 \leq s \leq n - a$, color it with $t$ if $1 \leq s \leq k$ with probability $p_i$ such that $\sum_{i=1}^{h} p_i = 1$, where $p_i$ will be determined later. Let the random variable $X_i = 1$ denote the event that the vertex $u_i$ is colored with $t$.

Let $g_0 = 0; g_s = G(V_1, V_2 \cup [v_1, \ldots, v_s])$ for $1 \leq s \leq n - a$; and $\eta_t = \deg_{g_s}(u_i)$ the number of vertices in $A_{s-i}^{t}$ adjacent to $u_i$ for $1 \leq s \leq n - a$ and $1 \leq t \leq k$.

For any $1 \leq t_1 < \cdots < t_h \leq k$, let $b_{i_1, \cdots, i_h} = \sum_{i=1}^{h} \deg_{g_s}(u_i)$ and $\sum_{i=1}^{h} \deg_{g_s}(u_i)$ for $1 \leq s \leq n - a$. Thus, $b_{i_1, \cdots, i_h} = \sum_{i=1}^{h} \deg_{g_s}(u_i) + \sum_{i=1}^{h} (\sum_{i=1}^{h} \deg_{g_s}(u_i))$ and the conditional expectation of $\sum_{i=1}^{h} \deg_{g_s}(u_i)$ under a given random vector $T_i = (T_1, \ldots, T_s) \subseteq \{1, \ldots, k\}$ is

$$E(\sum_{i=1}^{h} \deg_{g_s}(u_i) | T_{s-1}) = \left( \sum_{i=1}^{h} \sum_{j=1}^{h} \eta_{j,i} \right) \cdot \left( \sum_{i=1}^{h} p_i \right).$$

By the additive property of conditional expectation,

$$E(\sum_{i=1}^{h} \deg_{g_s}(u_i) | T_{s-1}) = \sum_{i=1}^{h} E(\sum_{j=1}^{h} \deg_{g_s}(u_i) | T_{s-1}) \cdot P(T_{s-1})$$

$$= \sum_{i=1}^{h} \left( \sum_{j=1}^{h} \eta_{j,i} \right) \cdot \left( \sum_{i=1}^{h} p_i \right) \cdot P(T_{s-1})$$

$$= \sum_{i=1}^{h} \left( \sum_{j=1}^{h} \eta_{j,i} \cdot P(T_{s-1}) \right) \cdot \left( \sum_{i=1}^{h} p_i \right) \cdot P(T_{s-1})$$

$$= \sum_{i=1}^{h} \left( \sum_{j=1}^{h} \eta_{j,i} \cdot P(T_{s-1}) \right) \cdot \left( \sum_{i=1}^{h} p_i \right).$$
Let $c_i = \sum_{s=1}^{T_{s-1}} \eta_i \cdot P(T_{s-1})$ for $1 \leq i \leq k$; then $c_i \geq 0$, $\Sigma_{i=1}^{k} c_i = \Sigma_{i=1}^{k} (\sum_{s=1}^{T_{s-1}} \eta_i \cdot P(T_{s-1})) = \sum_{s=1}^{T_{s-1}} \eta_i \cdot P(T_{s-1}) = \sum_{s=1}^{T_{s-1}} \eta_i \cdot \text{deg}_{\hat{g}_{s-1}}(u_i) = \text{deg}_{\hat{g}_{s-1}}(u_i) > 0$ and $E(\nabla \hat{\beta}_{12}^{i-1}) = \left( \frac{\Sigma_{i=1}^{k} c_i}{\nabla \hat{\beta}_{12}^{i-1}} \right) \cdot \left( \frac{\Sigma_{i=1}^{h} p_i}{\nabla \hat{\beta}_{12}^{i-1}} \right)$. By Lemma 2.6, there exists $p_i \in [0, \frac{1}{\lambda}]$ such that $\Sigma_{i=1}^{k} p_i = 1$ and for any $1 \leq i_1 < \cdots < i_h \leq k$ and $1 \leq s \leq n - a$,

$$E\left( \nabla \hat{\beta}_{12}^{i-1} \right) \leq \Delta_{n,k} \sum_{i=1}^{k} c_i = \Delta_{n,k} \text{deg}_{\hat{g}_{s-1}}(u_i),$$

where $\Delta_{n,k} = \frac{n - 2h}{2h - 1}$, $\Delta_{n,k}$ is defined in (1) for $k \geq 2h + 1$, and $\Delta_{n,k} < \frac{2h^2 - 2h - 1}{2h - 1}k$ for $k \geq 2h$.

By $\beta_{12}^{n-a} = \beta_{12}^{n-a} + \sum_{s=1}^{n-a} \left( \frac{\nabla \hat{\beta}_{12}^{m-1}}{\nabla \hat{\beta}_{12}^{m-1}} \right)$ and the linear property of expectation,

$$E\left( \beta_{12}^{n-a} \right) = \beta_{12}^{n-a} + \sum_{s=1}^{n-a} \left( E\left( \nabla \hat{\beta}_{12}^{m-1} \right) \right) \leq e(V_1) + \Delta_{n,k} \sum_{s=1}^{n-a} \text{deg}_{\hat{g}_{s-1}}(u_i) = e(V_1) + \Delta_{n,k} (m - e(V_1)) = \Delta_{n,k} m + (1 - \Delta_{n,k}) e(V_1) < \Delta_{n,k} m + \left( 1 - \Delta_{n,k} \right) \frac{1}{2} a^2 < \Delta_{n,k} m + \frac{1}{2} \left( 1 - \Delta_{n,k} \right) m^{2}. $$

Since $\beta_{12}^{n-a}$ is a function from $(T_1, \ldots, T_{n-a}) \subseteq \{1, \ldots, k\}^{n-a}$ to the positive integers, and since changing the color of $u_s$ for $1 \leq s \leq n - a$, that is, changing the value of $\hat{g}_s$ affects $\beta_{12}^{n-a}$ by at most $\text{deg}(u_s)$, by Lemma 2.1, for all $\lambda > 0$,

$$P\left( \beta_{12}^{n-a} \geq E\left( \beta_{12}^{n-a} \right) + \lambda \right) \leq e^{-\lambda^2 / \left( 2 \sum_{i=1}^{n-a} \text{deg}(u_i)^2 \right)} \leq e^{-\lambda^2 / \left( 2 \sum_{i=1}^{n-a} \text{deg}(V_{i+1}) \right)} < e^{-\lambda^2 / (2n - 2m^{1-b})} = e^{-\lambda^2 / (2n - 2m^{1-b})}.$$ 

Take $\lambda = \sqrt{\frac{1}{2 \ln \left( \frac{k}{n} \right)} \cdot m^{1-b/2}}$, so we have $P\left( \beta_{12}^{n-a} \geq E\left( \beta_{12}^{n-a} \right) + \lambda \right) < 1 / \left( \frac{k}{n} \right)$. Hence, for any $1 \leq i_1 < \cdots < i_h \leq k$, there exists a $k$-partition $V(G) = \bigcup_{i=1}^{k} V_i$ satisfying

$$e(V_1 \cup \cdots \cup V_h) \leq \nabla \beta_{12}^{n-a} + \lambda < \Delta_{n,k} m + \frac{1}{2} \left( 1 - \Delta_{n,k} \right) m^{2b} + \sqrt{\ln \left( \frac{k}{n} \right)} \cdot m^{1-b/2} \leq \Delta_{n,k} m + O(m^{2b}).$$

where the last inequality comes from the minimum of $\max(m^{2b}, m^{1-b/2})$ occurring at $b = \frac{2}{5}$. 

**Remark 3.2.** The following example shows that $\Phi(h, k, m) \geq \frac{(h-1)m}{k}$, which is approximately equal to the upper bound $(h - \frac{1}{2h - 1}) \frac{n}{k}$ in Theorem 3.1. Take the star graph $K_{1,n}$ whose vertex with degree $n$ is denoted by $a$ and any $k$-partition $V(K_{1,n}) = \bigcup_{i=1}^{k} V_i$ such that $a \in V_k$ and $|V_1| \geq |V_2| \geq \cdots \geq |V_{n-1}|$. Thus $|V_1 \cup V_2 \cup \cdots \cup V_{n-1}| \geq \frac{(h-1)(n+1)-|V_k|}{k-1}$. Hence, $e(V_1 \cup \cdots \cup V_{n-1}) \geq |V_1 \cup V_2 \cup \cdots \cup V_{n-1}| + |V_k| - 1 \geq \frac{(h-1)(n+1)-|V_k|}{k-1} + \frac{(h-1)m}{k-1} \geq \frac{(h-1)m}{k-1}$. 

**Remark 3.3.** In fact, the general framework in Theorem 3.1 is a refinement of that of Bollobás and Scott [9] and Ma and Yu [11,13], where the key to our improvement is to pick a tougher probability structure and design a better iterative process in Section 2. For $h = 2$, our result is $\Phi(2, k, m) \leq 1.5m/k + O(m^{2/5})$. 

4. Conclusions and future research

This work considered a paired bottleneck $k$-partitioning problem in a general format and derived the following conclusion: for any graph $G = (V, E)$ involving $m$ edges and any integer $k \geq 2h$ for $h \geq 2$, there is a partition $V = \bigcup_{i=1}^{k} V_i$ such that for any $1 \leq i_1 < i_2 < \cdots < i_h \leq k$, $e(V_{i_1} \cup \cdots \cup V_{i_h}) \leq \Phi(h, k, m)$, where $(h - 1) \frac{m}{k-1} \leq \Phi(h, k, m) \leq (h - \frac{1}{k}) \frac{m}{k} + O(m^2)$. This is like the conjectures proposed by Bollobás and Scott in [8], and some of them are considered in the affirmative in [11,14,15]; as a next step, we would like to further investigate some simultaneous bounds for generalized paired bottleneck $k$-partitioning.

Acknowledgments

The authors would like to thank the anonymous referees very much for valuable suggestions, comments and corrections which resulted in the improved final manuscript.

References