1 PROOFS

Here we present the proofs of the lemmas in the main paper. They are restated for ease of reference. WLOG is used as the abbreviation for without loss of generality. We use 2 as the base for the logarithm.

Lemma 1: The entropy rate of the random walk on the graph $\mathcal{H}: 2^E \rightarrow \mathbb{R}$ is a monotonically increasing submodular function under the proposed graph construction.

Proof: The proof is divided into two parts. The first part proves the monotonicity. In the second part, we show the submodularity.

Monotonicity: $\delta\mathcal{H}_{a_1}(A) \geq 0$ for all $A \subseteq E$ and $a_1 \in E \setminus A$. WLOG let assume $a_1 = e_{1,2}$. Under the selected set $A \cup \{e_{1,2}\}$, the loop weights for vertices v_1 and v_2 are given by $c_1 \equiv w_1 - \sum_{k:e_{1,k} \in A\cup\{e_{1,2}\}} w_{1,k}$ and $c_2 \equiv w_2 - \sum_{k:e_{2,k} \in A\cup\{e_{1,2}\}} w_{2,k}$ respectively. From Equation (11) of the main paper and with some simple algebraic manipulations, the marginal gain from adding $e_{1,2}$ to A is given by

$$\delta\mathcal{H}_{e_{1,2}}(A) = \Big\{\frac{w_{1,2}+c_1}{w_T}\log(\frac{w_{1,2}+c_1}{w_T}) - \frac{w_{1,2}}{w_T}\log(\frac{w_{1,2}}{w_T}) - \frac{c_1}{w_T}\log(\frac{c_1}{w_T})\Big\} + \Big\{\frac{w_{2,1}+c_2}{w_T}\log(\frac{w_{2,1}+c_2}{w_T}) - \frac{w_{2,1}}{w_T}\log(\frac{w_{2,1}}{w_T}) - \frac{c_2}{w_T}\log(\frac{c_2}{w_T})\Big\} \quad (1)$$

$$= \frac{w_{1,2}+c_1}{w_T}\Big\{\frac{w_{1,2}}{w_{1,2}+c_1}\log(\frac{w_{1,2}+c_1}{w_{1,2}}) + \frac{c_1}{w_{1,2}+c_1}\cdot \log(\frac{w_{1,2}+c_1}{c_1})\Big\} + \frac{w_{2,1}+c_2}{w_T}\Big\{\frac{w_{2,1}}{w_{2,1}+c_2}\log(\frac{w_{2,1}+c_2}{w_{2,1}}) + \frac{c_2}{w_{2,1}+c_2}\log(\frac{w_{2,1}+c_2}{c_2})\Big\} \geq 0. \quad (2)$$

Note that the terms in the two curly bracket pairs in (2) are the entropy values of the two binary random variables with distributions $(\frac{w_{1,2}}{w_{1,2}+c_1}, \frac{c_1}{w_{1,2}+c_1})$ and $(\frac{w_{2,1}}{w_{2,1}+c_2}, \frac{c_2}{w_{2,1}+c_2})$ respectively. Since the entropy of a discrete random variable is nonnegative, we show that $\delta\mathcal{H}_{e_{1,2}}(A) \geq 0$ and prove the monotonicity.

Submodularity: $\delta\mathcal{H}_{a_1}(A) \geq \delta\mathcal{H}_{a_1}(A \cup \{a_2\})$ for all $A \in E$ and $a_1, a_2 \in E \setminus A$. Based on whether these edges share a common vertex or not, we discuss the following two cases.

1.) a_1 and a_2 have no common vertex. WLOG, let assume $a_1 = e_{1,2}$ and $a_2 = e_{3,4}$. Since $e_{3,4}$ is not connected to either v_1 or v_2, the addition of $e_{3,4}$ has no effect on the loop weights of v_1 and v_2; therefore, we have the following equalities

$$c_1 = w_1 p_{1,1}(A \cup \{e_{1,2}, e_{3,4}\}) = w_1 p_{1,1}(A \cup \{e_{1,2}\}) \quad (3)$$
$$c_2 = w_2 p_{2,2}(A \cup \{e_{1,2}, e_{3,4}\}) = w_2 p_{2,2}(A \cup \{e_{1,2}\}) \quad (4)$$

where $p_{i,j}$'s are the transition probabilities given in Equation (11) of the main paper. As a result, $\delta\mathcal{H}_{e_{1,2}}(A \cup \{e_{3,4}\})$ can also be simplified to (1), and $\delta\mathcal{H}_{e_{1,2}}(A \cup \{e_{3,4}\}) = \delta\mathcal{H}_{e_{1,2}}(A)$.

2.) a_1 and a_2 share a common vertex. WLOG, let assume $a_1 = e_{1,2}$ and $a_2 = e_{1,3}$ where v_1 is the shared vertex. Due to the shared vertex, the loop weight for vertex v_1 and v_2 after the addition of $e_{1,3}$ are given by $d_1 \equiv w_1 - \sum_{k:e_{1,k} \in A\cup\{e_{1,2},e_{1,3}\}} w_{1,k} = c_1 - w_{1,3}$ and $d_2 \equiv w_2 - \sum_{k:e_{2,k} \in A\cup\{e_{1,2},e_{1,3}\}} w_{2,k} = w_2 - \sum_{k:e_{2,k} \in A\cup\{e_{1,2}\}} w_{2,k} = c_2$ respectively. The marginal gain from the addition of $e_{1,2}$ to $A \cup \{e_{3,4}\}$ is equal to

$$\delta\mathcal{H}_{e_{1,2}}(A \cup \{e_{3,4}\}) = \Big\{\frac{w_{1,2}+d_1}{w_T}\log(\frac{w_{1,2}+d_1}{w_T}) - \frac{w_{1,2}}{w_T}\log(\frac{w_{1,2}}{w_T}) - \frac{d_1}{w_T}\log(\frac{d_1}{w_T})\Big\} + \Big\{\frac{w_{2,1}+d_2}{w_T}\cdot \log(\frac{w_{2,1}+d_2}{w_T}) - \frac{w_{2,1}}{w_T}\log(\frac{w_{2,1}}{w_T}) - \frac{d_2}{w_T}\log(\frac{d_2}{w_T})\Big\} \quad (5)$$

By subtracting (5) from (1), we have

$$\delta\mathcal{H}_{e_{1,2}}(A) - \delta\mathcal{H}_{e_{1,2}}(A \cup \{e_{3,4}\}) \quad (6)$$
$$= \Big\{\frac{w_{1,2}+d_1+w_{1,3}}{w_T}\log(\frac{w_{1,2}+d_1+w_{1,3}}{w_T}) - \frac{d_1+w_{1,3}}{w_T}\log(\frac{d_1+w_{1,3}}{w_T})\Big\} - \Big\{\frac{w_{1,2}+d_1}{w_T}\log(\frac{w_{1,2}+d_1}{w_T}) - \frac{d_1}{w_T}\log(\frac{d_1}{w_T})\Big\} = g(\frac{d_1+w_{1,3}}{w_T}) - g(\frac{d_1}{w_T}) > 0 \quad (7)$$

The last equation is an application of the strictly increasing property of $g(x) \equiv (x+\xi)\log(x+\xi) - x\log(x)$ where $\xi = \frac{w_{1,2}}{w_T}$ in this case.

From the above case discussions, we show that $\delta\mathcal{H}_{e_{1,2}}(A) \geq \delta\mathcal{H}_{e_{1,2}}(A \cup \{e_{3,4}\})$ and complete the proof. □

Lemma 2: The balancing function $\mathcal{B} : 2^E \rightarrow \mathbb{R}$ is a monotonically increasing submodular function under the proposed graph construction.

Proof: We prove the monotonicity and submodularity separately.

Monotonicity: $\delta\mathcal{B}_{a_1}(A) \geq 0$ for all $A \subseteq E$ and $a_1 \in E \setminus A$. WLOG let assume $a_1 = e_{1,2}$, v_1 be in cluster S_i, and v_2 be in cluster S_j. From Equation (16) of the main paper, the probability that a randomly picked vertex is in S_i and S_j is given by $p_{Z_A}(i) = \frac{|S_i|}{|V|}$ and $p_{Z_A}(j) = \frac{|S_j|}{|V|}$ respectively.

To prove the monotonicity, we discuss the case where $i \neq j$ — v_1 and v_2 are in two different clusters given A. [1] Let $p_i = p_{Z_A}(i)$ and $p_j = p_{Z_A}(j)$. The addition of $e_{1,2}$ to A merges S_i and S_j, and the probability that a randomly picked vertex is in the merged cluster is equal to $p_{Z_A}(i) + p_{Z_A}(j) = \frac{|S_i|+|S_j|}{|V|}$. The increase in the balancing function is then given by

$$\delta\mathcal{B}_{e_{1,2}}(A) = H(A \cup \{e_{1,2}\}) - H(A) - N_{\mathcal{A}\cup\{e_{1,2}\}} + N_{\mathcal{A}}$$
$$= 1 - (p_i + p_j)\log(p_i + p_j) + p_i \log p_i + p_j \log p_j \quad (8)$$
$$= 1 + p_i \log\frac{p_i}{p_i + p_j} + p_j \log\frac{p_j}{p_i + p_j} \quad (9)$$
$$\geq 1 + (p_i + p_j)\log(\frac{p_i + p_j}{2(p_i + p_j)}) = 1 - (p_i + p_j) \geq 0. \quad (10)$$

Note that the first inequality in (10) is an application of the log-sum inequality. From (10) we prove the monotonically increasing property.

Submodularity: $\delta\mathcal{B}_{a_1}(A) \geq \delta\mathcal{B}_{a_1}(A \cup \{a_2\})$ for all $A \in E$ and $a_1, a_2 \in E \setminus A$. To prove the submodularity, we discuss the cases where the addition of a_1 to A combines two clusters. If it does not, both sides of the above equation equal zero and the submodularity holds. WLOG let assume $a_1 = e_{1,2}$ and $a_2 = e_{3,4}$. Depending on whether the addition of $e_{3,4}$ combines two clusters or not, there are three derivative cases.

1. If $i = j$, then $\delta\mathcal{B}_{a_1}(A) = 0$ and the monotonicity holds.

1.) Let assume the addition of $e_{3,4}$ combines the clusters S_i and S_j. This means v_1 and v_2 are in the same cluster given $A \cup \{e_{3,4}\}$. Therefore the addition of $e_{1,2}$ to $A \cup \{e_{3,4}\}$ has no effect on the graph partition. Both the number of clusters and the cluster membership distribution remain the same; hence, $\delta\mathcal{B}_{e_{1,2}}(A) \geq \delta\mathcal{B}_{e_{1,2}}(A \cup \{e_{3,4}\}) = 0$.

2.) In the case that $e_{3,4}$ combines some other clusters S_k and S_m where $|\{k,m\} \cap \{i,j\}| = \emptyset$ or $e_{3,4}$ does not combine any clusters. The addition of $e_{1,2}$ will still merge S_i and S_j. Moreover, $p_{Z_{A\cup\{e_{3,4}\}}}(i) = p_{Z_A}(i)$ and $p_{Z_{A\cup\{e_{3,4}\}}}(j) = p_{Z_A}(j)$. As a result $\delta\mathcal{B}_{e_{1,2}}(A) = \delta\mathcal{B}_{e_{1,2}}(A \cup \{e_{3,4}\})$.

3.) Suppose that the addition of $e_{3,4}$ combines S_k to S_i. Let $p_k = p_{Z_A}(k)$. The marginal gain obtained from adding $e_{1,2}$ to $A \cup \{e_{3,4}\}$ is given by

$$\delta\mathcal{B}_{e_{1,2}}(A \cup \{e_{3,4}\}) = -(p_i + p_j + p_k)\log(p_i + p_j + p_k) + (p_i + p_k)\log(p_i + p_k) + p_j \log(p_j) + 1. \tag{11}$$

By subtracting (11) from (8), we have

$$\begin{aligned} &\delta\mathcal{B}_{e_{1,2}}(A) - \delta\mathcal{B}_{e_{1,2}}(A \cup \{e_{3,4}\}) \\ &= (p_i + p_j + p_k)\log(p_i + p_j + p_k) - (p_i + p_j)\log(p_i + p_j) \\ &\quad - ((p_i + p_k)\log(p_i + p_k) - p_i \log p_i) \\ &= g(p_i + p_j) - g(p_i) \geq 0 \end{aligned} \tag{12}$$

Note that the last inequality is an application of strictly increasing property of the function $g(x) = (x+\xi)\log(x+\xi) - x\log(x)$.

From showing the diminishing return property for the above cases, we complete the proof. □

Lemma 3: Let E be the edge set, and let $\mathcal{I}$ be the set of subsets $A \subseteq E$ which satisfies: 1.) A is cycle-free and 2.) A constitutes a graph partition with more than or equal to K connected components. Then the pair $M = (E, \mathcal{I})$ is a matroid.

Proof: The proof is given by showing that M satisfies the three matroid conditions.

1.) It is straightforward to see that the empty set is an independent set of the matroid — it contains no cycles and constitutes a $|V|$-partition where $N_\emptyset = |V| > K$.

2.) Let $I \in \mathcal{I}$ and $I' \subseteq I$. This implies that I has no cycles and $N_I \geq K$. Since $I' \subseteq I$, the set I' can be obtained by removing edges from the set I. The removal of edges does not add cycles and increases the number of connected components. Therefore the set I' also contains no cycles and $N_{I'} \geq K$.

3.) Let I_1 and I_2 be two independent sets in $\mathcal{I}$ such that $|I_1| < |I_2|$. The independent set assumptions implies the associated connected components of I_1 and I_2 satisfy the following relations: $N_{I_1} = |V| - |I_1| \geq K$, $N_{I_2} = |V| - |I_2| \geq K$, and $|I_1| < |I_2| \implies |V| - |I_1| > |V| - |I_2| \geq K$. From the above statements we have the following equations.

$$N_{I_1} > N_{I_2} \tag{13}$$

$$N_{I_1} = |V| - |I_1| \geq K + 1 \tag{14}$$

We now prove that there exists some $e \in I_2 - I_1$ such that $N_{I_1 \cup \{e\}} \geq K$ and the set $I_1 \cup \{e\}$ is cycle-free; i.e., $I_1 \cup \{e\}$ is an independent set. Since adding one edge to a graph decreases the number of connected components by at most one, Equation (14) implies $N_{I_1 \cup \{e\}} \geq K$ for $e \in I_2 - I_1$. The remaining part of the proof is achieved by contradiction.

Let us assume there is no edge $e \in I_2 - I_1$ such that $I_1 \cup \{e\}$ is cycle-free. In other words, adding any edge e in $I_2 - I_1$ to the set I_1 will result in a cycle and leaves the number of connected components in the graph unchanged, $N_{I_1 \cup \{e\}} = N_{I_1}$. Thus by adding all the edges from $I_2 - I_1$ to the set I_1, the number of connected components will remain as N_{I_1}, i.e., $N_{I_1 \cup (I_2 - I_1)} = N_{I_1 \cup I_2} = N_{I_1}$.

We know that the set $I_1 \cup I_2$ can also be obtained by adding edges from $I_1 - I_2$ to I_2. Since adding edges to a graph can only decrease the number of connected components, we have the following relation $N_{I_1 \cup I_2} \leq N_{I_2} \implies N_{I_1} \leq N_{I_2}$. This contradicts Equation (13) and thus the lemma is proved. □