Homework 1

August 31, 2017

A graph G is a cluster graph[1] if every connected component of G is a clique. In the Cluster Editing problem, we are given as input a graph G and an integer k, and the objective is to check whether one can edit (add or delete) at most k edges in G to obtain a cluster graph. That is, we look for a set $F \subseteq \binom{V(G)}{2}$ of size at most k, such that the graph $(V(G), (E(G)\backslash F) \cup (F \backslash E(G)))$ is a cluster graph.

1. Show that a graph G is a cluster graph if and only if it does not have an induced path on three vertices (sequence of three vertices u, v, w such that uv and vw are edges and $uw \notin E(G)$).

2. Show a kernel for cluster editing with $O(k^2)$ vertices.

Notations. We first introduce some notations that we will use. We denote a graph $G = (V, E)$, where V is the vertex set of G and E is the edge set of G. For arbitrary vertex pair $(u, v) \in V$, we denote by $w \in V$ by a *common neighbor* of u and v such that $(w, u) \in E$ and $(w, v) \in E$. We denote w by a *non-common neighbor* of u and v, where $w \neq u$ and $w \neq v$, such that either $(w, u) \in E$ or $(w, v) \in E$ but not both.

Question 1. We prove the following lemma.

Lemma 1. *A graph G is a cluster graph if and only if it does not have an induced path P_3 on three vertices.*

Proof. We need to prove in two directions. Firstly, we prove that if G is a cluster graph, then G does not have an induced path P_3 on three vertices. Secondly, we prove that if G does not have an induced path P_3 on three vertices, then G is a cluster graph.

Claim 1. *If G is a cluster graph, then G does not have an induced P_3 on three vertices.*

Proof. We prove by contradiction. We assume that a graph G has an induced path P_3. Recall that an induced path P_3 is a sequence of vertices in G that each two adjacent vertices in the sequence are connected by an edge in G, and each two non-adjacent vertices in the sequence are note connected by an edge in G. It is easy to see that an induced path P_3 is definitely not a subgraph of a clique in G, and also an induced path P_3 is not a clique as well. Indeed, if P_3 is a subgraph of a clique, there always exists two adjacent vertices with no edge between them. This contradicts the definition of a clique. Furthermore, this contradicts to the fact that G is a cluster graph where all connected components are cliques. Because in each clique, every pair of distinct vertices is connected by a unique edge. □

---

[1]Regarding to this homework, we will assume, without loss of generality, that G is a connected graph. For a non-connected graph G, we can also process every connected component separately using our solutions.

Claim 2. *If G does not have an induced path P_3 on three vertices, then G is a cluster graph.*

Proof. We prove by contradiction as well. We assume that G is not a cluster graph. Then there must exist at least one subgraph $G' \subseteq G$, which is not a clique, and this subgraph G' must include at least one pair of vertices (u, v) such that there is no edge between them. Without loss of generality, we can assume that there only exist one pair such vertices (u, v). Therefore, there must exist a common neighbor w of u and v because in the subgraph G' the vertex u has edge connections with other vertices except with the vertex v. Therefore vertices u, v and w would form an induced path P_3, and this contradicts to the fact that G does not have an induced path P_3 on three vertices. □

This completes the proof. □

Question 2. We show a kernel for k-Cluster Editing problem with $O(k^2)$ vertices. More formally, we prove the following theorem.

Theorem 1. *The k-Cluster Editing problem has an $O(k^2)$-vertex kernel.*

Proof. Inspired by the lemma we proved above, we design our reduction rules as follows:
Reduction Rule 1: For each pair of vertices $u, v \in V$.

1.1: If u and v have more than k common neighbors, then the edge (u, v) has to belong to E. If $(u, v) \notin E$, the we add it to E.

1.2: If u and v have more than k non-common neighbors, then the edge (u, v) cannot belong to E. If $(u, v) \in E$, delete it.

1.3: If u and v have both more than k common and more than k non-common neighbors, then the given instance has no solution.

we first prove that the reduction rule 1 is correct. *For the reduction rule 1.1*, if u and v have more than k common neighbors. If we did exclude the edge (u, v) from E, then we would have to delete the edge (u, w) or (v, w) or both. However, this would require at least $k + 1$ edge deletions, which contradicts to the fact that at most k edge modifications are allowed. *For the reduction rule 1.2*, if u and v have more than k non-common neighbors. If we did include (u, v) in E, then we would have to either delete the edge (u, w) from E or to add (v, w) to E, with at least $k + 1$ non-common neighbors, this would require at least $k + 1$ edge modifications. This contradicts to the fact that at most k edge modifications are allowed. *For the reduction rule 1.3*, if u and v have more than k common neighbors and more than k non-common neighbors. From the reduction rule 1.1 and the reduction rule 1.2, it is clear that it would require more than k edge modifications both including (u, v) in E and when excluding (u, v) from E. We note that the reduction rule 1 applies to every vertex pair (u, v) for which the number of vertices which are neighbors of u or v (or both) is greater than $2k$.
Reduction Rule 2: Delete the connected components which are cliques from the graph. The correctness of this reduction rule is straightforward and we will use it in the following proof.

Now we are ready to show a kernel which contains $O(k^2)$ vertices. Firstly we note that since the reduction rule 2 deletes all isolated cliques from the graph G, so G is not clique and we need at least one edge modification to transform the graph $G = (V, E)$ into a reduced graph $G' = (V', E')$, consisting of disjoint cliques.

Let $k = k_A + k_D$, where k is the minimum number of edge modifications, k_A is the number of edge additions and k_D is the number of edge deletions. Under the assumption that G is reduced

with respect to reduction rules 1 and 2, we show by contradiction that $|V| \leq 2k^2 + k$. We assume that $|V| > (2k + 1) \cdot k$ and consider following two cases.

Case 1: $k_A = 0$ and $1 \leq k_D = k$. Let $S \subset V$ be a vertex set of a largest clique in G'. Then the vertices in S also form a clique in G since $k_A = 0$. Since G is connected, at least one vertex $u \in S$ is connected to a vertex $v \notin S$. We further consider two subcases.

(a) v is not connected to any other vertex $z \in S$, where $z \neq u$, we conclude that $|S| \geq |V|/(k_D + 1)$. By k_D edge deletions, the graph G is transformed to the graph G' and G' contains at most $k_D + 1$ cliques. Therefore, a largest clique in G' contains at least $|V|/(k_D + 1)$ vertices. We have

$$|S| \geq \frac{|V|}{(k_D + 1)} > \frac{(2k + 1) \cdot k}{(k_D + 1)} = \frac{k^2 + k}{k + 1} + \frac{k^2}{k + 1} > k + 1$$

Therefore we have $|S| \geq k + 2$ and u has at least $k + 1$ neighbors, all of which are from the set S and are not equal to u. Furthermore, these $k + 1$ neighbors are not neighbors of v. This situation contradicts the reduction rule 1.

(b) There exists a vertext $z \in S$, $z \neq u$ and $(z, v) \in E$. In this case we conclude that the clique size is at least $|V|/k_D$. Since G' contains at most k_D cliques, a largest clique in G' contains at least $|V|/k_D$ vertices. Therefore, we have

$$|S| \geq \frac{|V|}{k_D} > \frac{(2k + 1) \cdot k}{k_D} = 2k + 1$$

Thus S contains at least $2k + 1$ vertices and at most k of them are connected to V. Thus u has at least $k + 1$ neighbors in this clique which are not neighbors of V, contradicting the reduction rule 1.

Case 2: $1 \leq k_A \leq k$ and $k_D < k$. We also consider a vertex set $S \subset V$ of a largest clique in G'. Since G' contains at most $k_D + 1$ cliques, we have $|S| \geq |V|/(k_D + 1)$, and because $k_D < k$ we have $|S| \geq |V|/k$. Since we assume that $|V| > (2k + 1) \cdot k$, we have $|S| > 2k + 1$. Since vertices in S form a clique in G' and at most k edges are added in the transformation from G to G'. In the graph G, at most k vertex pairs in the set S which are not connected by an edge, we also consider two cases and prove the theorem by contradiction.

(a) we consider the case $k_A < k$, we conclude that there exists a vertex $u \in S$, $v \notin S$ such that the edge $(u, v) \in E$. Since $|S| > 2k + 1$, there are at least $(2k + 1) - k_A - 1$ vertices z in S with $(z, u) \in E$. Therefore, at most $k_D - 1$ vertices z are included in the set S with $(z, u) \in E$. Thus we have at least $(2k + 1) - (k_A + k_D) = (2k + 1) - k = k + 1$ vertices $z \in S$ with $(z, u) \in E$ but $(z, v) \notin E$. Therefore, the vertex u has at least $k + 1$ neighbors in G which are not neighbors of v and this contradicts to the reduction rule 1.

(b) We consider the case $k_A = k$ and conclude that $S = V$ and G' consists of only one clique. Since $k_A = k \geq 1$, there exists a pair of vertices $u, v \in V$ such that $(u, v) \notin E$. Because we assume that $|S| = |V| > 2k + 1$, then there exists more than $\binom{2k + 1}{2}$ vertex pairs. Therefore, out of these pairs there are at most k pairs, including the vertex pair (u, v), are not connected by an edge. Thus u and v have at least $k + 1$ neighbors in G by counting arguments. This contradicts to the reduction rule 1.

3

In conclusion, the vertex set V contains at most $2k^2 + k$ vertices. Namely, we obtain a kernel for the Cluster Editing problem with $O(k^2)$ vertices. $\qquad\square$