Spring 4-3-2019

Finding Connected-Dense-Connected Subgraphs and variants is NP-Hard

Dhara Shah
dshah8@student.gsu.edu

Sushil Prasad
sprasad@gsu.edu

Yubao Wu
ywu28@gsu.edu

finding Connected-Dense-Connected Subgraphs and variants is NP-Hard

Dhara Shah[1], Yubao Wu[1], and Sushil Prasad[1]

Department of Computer Science, Georgia State University, Atlanta 30303, USA
{dshah8, ywu28, sprasad}@gsu.edu

Abstract. Finding Connected-Dense-Connected (CDC) subgraphs from Triple Networks is NP-Hard. finding One-Connected-Dense (OCD) subgraphs from Triple Networks is also NP-Hard. We present formal proofs of these theorems hereby.

Keywords: Triple Networks · Connected-Dense-Connected subgraphs · One-Connected-Dense subgraphs · NP-Hard

Theorem 1. *Finding a CDC subgraph in a Triple Network is NP Hard.*

Proof. We prove that finding a CDC subgraph is a reduction of set-cover problem. Let $R = \{r_1, \cdots, r_p\}$ be a set and and $C = \{C_1, \cdots, C_q\}$ be its cover with $R = \cup_{i=1}^{q} C_i$. The aim of this set cover problem is to find minimum subset $C_{opt} \subset C$, known as optimal set-cover, such that each $r_j \in R$ belongs to at least one set of C_{opt}. This problem is proved to be NP complete.

Let $T = \{t_1, \cdots, t_p\}$ be a set of points, having the same cardinality as R. Let $D = \{D_1, \cdots, D_q\}$ be a set-cover of T, analogous to C, such that if $r_i \in C_j$, then $t_i \in D_j$. Hence, T, D can be considered as a copy of R, C.

We construct the Triple Network as follows. Let $V_a = \{h, r_1, \cdots, r_p, C_1, \cdots, C_q\}$, where node h is connected to every $C_i \in C$ and node r_i is connected to node C_j if $r_i \in C_j$ in the set-cover problem. Similarly, let $V_b = \{k, t_1 \cdots, t_p, D_1, \cdots, D_q\}$ be the analogous set to V_a. We connect V_a and V_b by connecting all nodes $\{r_1, \cdots, r_p, h\}$ to all nodes $\{t_1, \cdots, t_p, k\}$.

Construction of this Triple Network is illustrated in figure 1 from an instance of set-cover problem $C_1 = \{r_1, r_2\}, C_2 = \{r_1\}, C_3 = \{r_2, r_4\}, C_4 = \{r_2, r_3\}, C_5 = \{r_4\}$.

Let $C_{opt} \subset C$ be an optimal solution to the set-cover problem of C and $|C_{opt}| = q^* \leq q$. Similarly, let D_{opt} be the analogous optimal solution to D and $|D_{opt}| = q^* \leq q$. Let $H = \{h, r_1, \cdots, r_p\}$ and $J = \{k, t_1, \cdots, t_p\}$. The subgraph induced by $S_a = H \cup C_{opt}$ is connected in V_a, and similarly, the subgraph induced by $S_b = J \cup D_{opt}$ is connected in V_b. Hence, the sub Triple Network $G[S_a, S_b]$ has density $\rho(S_a, S_b) = \frac{(p+1)^2}{(p+q^*+1)}$.

Let S_1 and S_2 be any nonempty node sets where $G_a[S_1]$ and $G_b[S_2]$ are connected. In general, $S_1 = H' \cup C'$ where $H' \subset H$ and $C' \subset C$. Similarly, $S_2 = J' \cup D'$ where $J' \subset J$ and $D' \subset D$. We show that $\rho(S_1, S_2) \leq \rho(S_a, S_b)$,

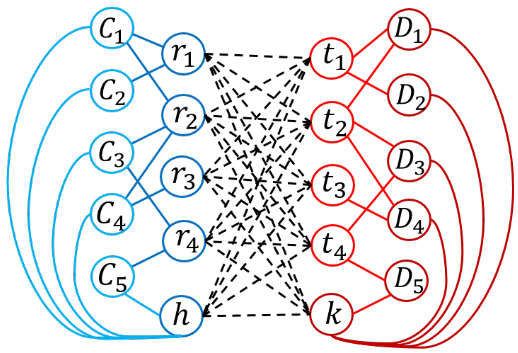

Fig. 1. Triple Network from set-cover

making $G[S_a, S_b]$ a CDC subgraph. Let $|H'| = p_1$, $|C'| = q_1$, $|J'| = p_2$ and $|D'| = q_2$. Hence, $\rho(S_1, S_2) = \frac{p_1 p_2}{\sqrt{(p_1+q_1)(p_2+q_2)}}$.

First, we consider the case when S_1 contains all the nodes of H and S_2 contains all the nodes of J. In this case, $p_1 = p_2 = p + 1$. Also, by definition of optimal set-cover, $q^* \leq q_1$ and $q^* \leq q_2$. Hence, $\rho(S_1, S_2) = \frac{(p+1)^2}{\sqrt{(p+q_1+1)(p+q_2+1)}} \leq \frac{(p+1)^2}{(p+q^*+1)} = \rho(S_a, S_b)$.

Second, we consider the case when S_1 contains a subset of nodes $H' \subset H$. In this case, we first show that adding elements from $H \setminus H'$ to S_1 will only increase its density.

If $h \notin S_1$, then after adding h to S_1, the resulting subgraph has density $\frac{(p_1+1)p_2}{\sqrt{(p_1+q_1+1)(p_2+q_2)}} > \frac{p_1 p_2}{\sqrt{(p_1+q_1)(p_2+q_2)}} = \rho(S_1, S_2)$. This subgraph is also connected in G_a, since h is connected to every $C_i \in C$. To add a node $r_j \in H \setminus H'$ and making it still connected, we need to add at most one node C_i to C' with $r_j \in C_i$. Hence, the density of this resulting subgraph is $\frac{(p_1+1)p_2}{\sqrt{(p_1+q_1+2)(p_2+q_2)}} > \frac{p_1 p_2}{\sqrt{(p_1+q_1)(p_2+q_2)}} = \rho(S_1, S_2)$. We can repeat this process by adding remaining nodes of $H \setminus H'$ to S_1, while density of the resulting subgraphs keeps increasing.

Similarly, adding elements from $J \setminus J'$ to S_2 increases density of the resulting subgraphs. Since we proved in the first case that the density $\rho(S_1, S_2)$ when $H \subset S_1$ and $J \subset S_2$, we have hence completed the proof of the second case.

In summary, we proved that for any nonempty sets $S_1 \subset V_a$ and $S_2 \subset V_b$, $\rho(S_1, S_2) \leq \rho(S_a, S_b)$, making $G[S_a, S_b]$ a CDC subgraph. Also,$G[S_a, S_b]$ is the solution inducted by optimal set covers, an instance being $S_a = \{r_1, r_2, r_3, r_4, h, C_1, C_3, C_4\}$ and $S_b = \{s_1, s_2, s_3, s_4, k, D_1, D_3, D_4\}$ hence proving that finding a CDC subgraph is NP hard.

Lemma 1. *Finding OCD subgraph in triple network is NP hard*

Proof. We prove that finding OCD subgraph is also reduction of the set cover problem. We first construct the triple network same as in theorem 1. Let $S_a = H$ and $S_b = J \cup D_{opt}$. The subgraph $G[s_a, S_b]$ hence has density $\rho(S_a, S_b) =$

$\frac{(p+1)^2}{\sqrt{(p+1)(p+q^*+1)}}$ We claim that $G[S_a, S_b]$ is an OCD subgraph. We observe that $G[S_b]$ is connected.

Let S_1 and S_2 be any nonempty node sets where either $G[S_1]$ or $G[S_2]$ is connected. In general, $S_1 = H' \cup C'$ where $H' \subset H$. Similarly, $S_2 = J' \cup D'$ where $J' \cup J$. We show that $\rho(S_1, S_2) \leq \rho(S_a, s_b)$.

First, we consider the case when S_1 contains all the nodes of H and S_2 contains all the nodes of J. In this case, $p_1 = p_2 = p + 1$. Also, by definition of optimal set-cover, $q^* \leq q_1$ and $q^* \leq q_2$. Hence, $\rho(S_1, S_2) = \frac{(p+1)^2}{\sqrt{(p+q_1+1)(p+q_2+1)}} \leq \frac{(p+1)^2}{\sqrt{(p+q^*+1)(p+1)}} = \rho(S_a, S_b)$.

Second, we consider the case when S_1 contains a subset of nodes $H' \subset H$. In this case, we first show that adding elements from $H \setminus H'$ to S_1 will only increase its density. Suppose, $G_a[S_1]$ is not connected and $G_b[S_2]$ is connected. Then, after adding element from $H \setminus H'$, the resulting subgraph has density $\frac{(p_1+1)p_2}{\sqrt{(p_1+q_1)(p_2+q_2)}} > \frac{p_1 p_2}{(p_1+q_1)(p_2+q_2)} = \rho(S_1, S_2)$. This includes adding h to S_1 if $h \notin H'$, making resultant subgraph connected in V_a. Now suppose $G_a[S_1]$ is connected. Then, following the same case of theorem 1, we first add h if it is not in H' and then add element from $H \setminus H'$ and still show that the resultant subgraph is connected in V_a and its density increases. Similarly, we conclude that when S_2 contains a subset of nodes in $J' \subset J$, adding elements from $J' \setminus J$ also increases the density of the resultant subgraph.

At last, we observe that if $G_a[S_2]$ is connected, then the resultant subgraph obtained by removing elements from C' has density $\frac{p_1 p_2}{\sqrt{(p_1+q_1-1)(p_2+q_2)}} > \rho(S_1, S_2)$.

In summary, we have proved that for any nonempty sets $S_1 \subset V_a$ and $S_2 \subset V_b$ with either $G_a[S_1]$ or $G_b[S_2]$ connected has density $\rho(S_1, S_2) \leq \rho(S_a, S_b)$, making $G[S_a, S_b]$ an OCD subgraph. Also, $G[S_a, S_b]$ is the solution induced by optimal set cover, an instance being $S_a = \{r_1, r_2, r_3, r_4, h\}, S_b = \{s_1, s_2, s_3, s_4, k, D_1, D_3, D_4\}$ hence proving that finding OCD subgraphs is NP hard.