A Title

An Author[1] and Another Author[2]

[1]An affiliation
[2]Another affiliation

Abstract

An abstract.

1 Introduction

Social networks provide a powerful framework for modeling real-world interactions, capturing the flow of information among individuals, organizations, and entities. As a result, inherent to social networks is the concept of *network fairness* [6, 2, 8] — if LinkedIn users share resources and job opportunities with their followers, who you follow is directly related to how much information you recieve, and how soon this information reaches you. In a social network, certain nodes have more advantage than others simply due to their connectivity to the rest of the network. Understanding, quantifying, and ultimately mitigating these disparities is crucial for promoting fairness in the environments that social networks represent.

A significant body of research focuses on maximizing the information access of nodes. Specifically, [2, 3] study the problem of Broadcast Improvement, where the goal is to add edges to a graph to maximize the graph's broadcast, the minimum probability over all vertex pairs u, v that information starting at u will reach v. This notion is similar to the diameter of a graph, but adapted to Independent Cascade, where edges each have a probability of transmitting information. However, less attention has been given to *equalizing* information access across different parts of a network, which is the focus of our work.

At a high level, we aim to add a budgeted number of edges to a graph to make two vertices equally important in the graph. Formally, given a graph and a budget k, our goal is to augment the graph by adding at most k edges to make the ratio between the closeness centrality of two designated vertices as close to 1 as possible. The closeness centrality of a vertex v is defined as the sum of shortest path distances from v to all other vertices, making it a useful measure of how efficiently a node can access information in the shortest-path metric — if a node is further away from more vertices (higher closeness centrality), it is less likely to receive information quickly, assuming information can arise from any part of the network.

To tackle this problem, we draw insights from existing work on edge augmentation in shortest-path settings. Previous research has studied strategies to maximize the centrality of a single node or a group of nodes [5, 10], although this is still a notably different problem from ours, which seeks to equalize closeness centralities. We also study Diameter Minimization [1, 4, 7, 9], which seeks to add a limited number of edges to minimize a graph's diameter, as well as broader edge augmentation approaches using k-center strategies [11]. These existing approaches provide theoretical guarantees and algorithmic strategies which inform our work, connecting network fairnesss with more classical graph optimization problems and techniques.

1.1 Our Contributions

For the problem of making the ratio of the closeness centrality of two vertices as close to 1 as possible, we present a simple algorithm that always achieves a ratio of $\frac{1}{2}$. Reducing from Set Cover, we then show that achieving any ratio $\tau \in \left(\frac{1}{2}, 1\right]$ is NP-hard. As the best possible ratio is 1, this also implies that our algorithm for achieving a ratio of $\frac{1}{2}$ is a $\frac{1}{2}$-approximation for our problem.

hardness of approximation? Group problem?

2 Preliminaries

The shortest path length between two vertices $u, v \in V$ within a graph $G = (V, E)$ is represented as $d_G(u, v)$. A graph $G = (V, E)$ augmented with an edge set $S \subseteq V^2 \backslash E$ is denoted as $G + S = (V, E \cup S)$. Finally, we define the closeness centrality of a vertex $v \in V$ in a graph $G = (V, E)$ as $c_G(v) = \sum_{u \in V} d_G(u, v)$, the sum of the shortest paths to each other vertex in the graph. Note that having a smaller closeness centrality value means a node is closer to more vertices in the graph, i.e. more important in the graph.

Given a graph G and vertices a, b, we want to find the k edges which will make the ratio of their closeness centralities as close to 1 as possible.

Closeness Ratio Improvement

Input: A graph $G = (V, E)$, vertices $a, b \in V$, and a positive integer $k \in \mathbb{N}$.

Task: Find a set of edges S of size at most k which maximizes $\frac{\min(c_{G+S}(a), c_{G+S}(b))}{\max(c_{G+S}(a), c_{G+S}(b))}$

Why have we chosen closeness centrality as our measure of node importance, when many other such measures exist? Primarily, edge additions cannot increase/worsen the closeness centrality of any vertex in the graph. This is different than betweenness centrality, where adding an edge to increase the betweenness centrality of a vertex can inadvertenly decrease the betweenness centrality of another vertex.

Closeness centrality also simplifies the concept of being close to *all vertices* in a graph — a vertex needs to be reasonably close to all other vertices to achieve a low centrality value. Furthermore, if a vertex is close to most of the vertices in the graph but very far from a few, this vertex's centrality score can blow up just the same as if it was far from everything. This agrees with our assumptions about information flow in a network, new information can arise from anywhere and we want to maximize a vertex's ability to receive this information as soon as possible.

3 Hardness of Closeness Ratio Improvement

In this section, we establish the NP-hardness of CRI, over a range of target ratios.

Theorem 1. *Closeness Ratio Improvement is NP-Hard for $\tau = 1$*

Proof. Given an instance of Set Cover with universe $U = \{e_1, e_2, \ldots, e_n\}$, a collection of non-empty subsets $S_1, S_2, \ldots, S_m \subseteq U$, and a positive integer $k \in \mathbb{N}$, construct a decision instance of Closeness Ratio Improvement as follows. Vertices a, b, c form a clique. Each of the n elements from the universe have a corresponding vertex $e_1, e_2, \ldots, e_n$. For each set S_i, create vertex s_i and connect it to c. Next, connect s_i to each vertex that represents an element contained within S_i. Finally, create an independent set of $X = n + k$ vertices each

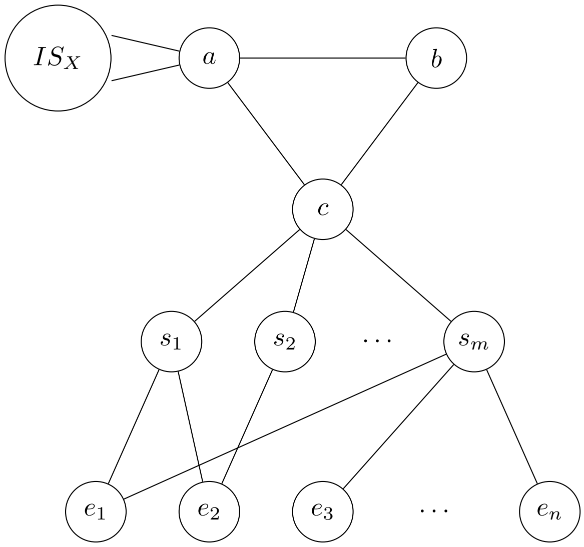

Figure 1: Construction of a CRI instance with vertices a, b from Set Cover.

connected to a. The construction of G is depicted in Figure 3.

We claim that there exists a k-sized set cover of U if and only if there is a set T of at most k edges such that $\frac{\min(c_{G+T}(a), c_{G+T}(b))}{\max(c_{G+T}(a), c_{G+T}(b))} \geq \tau = 1$. Note that in our construction, a is initially more central than b. As we cannot increase the closeness centrality of a vertex by adding edges, the task of making their ratio closer to 1 is initially analagous to decreasing b's centrality.

k-sized set cover $\implies \frac{\min(c_{G+T}(a), c_{G+T}(b))}{\max(c_{G+T}(a), c_{G+T}(b))} = 1$. For each set S_i in the cover, construct the edge bs_i. Now every element vertex must be adjacent to a set vertex in the cover (by definition of a cover), and then each of these set vertices has a newly constructed edge to b. Thus each element vertex is distance 2 from b, and k set vertices are distance 1 away from b. Now $c_{G+T}(b) = 2(n+k) + 1(2) + 1(k) + 2(m-k) + 2(n) = 4n + 2m + k + 2$. Similarly, $c_{G+T}(a) = 1(n+k) + 1(2) + 2(m) + 3(n) = 4n + 2m + k + 2$. Then the ratio of a and b's closeness centrality must be 1, as desired.

No k-sized set cover $\implies \frac{\min(c_{G+T}(a), c_{G+T}(b))}{\max(c_{G+T}(a), c_{G+T}(b))} < 1$. As previously stated, in order to maximize the closeness ratio of a and b, we decrease b's closeness centrality as much as possible with k edges.

Lemma 1. *3 In this construction, if there is no k-sized set cover, there is always a set $T \subseteq V^2 \backslash E$ of at most k edges, all incident on both $b \in V$ and any set vertex, whose addition to the graph optimally improves the closeness centrality of b.*

Proof. Deferred to appendix. $\square$

In other words, the set of edges which will optimally decrease b's closeness centrality are all incident on b and some set vertex. Therefore, the best b can reduce its closeness centrality is by connecting itself to set vertices. This method can make k set vertices distance 1 from b, and thus at most $n-1$ element vertices distance 2 from b, as there is no cover.

Claim. *In this construction, if there is no k-sized set cover, then there must exist some element vertex with no newly-added edge incident on it or any vertex of a set containing it.*

Proof. Deferred to appendix. $\square$

However, using the above claim, there must be some element which remains distance 3 from b. Then the closeness ratio of a, b is:

$$\frac{\min(c_{G+T}(a), c_{G+T}(b))}{\max(c_{G+T}(a), c_{G+T}(b))} < \frac{(n+k)+2+2m+3n}{2(n+k)+2+2m-k+2n+1} = \frac{4n+2m+k+2}{4n+2m+k+3} < 1$$

$\square$

We have shown that it is NP-hard to achieve a closeness centrality ratio of 1, but are smaller ratios achievable in polynomial time? By manipulating the size of the independent set connected to a (for $\tau = 1$, it was $n+k$ vertices) we can in fact prove a much stronger hardness result.

Theorem 2. *Closeness Ratio Improvement is NP-Hard for $\tau \in (\frac{1}{2}, 1)$.*

We will go through the construction of a Closeness Ratio Improvement instance from Set Cover, but the analysis of yes and no cases is deferred to the appendix.

Proof. Fix an arbitrary $\tau \in (\frac{1}{2}, 1)$. Consider an instance of Set Cover with m sets, n elements, and $k \in \mathbb{N}$ which satisfies $\frac{2m+4n+k}{1+2m+4n+k} \geq \tau^2$. This fraction converges to 1 as we increase m, n, k, and $\tau^2 < 1$ when $\tau \in (\frac{1}{2}, 1)$, so we should always be able to find m, n, k that satisfy this inequality. Use the same construction of G described in the proof of Theorem and depicted in Figure. However, now let X (the number of vertices in the independent set attached to $a \in V$) be an integer in the interval $(\frac{2+2m+3n-\tau(2+2m-k+2n+1)}{2\tau-1}, \frac{2+2m+3n-\tau(2+2m-k+2n)}{2\tau-1}]$. We know such an integer always exists by Lemma.

We claim that there exists a k-sized set cover of U if and only if there is a set T of at most k edges such that $\frac{\min(c_{G+T}(a), c_{G+T}(b))}{\max(c_{G+T}(a), c_{G+T}(b))} \geq \tau$. $\square$

4 $\frac{1}{2}$-Approximation

To restate the problem, we are given a graph G, vertices a, b and an edge budget k. In the decision variant of this problem, we are given a target ratio of τ, but generally we would like to make the ratio of a and b's closeness centrality in the augmented graph as close to 1 as possible.

Observe that if a and b are ever connected, the worst possible ratio is $\frac{1}{2}$. This happens when every vertex other than a, b in the graph is directly connected to b, and a's only neighbor is b itself (see Figure). For large n, b's closeness centrality is 1 times n vertices, while a's is 2 times n vertices, giving us closeness ratio of $\frac{1}{2}$. If a was closer than b to any vertex, then the numerator would increase and denominator would decrease, giving us a better ratio. Similarly, if any vertex was further than distance 1 from b, we would get a better ratio (think $\frac{2}{3}$, $\frac{3}{4}$, etc.). In short, with a and b connected, we guarantee a closeness ratio at least $\frac{1}{2}$.

Now consider the algorithm, given $k \geq 1$ edges, which simply adds the edge ab (if it does not already exist in the graph). As we just argued, our closeness ratio is now at least $\frac{1}{2}$ (though it might be better). This agrees very well with our hardness results, as we showed getting a target ratio greater than $\frac{1}{2}$ in all cases is NP-hard, but getting a target ratio of exactly $\frac{1}{2}$ is easy (in fact, it only requires one edge).

Furthermore, this strategy is a $\frac{1}{2}$-approximation for CRI. Whatever ratio an optimal efficient

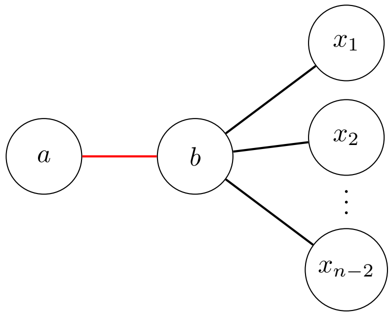

Figure 2: Worst case of closeness ratio when a and b are connected.

algorithm (which we do not know how to find, as this problem is NP-hard) would achieve on an instance of CRI, our algorithm will never get a ratio less than half of that value. This follows from the observation that, simply by our definition of closeness ratio, a ratio better than 1 is not possible. Thus is we guarantee a ratio of $\frac{1}{2}$, we guarantee that our algorithm produces no worse than $\frac{1}{2}$ of the optimal ratio.

An important distinction in this section is the difference between *target* ratio and *approximation* ratio. The approximation ratio is the ratio of the closeness ratio our algorithm achieves over the closeness ratio achieved by an optimal algorithm. Connecting ab immediately achieves a target ratio of $\frac{1}{2}$, and as the best possible closeness ratio achievable is 1, this strategy necessarily forms *at least* at $\frac{1}{2}$ approximation. However, it is possible that this strategy is in fact better than a $\frac{1}{2}$-approximation. That is, all the problem instances where this strategy gets a ratio of $\frac{1}{2}$, an optimal strategy actually cannot get a ratio of 1 (meaning then the ratio of our ratio over optimal ratio is greater than $\frac{1}{2}$).

TIGHTNESS OF 1/2 EXAMPLE.

5 Hardness of Approximation

Theorem 3. *hardness of approximation.*

References

[1] F. Adriaens and A. Gionis. Diameter minimization by shortcutting with degree constraints. 2022. doi: 10.48550/arXiv.2209.00370.

[2] A. Bashardoust, S. Friedler, C. Scheidegger, B. D. Sullivan, and S. Venkatasubramanian. Reducing access disparities in networks using edge augmentation. *FAccT '23*, pages 1635–1651, 2023.

[3] A. Bhaskara, A. Crane, S. Jain, M. M. H. U. Mazumder, B. D. Sullivan, and P. Yalamanchili. Optimizing information access in networks via edge augmentation. 2024. URL `https://arxiv.org/abs/2407.02624`.

[4] D. Bilò, L. Gualà, and G. Proietti. Improved approximability and non-approximability results for graph diameter decreasing problems. *Theoretical Computer Science*, 417: 12–22, 2012.

[5] P. Crescenzi, G. D'Angelo, L. Severini, and Y. Velaj. Greedily improving our own closeness centrality in a network. *ACM Transcations on Knowledge Discovery from Data*, 11(1):1–32, 2016.

[6] danah boyd, K. Levy, and A. Marwick. The networked nature of algorithmic discrimination. *Data and Discrimination: Collected Essays*, pages 43–57, 2014.

[7] E. Demaine and M. Zadimoghaddam. Minimizing the diameter of a network using shortcut edges. *Algorithm Theory - SWAT 2010*, 6139:420–431, 2010.

[8] B. Fish, A. Bashardoust, D. Boyd, S. A. Friedler, C. Scheidegger, and S. Venkatasubramanian. Gaps in information access in social networks. 2019. doi: 10.1145/3308558.3313680.

[9] C.-L. Li, S. T. McCormick, and D. Simchi-Levi. On the minimum-cardinality-bounded-diameter and the bounded-cardinality-minimum-diameter edge addition problems. *Operations Research Letters*, 11:303–308, 1991.

[10] S. Medya, A. Silva, A. Singh, P. Basu, and A. Swami. Group centrality maximization via network design. pages 126–134, 2018. ISBN 978-1-61197-532-1.

[11] A. Meyerson and B. Tagiku. Minimizing average shortest path distances via shortcut edge addition. *APPROX 2009*, pages 272–285, 2009.

A Appendix Section

Claim. *Given a graph $G = (V, E)$, $v \in V$, $k \in \mathbb{N}$, and a set of edges $S \subseteq V^2 \backslash E$ of size at most k where at least one edge in S is not incident on v, there must always exist a set $T \subseteq V^2 \backslash E$ such that $|T| \leq |S|$ and $c_{G+T}(v) < c_{G+S}(v)$.*

Proof. Define the edge in S not incident on v as xy, using its endpoints $x, y \in V$. Without loss of generality, suppose $d_{G+S}(v, x) \leq d_{G+S}(v, y)$. For any vertex $z \in V$, we have two cases: either the shortest path from v to z in $G + S$ uses the edge xy or it does not. In either case, with the edge set $T = (S \backslash \{xy\}) \cup \{vy\}$, we will show that $d_{G+T}(v, z) \leq d_{G+S}(v, z)$.

If the path from v to z in $G + S$ uses the edge xy, then $d_{G+S}(v, z)$ can be expressed as $d_{G+S}(v, x) + 1 + d_{G+S}(y, z)$. Note that the assumption about the relative distances of x, y to v guarantees that the shortest path from v to z crosses over xy instead of yx, as $d_{G+S}(v, x) \leq d_{G+S}(v, y) < d_{G+S}(v, y) + d_{G+S}(y, x)$. Then in $G + T$,

$$d_{G+T}(v, z) = d_{G+T}(v, y) + d_{G+T}(y, z) = 1 + d_{G+S}(y, z)$$

$$< d_{G+S}(v, x) + 1 + d_{G+S}(y, z) = d_{G+S}(v, z).$$

If the path from v to z in $G + S$ does not use the edge xy, then removing it and replacing it with a new edge cannot make the distance between v and z any greater, it can only decrease it. Thus in either case, $d_{G+T}(v, z) \leq d_{G+S}(v, z)$. As this is true for every vertex $z \in V$, and this inequality is strict when $z = y$, $c_{G+T}(v) < c_{G+S}(v)$, as desired. □

Lemma 2. *Given a graph $G = (V, E)$, $v \in V$ and $k \in \mathbb{N}$, there is always a set $T \subseteq V^2 \backslash E$ of at most k edges, all incident on v, whose addition to the graph optimally improves the closeness centrality of v.*

Proof. Assume not. That is, suppose the set S of k edges which optimally improved the closeness centrality of v were not all incident on v. This means at least one edge in S is not incident on v, implying the existence of some set T of size at most k such that $c_{G+T}(v) < c_{G+S}(v)$ (by Claim), contradicting the assumption that S optimally improved the closeness centrality of v. □

Claim. *Given a graph of the construction provided in Figure* **??** *derived from a set cover instance with m sets, n elements and $k \in \mathbb{N}$, if there is no k-sized set cover, then there must exist some element vertex with no newly-added edge incident on it or any vertex of a set containing it.*

Proof. Assume not. That is, every element vertex *does* have a new edge incident on it or one of the set vertices containing it. For any element vertex that is the endpoint of a new edge, simply reconstruct this edge to one of the set vertices adjacent to this element. Now we have identified k edges incident only on set vertices such that every element is the neighbor of at least one such set vertex, and thus we have identified a k-sized set cover (a contradiction). □

Claim. *Given a graph of the construction provided in Figure* **??** *derived from a set cover instance of m sets, n elements and $k \in \mathbb{N}$ with no k-sized set cover, given a set of at most k edges $S \subseteq V^2 \backslash E$, all incident on $b \in V$, if at least one edge is not incident on a set vertex, then there must exist a set $T \subseteq V^2 \backslash E$ such that $|T| = |S|$ and $c_{G+T}(b) \leq c_{G+S}(b)$.*

Proof. Suppose $S \subseteq V^2 \backslash E$ is a set of k edges which contains only edges incident on b, but at least one edge is not incident on a set vertex. Given b is connected to a and c, this implies that there must exist an edge in S which is connected a vertex x_i in the independent set or an element vertex e_i.

In the first case, this edge reduces the closeness centrality of b by at most 1 — it has changed $d(x_i, b)$ from 2 to 1, but not made b closer to any other vertices. If we instead connect b to a set vertex s_i, then $d(s_i, b)$ has been changed from 2 to 1, meaning this edge reduces the closeness centrality of b by at least as much as the edge $b - x_i$.

If the edge not incident on a set vertex instead connects b to an element vertex e_i, this edge reduces the closeness centrality of b by at most 2 — it has changed $d(e_i, b)$ from 3 to 1, but not made b closer to any other vertices. As there is no k-sized set cover, Claim tells us there must be some element vertex e_j with no new edges incident on it or any set containing it. So instead connect b to a set vertex s_j, where $e_j \in S_j$. then both $d(s_j, b)$ and $d(e_j, b)$ have been changed from 2 to 1, meaning $b - s_j$ reduces the closeness centrality of b by at least as much as the edge $b - e_i$.

If we define the set T as the set S, but with the above case-dependent substitutions made, then we have that $|T| = |S|$ and $c_{G+T}(b) \leq c_{G+S}(b)$, as desired. $\square$

Lemma 3. *Given a graph of the construction provided in Figure* **??** *derived from a set cover instance of m sets, n elements and $k \in \mathbb{N}$ with no k-sized set cover, there is always a set $T \subseteq V^2 \backslash E$ of at most k edges, all incident on both $b \in V$ and any set vertex, whose addition to the graph optimally improves the closeness centrality of b.*

Proof. From Lemma 2, we know that there exists a set T of k edges which optimally improves the closeness centrality of b such that all edges in T are incident on b itself. What remains to be shown is that there exists such a set of edges where they are all incident on b *and* some set vertex s_i.

Assume not. That is, all the sets of edges which optimally improve b's closeness centrality contain only edges incident on b, but at least one edge not incident on a set vertex. Now consider one such set of edges S. Claim gives us a set S' with $c_{G+S'}(b) \leq c_{G+S}(b)$. If S' still contains an edge incident on b, but not a set vertex, we can reapply Claim to get a new set that improves b's closeness centrality just as well. In fact, we can keep applying this claim until we have a set of k edges T such that $|T| = |S|$ and $c_{G+T}(b) \leq c_{G+S}(b)$ and all edges in T are incident on both b and any set vertex. However, we assumed that all the sets of edges which optimally improve b's closeness centrality contain only edges incident on b, but must have at least one edge not incident on a set vertex, a contradiction. $\square$

Lemma 4. *For $m, n, k \in \mathbb{N}$ and $\tau \in \left(\frac{1}{2}, 1\right]$, there exists an $X \in \left(\frac{2+2m+3n-\tau(2+2m-k+2n+1)}{2\tau-1}, \frac{2+2m+3n-\tau(2+2m-k+2n)}{2\tau-1}\right]$ such that $X \in \mathbb{N}$.*

Proof. Taking the difference of the bounds of the interval $\left(\frac{2+2m+3n-\tau(2+2m-k+2n+1)}{2\tau-1}, \frac{2+2m+3n-\tau(2+2m-k+2n)}{2\tau-1}\right]$, we get that the length of the interval is $\frac{\tau}{2\tau-1}$, which is greater than or equal to 1 for all $\tau \in \left(\frac{1}{2}, 1\right]$. Therefore, the length of our interval is bounded below by 1, and so there must exist some integer X within this interval. Furthermore, as the numerator and denominator of the bounds of this interval are positive when $m, n, k \in \mathbb{N}$ and $\tau \in \left(\frac{1}{2}, 1\right]$, $X \in \mathbb{N}$ as desired. $\square$

Theorem 4. *Closeness Ratio Improvement is NP-Hard for $\tau \in (\frac{1}{2}, 1)$.*

Proof. (Concluding the argument started earlier, with construction given).

k-sized set cover $\implies \frac{\min(c_{G+T}(a), c_{G+T}(b))}{\max(c_{G+T}(a), c_{G+T}(b))} \geq \tau$. If there is a set cover, follow the same protocol as in the proof of Theorem **??** (connect b to the k set vertices representing sets in the cover). In this case, $c_{G+T}(a) = X + 2 + 2m + 3n$ and $c_{G+T}(b) = 2X + 2 + 2m - k + 2n$.

We want to show that the ratio we achieve is greater than or equal to τ, but we must still ensure that our ratio is less than or equal to 1, else we violate our min/max definition of closeness ratio. Note that this was not a conceren when $\tau = 1$, because we showed our ratio when there was a set cover was 1 exactly. So we consider the case where $c_{G+T}(b) \geq c_{G+T}(a)$ and the case where $c_{G+T}(b) < c_{G+T}(a)$, and show that in both cases the ratio of $\frac{\min(c_{G+T}(a), c_{G+T}(b))}{\max(c_{G+T}(a), c_{G+T}(b))} \geq \tau$.

If $c_{G+T}(b) \geq c_{G+T}(a)$, then

$$\frac{\min(c_{G+T}(a), c_{G+T}(b))}{\max(c_{G+T}(a), c_{G+T}(b))} = \frac{X + 2 + 2m + 3n}{2X + 2 + 2m - k + 2n}$$

$$\geq \frac{(\frac{2+2m+3n-\tau(2+2m-k+2n)}{2\tau-1}) + 2 + 2m + 3n}{2(\frac{2+2m+3n-\tau(2+2m-k+2n)}{2\tau-1}) + 2 + 2m - k + 2n} \geq \tau$$

Alternatively, if $c_{G+T}(b) < c_{G+T}(a)$, then

$$\frac{\min(c_{G+T}(a), c_{G+T}(b))}{\max(c_{G+T}(a), c_{G+T}(b))} = \frac{2X + 2 + 2m - k + 2n}{X + 2 + 2m + 3n}$$

$$> \frac{2(\frac{2+2m+3n-\tau(2+2m-k+2n+1)}{2\tau-1}) + 2 + 2m - k + 2n}{(\frac{2+2m+3n-\tau(2+2m-k+2n+1)}{2\tau-1}) + 2 + 2m + 3n} = \frac{(2-2\tau) + 2m + 4n + k}{\tau + 2m\tau + 4n\tau + k\tau}$$

As $\tau \in (\frac{1}{2}, 1)$, $2 - 2\tau > 0$, implying:

$$\frac{(2-2\tau) + 2m + 4n + k}{\tau + 2m\tau + 4n\tau + k\tau} > \frac{2m + 4n + k}{\tau + 2m\tau + 4n\tau + k\tau} = \frac{1}{\tau}(\frac{2m + 4n + k}{1 + 2m + 4n + k})$$

We chose our Set Cover instance such that $\frac{2m+4n+k}{1+2m+4n+k} \geq \tau^2$. Thus we have that:

$$\frac{1}{\tau}(\frac{2m + 4n + k}{1 + 2m + 4n + k}) \geq \frac{1}{\tau}(\tau^2) = \tau$$

Thus in either case, if there is a set cover of U, we can add k edges to G to get a closeness ratio greater than or equal to τ.

No k-sized set cover $\implies \frac{\min(c_{G+T}(a), c_{G+T}(b))}{\max(c_{G+T}(a), c_{G+T}(b))} < \tau$. If there is no k-sized set cover, we use the same analysis from the proof of Theorem **??** to argue that the best way b can reduce its closeness centrality is by connecting itself to set vertices. Furthermore, this method makes k set vertices distance 1 from b, and thus $n - 1$ element vertices distance 2 from b, but (using Claim) there must be some element which remains distance 3 from b. Then the closeness ratio of a, b is:

$$\frac{\min(c_{G+T}(a), c_{G+T}(b))}{\max(c_{G+T}(a), c_{G+T}(b))} = \frac{X + 2 + 2m + 3n}{2X + 2 + 2m - k + 2n + 1}$$

$$< \frac{(\frac{2+2m+3n-\tau(2+2m-k+2n+1)}{2\tau-1}) + 2 + 2m + 3n}{2(\frac{2+2m+3n-\tau(2+2m-k+2n+1)}{2\tau-1}) + 2 + 2m - k + 2n + 1} = \tau$$

□