

Summer Report On The Capacitated K-Center Problem

Anya Chaturvedi
Ketaki Vaidya

K. R. Prajwal
Sagar Sahni

July 1, 2016

1 Introduction to our Problem

The capacitated K-center problem is basically a facility location problem, where one is asked to locate K facilities in a graph and to assign vertices to these facilities. In doing this we minimize the maximum distance from a vertex to the facility to which it is assigned while keeping in mind that each facility may be assigned at most L vertices including itself. This problem is known to be NP-hard.

We show our attempts at solving this problem, counter examples as learnt by us and few basic implementations of the same. The implementations include finding the optimal using the integer program and a partially accurate local search algorithm.

2 A Look At The K-center problem

The problem taken up by us is actually a generalization of the K-center problem which we will explain here. Given n vertices with specified distances, one wants to build k facilities at different vertices and minimize the maximum distance of a vertex to a facility. This problem is NP-hard. An approximation algorithm with a factor of η , for a minimization problem, is a polynomial time algorithm that guarantees a solution with cost at most η times the cost of an optimal solution. For the basic K-center methods have been presented for obtaining an approximation factor of 2. Given a complete undirected graph $G = (V, E)$ with distances $d(v_i, v_j) \in N$ satisfying the triangle inequality, find a subset $S \subseteq V$ with $|S| = k$.

Input:

1. The vertices must be in a metric space, or in other words a complete graph that satisfies the triangle inequality.
2. An upper bound on the number of centers K.

Formal Definition:

$$\min_{S \subseteq V} \max_{v \in V} \min_{s \in S} d(v, s)$$

where d is the distance function.

This is much similar to the problem of placing k disks such that all points are covered in the set V and thus finding the minimum radius for the disks.

3 The Capacitated K-Center

The capacitated K-center problem is nothing but a generalization of the K-center problem. We have to output a set of at most K centers, as well as an assignment of vertices to centers. No

more than L vertices may be assigned to a single center. Under these constraints, we wish to minimize the maximum distance between a vertex u and its assigned center $\varphi(u)$.

Input:

1. The vertices must be in a metric space, or in other words a complete graph that satisfies the triangle inequality.
2. An upper bound on the number of centers K .
3. A maximum load L .

Formal Definition:

$$\min_{S \subseteq V} \max_{u \in V} d(u, \varphi(u))$$

such that,

$$|\{u | \varphi(u) = v\}| \leq L \forall v \in S,$$

where,

$$\varphi : V \rightarrow S.$$

The first polynomial time approximation algorithm for this problem was with an approximation factor of 10. Later a 6-approximation was found which is the best up till date while a 5-approximation works when we can assign multiple centers at a single vertex without including the vertex in L . This was our problem for the summer.

Initially we went through the 2.6 and 9.3 sections of Williamson Schmoys which solve the min degree spanning tree problem which is also NP-hard and got to know of methods like local search used to create approximation algorithms.

Certifying a graph for an algorithm not only has to ensure that we can show whether a solution is possible for G or not it also involves showing that if a solution is not possible in G then it is not even possible in G^2, G^3, \dots

4 Checking the Feasibility of the Input Graph

We say that, if we solve the capacitated K -center problem then it will be equivalent to getting the minimum value of W for which $G(W)$ has a capacitated dominating set of size k , where $G(W)$ refers to the graph having all edges with edge weight less than or equal to W .

We can easily set up such an equivalence if we slightly modify our problem. Given a graph $G = (V, E)$ and a capacity function $c : V \rightarrow N$, $S \subseteq V$ is a capacitated dominating set if there exists a mapping $f : V \setminus S \rightarrow S$ (domination mapping) such that:

1. $\forall v \in V \setminus S, \forall s \in S, f(v) = s \implies (v, s) \in E$
2. $\forall s \in S, |\{v \in V \setminus S | f(v) = s\}| \leq c(s)$

In the new version of the k -center problem, the goal is to get a set S of k vertices and an assignment $h : V \setminus S \rightarrow S$ of every vertex (in $V \setminus S$) to an open center such that the longest distance between a vertex and a center it is assigned to is minimum and no facility is assigned more vertices than its capacity. Note that, the assignment is not for all vertices in V . We are making the assumption that every center serves itself and the capacity ($L(v)$) of a vertex $v \in V$ is the maximum number of clients it can serve (excluding itself).

Now we can say that, solving the capacitated k -center problem is equivalent to getting the minimum value of W for which $G(W)$ has a capacitated dominating set of size k . $S \subseteq V$ is an

optimal solution to the capacitated K-center problem if and only if $w(S)$ is the minimum value of W for which the graph $G(W)$ has a capacitated dominating set of size k . It is also easy to verify that a capacitated dominating set of size k in $G(w(S))^i$, for some i and some optimal solution S , forms an i -approximation for the k -center problem.

But the capacitated dominating set problem is an NP-complete problem. So, we cannot get a capacitated dominating set of size k in polynomial time, unless $P = NP$. Similar to the uncapacitated version the capacitated version too is hard to approximate within a factor of 2.

4.1 Verification using a Network Flow

Given a graph $G = (V, E)$ and a capacity function $c : V \rightarrow N$, checking whether $S \subseteq V$ is a capacitated dominating set can be viewed as a network problem. Construct a directed graph $G_S = (V_S, E_S)$ as follows:

1. $V_S = V \cup \{s, t\}$
2. $E_S = \{(s, v_i) | v_i \in V \setminus S\} \cup \{(s_i, t) | s_i \in S\} \cup \{(v_i, s_i) | v_i \in V \setminus S \wedge s_i \in S \wedge (v_i, s_i) \in E\}$

Let $c_S : E_S \rightarrow N$ be the capacity function such that,

$$c_S(e_{uv}) = \begin{cases} c(u), & v = t \\ 1, & \text{otherwise} \end{cases}$$

In the final graph G_S we add two more vertices s (source) and t (sink) to the vertex set of G . We add directed edges from s to all vertices in $V \setminus S$, from each vertex in $V \setminus S$ to its neighbours in S and each vertex in S to t . We take the capacity function to be equal to the capacity of the source vertex for edges incident on t and to be equal to 1 for every other edge in E_S .

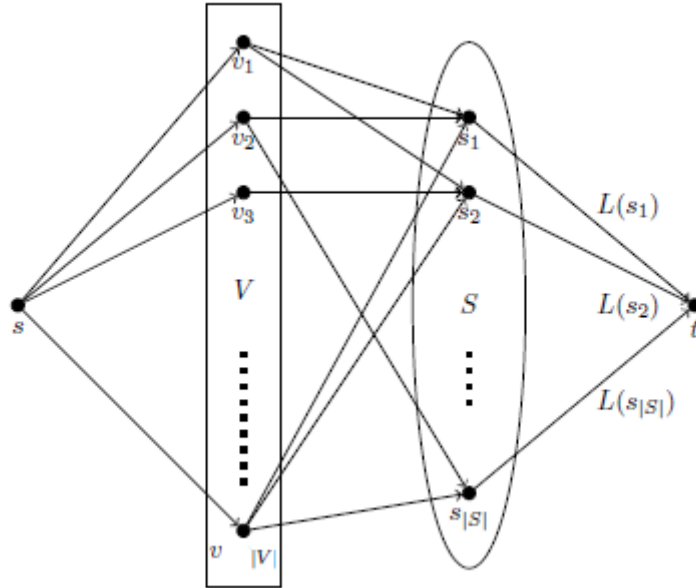


Figure 1: The flow network

Now it is easy to verify that, S is a capacitated dominating set in G if and only if G_S has a maximum coverage of $|V \setminus S|$ from s to t equal to the total vertices in the graph.

5 The Integer Program

Linear programming is a mathematical technique for maximizing or minimizing a linear function of several variables such as output or cost. The decision variables with respect to which the optimum is obtained can come out to be fractional or any real number. An integer programming problem is a mathematical optimization or feasibility program in which some or all of the variables are restricted to be integers which is the major difference it has from linear programming. Ours is a problem where we wish to minimize the maximum distance between a vertex u and its assigned center thus our integer program is framed as:

Variables:

1. ***cen*** - an integer dictionary which specifies the centers among the vertices where if $cen[i] = 1$ then the i^{th} index vertex is taken to be a center otherwise not.
2. ***assignments*** - dictionary $assignments[i][j] = 1$ will represent that the j^{th} index vertex is assigned to the center created at the i^{th} index vertex, we initialize it with the adjacency matrix
3. ***N*** - represents the total number of vertices
4. ***adj*** - a dictionary which stores the adjacency matrix of the graph

```

1  import snap
2  from pulp import *
3
4  def randomGraph(n,edges):
5      UGraph = snap.GenRndGnm(snap.PUNGraph, n, edges)
6      adj = dict()
7      for x in range(n):
8          for y in range(n):
9              adj[(x,y)] = 0
10     for EI in UGraph.Edges():
11         adj[(EI.GetSrcNId(),EI.GetDstNId())] = 1
12         adj[(EI.GetDstNId(), EI.GetSrcNId())] = 1
13     for i in xrange(n):
14         adj[(i,i)] = 1
15     return adj
16
17  N=input("No. of vertices (n<4039):")
18  edges=input("No. of edges:")
19  adj = randomGraph(N,edges)
20  edges = sum([sum(x) for x in adj])/2
21  L = N/10
22
23  # declare your variables
24  cen = dict(zip([i for i in range(N)],[0 for i in range(N)]))
25  cen = LpVariable.dicts("cen",cen, 0, 1,LpInteger)
26
27  assignments = {}
28  for i in range(N):
29      for j in range(N):
30          assignments[(i,j)] = adj[(i,j)]
31
32  assignments = LpVariable.dicts("assignments",assignments,0,1)
33

```

Figure 2: Code

```

34 # defines the problem
35 prob = LpProblem("problem", LpMinimize)
36
37 # defines the objective function to minimize : sum(y)
38 prob += lpSum([cen[i] for i in range(N)])
39
40 ##### Constraint definitions #####
41
42 # sum(x_ij) <= L*y[i]
43 for i in xrange(N):
44     prob += lpSum([assignments[(i,j)] for j in xrange(N)]) <= L*cen[i]
45
46 # sum(x_ij) = 1 for all j
47 for j in xrange(N):
48     prob += lpSum([assignments[(i,j)] for i in xrange(N)]) == 1
49
50 # y_i >= x_ij for all i,j
51 for i in xrange(N):
52     for j in xrange(N):
53         prob += assignments[(i,j)] <= cen[i]
54 #the assignment can be done only when the edge is present
55 for i in xrange(N):
56     for j in xrange(N):
57         prob += assignments[(i,j)] <= adj[(i,j)]
58
59 # solve the problem using GLPK
60 status = prob.solve(GLPK(msg=0))
61 print LpStatus[prob.status]
62 list_var=prob.variables()
63 list_cen=list_var[(len(list_var)-N):]
64 count=0
65 for v in list_cen:
66     if(v.varValue==1.0):
67         count+= 1
68 print N,edges,count

```

Figure 3: Code

Objective Function:

$$z = \sum_{i=1}^N cen[i]$$

Constraints:

1. $\forall i \sum_{j=1}^N assignments[i][j] \leq L * cen[i]$
...makes sure no more than L vertices are assigned to a center
2. $\forall j \sum_{i=1}^N assignments[i][j] = 1$
...a vertex can be assigned only to a single center

3. $\forall i \forall j \text{ assignments}[i][j] \leq \text{cen}[i]$
... a vertex can only be assigned to a second vertex if the second vertex is a center
4. $\forall i, \forall j \text{ assignments}[i][j] \leq \text{adj}[i][j]$
... assignment of vertices can only be done if the edge is present in the original graph

6 Local Search Algorithm

Local search can be used on problems that can be formulated as finding a solution maximizing a criterion among a number of candidate solutions. A local search algorithm starts from a candidate solution and then iteratively moves to a neighbour solution. We move from solution to solution in the space of candidate solutions (the search space) by applying local changes, until a solution optimal is found. We have used a local search algorithm which was formed by Aounon. It doesn't work on a set of graphs but we get the solution for some general graphs we tested on. Here, we start with a set S of random k vertices in graph G . Let S be the set of all k vertex sets which are formed by replacing one vertex in S by a vertex in V (one-swap vertex sets). We look for the set in S which maximizes the number of dominated vertices. If the optimum set dominates more vertices than S then we set S to that set and continue the process till we have reached a local optima. We do the assignments through a network flow graph.

Algorithm:

```

 $S \leftarrow v_1, v_2, \dots, v_k$ 
 $S' \leftarrow \text{one-swap vertex sets of } S_g$ 
 $V_{max} \leftarrow \arg \max_{V \in S} \text{MaxFlow}(G_V, c_V)$ 
while  $\text{MaxFlow}(G_{V_{max}}; c_{V_{max}}) > \text{MaxFlow}(G_S; c_S)$  do
     $S \leftarrow V_{max}$ 
     $S' \leftarrow \text{one-swap vertex sets of } S$ 
     $V_{max} \leftarrow \arg \max_{V \in S} \text{MaxFlow}(G_V, c_V)$ 
end while
if  $\text{MaxFlow}(G_S, c_S) = n - k$  then
    for all  $v \in V$  do
         $h(v) = s$ : if  $\text{fmax}(v, s) = 1$ 
    end for
    return  $S, h$ 
end if

```

```

1  import networkx as nx
2  import snap
3  from random import shuffle
4  from collections import defaultdict
5
6  #Generating a random graph using snap
7  def randomGraph(n, edges):
8      UGraph = snap.GenRndGnm(snap.PUNGraph, n, edges)
9      adj = dict()
10     for x in range(n):
11         for y in range(n):
12             adj[(x,y)] = 0

```

Figure 4: Code

```

13     for EI in UGraph.Edges():
14         adj[(EI.GetSrcNid(),EI.GetDstNid())] = 1
15         adj[(EI.GetDstNid(), EI.GetSrcNid())] = 1
16     for i in xrange(n):
17         adj[(i,i)] = 1
18     for x in range(n):
19         for y in range(n):
20             print adj[(x,y)],
21         print
22     return adj
23
24 #Storing the random graph in G
25 def getGraph(N,adj):
26     G = nx.Graph()
27     G.add_nodes_from([x for x in xrange(N)])
28     for i in xrange(N):
29         for j in xrange(i,N):
30             if adj[(i,j)]:
31                 G.add_edge(i,j)
32     return G
33
34 #Obtaining a random set S of initial k centers
35 def gets(G,k):
36     nodes = G.nodes()
37     shuffle(nodes)
38     return nodes[:k]
39
40 #making FG a directed flowgraph to give as input to max flow functions ( s-->V-->S-->t )
41 def getFlowGraph(G, S, L):
42     FG = nx.DiGraph()
43     N = len(G.nodes())
44     #for common nodes in V and S make separate keys for S
45     for x in G.nodes():
46         FG.add_node(x)
47         if x in S:
48             FG.add_node(x + N)
49
50     FG.add_nodes_from(['s','t'])
51     edges = set(G.edges())
52     for s in S:
53         for v in G.nodes():
54             if (min(s,v),max(s,v)) in edges:
55                 FG.add_edge(v, s + N, capacity=1)
56     for v in G.nodes():
57         FG.add_edge('s',v,capacity=1)

```

Figure 5: Code


```

58         for s in S:
59             FG.add_edge(s + N, 't', capacity=L)
60         return FG
61
62 #Obtaining the set Vmax with one swap from S to get the max flow value set
63 def getVmax(G, S, L):
64     V = G.nodes()
65     # flow[i][j] = max-flow in (S, V) when S[i] is swapped with V[j]
66     flow = [[None for _ in xrange(len(V))] for _ in xrange(len(S))]
67     sets = set(S)
68     for i in xrange(len(S)):
69         for j in xrange(len(V)):
70             if V[j] not in sets:
71                 temp = S[i]
72                 S[i] = V[j]
73                 flow_graph = getFlowGraph(G, S, L)
74                 flow[i][j] = nx.maximum_flow_value(flow_graph, 's', 't')
75                 S[i] = temp
76
77     max_value = 0
78     swap_pair = None
79     for i in xrange(len(S)):
80         for j in xrange(len(V)):
81             if max_value < flow[i][j]:
82                 swap_pair = (i,j)
83                 max_value = flow[i][j]
84
85     i,j = swap_pair
86     S[i] = V[j]
87     return S
88
89 #changing S and Vmax by one swaps so as to achieve the max flow possible after each obtained S
90 def doOneSwaps(G, S, L):
91     Vmax = getVmax(G, S, L)
92     while nx.maximum_flow_value(getFlowGraph(G, Vmax, L), 's', 't') > nx.maximum_flow_value(getFlowGraph(G, S, L), 's', 't'):
93         S = Vmax
94         Vmax = getVmax(G, S, L)
95     return S
96
97 N, M = 10,30
98 adj = randomGraph(N,M)
99 k, L = 5,3
100 G = getGraph(N,adj)

```

Figure 6: Code

```

100 S = getS(G,k)
101 S = doOneSwaps(G, S, L)
102 H = defaultdict(list)
103 #max flow value cannot be greater than the number of nodes in the graph or set V
104 if nx.maximum_flow_value(getFlowGraph(G, S, L), 's', 't') == N:
105     _,flows = nx.maximum_flow(getFlowGraph(G, S, L), 's', 't')
106     edges = set(G.edges())
107     #if a flow exists along a particular path assigning v to that vertex in S
108     for v in G.nodes():
109         for s in S:
110             if (min(v,s),max(v,s)) in edges:
111                 unitflow = False
112                 try:
113                     if flows[max(v,s + N)][min(v,s + N)] == 1:
114                         unitflow = True
115                 except KeyError:
116                     try:
117                         if flows[min(v,s + N)][max(v,s + N)] == 1:
118                             unitflow = True
119                     except KeyError:
120                         pass
121                 if unitflow:
122                     H[s].append(v)
123 else:
124     print "Failed!"
125 print S
126 print H

```

Figure 7: Code

7 Star Graph Generation

In an attempt to try our programs on a more clustered data set, this program was implemented such that in this star graphs are formed which are sparsely connected to each other by edges which are added to the graph based on a probability p .

```

1  import snap
2  import random
3  import networkx as nx
4  from collections import namedtuple
5
6  class Graph(object):
7      def __init__(self, specs):
8          self.k = specs.k
9          self.l = specs.l
10
11         if specs.struct == 'stars':
12             self.N = self.k*self.l
13             self.p = specs.p# supplementary edge probability
14             self.starGraph()
15
16         elif specs.struct == 'random':
17             self.N = specs.N
18             self.edges = specs.M
19             self.randomGraph()
20
21         else:
22             raise ValueError('Invalid specifications!')
23
24         self.nxGraph = self.NXify()
25
26     def initAdj(self):
27         self.adj = dict()
28         for x in range(self.N):
29             for y in range(self.N):
30                 self.adj[(x,y)] = 0
31         for i in xrange(self.N):
32             self.adj[(i,i)] = 1
33

```

Figure 8: Code

```

34     def randomGraph(self):
35         UGraph = snap.GenRndGnm(snap.PUNGraph, self.N, self.edges)
36         self.initAdj()
37
38         for EI in UGraph.Edges():
39             self.adj[(EI.GetSrcNId(),EI.GetDstNId())] = 1
40             self.adj[(EI.GetDstNId(), EI.GetSrcNId())] = 1
41
42     def starGraph(self):
43         self.initAdj()
44         non_centers = set([x for x in xrange(self.N)])
45         star_centers = random.sample(list(non_centers), self.k)
46         for x in star_centers: non_centers.remove(x)
47         non_centers = list(non_centers)
48         random.shuffle(non_centers)
49         star_clients = [non_centers[i:i + self.l - 1] for i in xrange(0, len(non_centers), self.l)]
50
51         for center, clients in zip(star_centers, star_clients):
52             for client in clients:
53                 self.adj[(center,client)] = self.adj[(client, center)] = 1
54
55         for x in range(self.N):
56             for y in range(x + 1,self.N):
57                 if self.adj[(x,y)] != 1:
58                     self.adj[(x,y)] = int(random.random() < self.p)
59                     self.adj[(y,x)] = int(random.random() < self.p)
60
61
62     #Storing the random graph in G
63     def NXify(self):
64         G = nx.Graph()
65         G.add_nodes_from([x for x in xrange(self.N)])
66         for i in xrange(self.N):
67             for j in xrange(i,self.N):
68                 if self.adj[(i,j)]:
69                     G.add_edge(i,j)
70
71     return G

```

Figure 9: Code

8 Observations made

This section basically focuses on all our trials and errors we had during our intern on this topic along with few of the concepts learnt which provide the base for these trials.

8.1 Lower Bound on K

Lets say we make some partition on our input graph such that each partition has two types of vertices. Inner and boundary, the inner vertices are the ones within the particular partition while the boundary vertices are inner vertices which even have edges to other partitions. Let the i^{th} partition be represented by its inner vertex C_i and its boundary vertices as a function of C_i as $B(C_i)$.

Since we know K vertices can cover up to L vertices so the max vertices than can be covered is $V = K * L$. Hence K can be taken to be approximately $\lceil V/L \rceil$. Thus if we try forming a lower bound so as to claim that the graph requires atleast these many vertices as equal to the bound then it might be of help.

Here, in the partitioning as described above, for the vertices other than the boundary vertices we can say that they can only be covered by vertices of their own partition thus we need to open centers in the partition itself for such vertices. Thus, if we say there are p partitions then $\sum_{i=1}^p \lceil (C_i - B(C_i))/L \rceil$ vertices are required atleast to cover the graph and hence this comes to be a lower bound.

This has been presented in a research paper already but the fact that we are not really involving the boundary vertices did not sink in and though it did not work out but we toyed for some time with ideas of somehow involving the boundary vertices. We thought with the current vertices the remaining capacity i.e. $\lceil (C_i - B(C_i))/L \rceil * L - (C_i - B(C_i))$ we may be able to cover some or even all of the boundary vertices. Through this we can decide opening more centers or shifting the centers so as to cover more boundary vertex. But it always depends upon the connectivity of the particular graph and what kind of partition we take and that was where we left this.

8.2 Our Attempts

- In our attempt we start by applying the 2-approximation K -center algorithm on the given graph. We then tried to apply what we had learnt by reading the $\text{OPT} + 1$ local search algorithm of the min-degree spanning tree from section 2.6 of Williamson & Schmoy. Effectively, if possible we transferred the load from the highest capacity nodes to lower ones through a contiguous sequence of edge assignment swaps until we reach a vertex with lower capacity as visible in the figure.

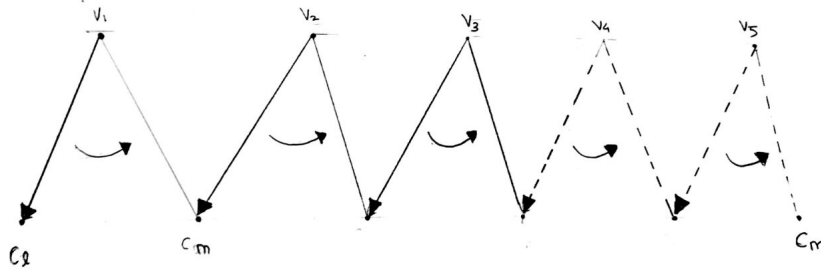


Figure 10: reassignment through propagation

Algorithm:

1. Start with k' centers of G^2 with no regard to capacity L using the 2-approximation algorithm for uncapacitated k -center.
2. $C \leftarrow$ set of all k -centers
3. $C^l \leftarrow$ subset of C such that $\forall C \in C^l, C$ covers exactly l clients.
4. In each phase:
 $l \leftarrow$ the maximum degree of any center
 $C_l \leftarrow$ set of all centers with degree l
5. Subphase (while there are center with degree l do:)

- (a) Choose a client v assigned to some center in C_l .
- (b) Choose another client C_m (mjl) of minimum degree among all the centers v is connected to or some C_m which is connected through some alternate manner of center then non center and finally with v .

This is done in order to reduce the maximum number of assignments to a particular center so as to reach the capacity L . This propagation is possible only when a non center is connected to two or more centers and hence reassignment can be done.

But we still might have too few centers in dense cliques while just sufficient in other parts or maybe even having residual capacities in sparse areas. We also have a counter example for the same as shown in the figure below. The centers are represented by the rectangular marks on the vertices. The left most center is covering many clients within a distance of $j=2$ where as the same cannot be said for the other centers.

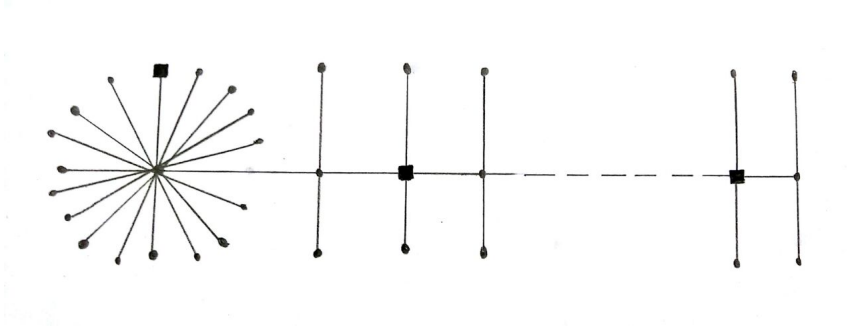


Figure 11: Counter Example

Maybe moving to higher powers of G may contribute to requiring less centers at regions where at present the centers are suffice.

Hence, from G^2 we move to G^4 . Here, each center with its clients forming a star graph will transform into a clique. The same shown in the figure below.

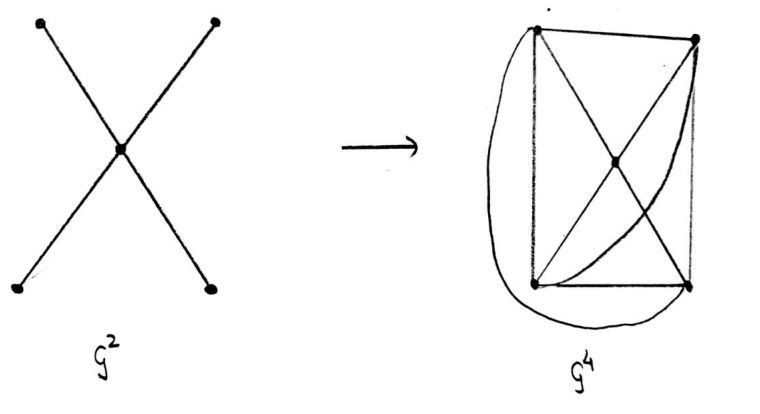


Figure 12: Clique formation

With such a change we might be able to merge few cliques by placing a center on a common node of two cliques (while removing their respective centers) such that both the cliques add up to give vertices less than L .

Following this we have free centers available which we can place in dense cliques with high degree of vertices. Like in the figure below in the nodes that were centers in G^2 are represented by square labels in the left figure and diamond label on the right while the one which we replace them by in G^4 is represent by the square in the right figure.

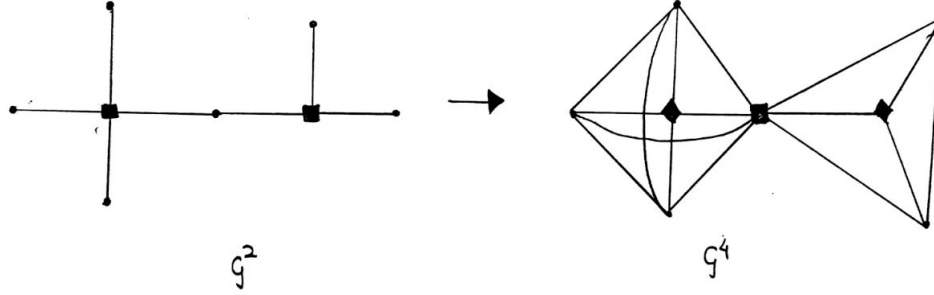


Figure 13: Merging

Like here we obtain one extra center. The problem still remains on how to keep obtaining such centers other than by merging and then placing them at an appropriate position.

Further, we try to come up with some kind of data structure that might help us make some claims on the lower bound on K . As above we make a level structure with the cliques in set C_l as taken before at level l . After a reassignment is performed, the structure changes, because, always the cliques are taken to be a center and its neighbors. After a reassignment, the set of neighbors of (more than or equal to 2) centers change, and also, their capacities change, thus moving them to different levels than present.

Theorem: When there can be no more edge reassignments, then no center at level l , has an edge of length $j = 4$ to any node in level $l' = l + 2$ for all l .

Proof: After there can be no more reassignments, let there exists such a center C at a level l which has an edge to some node P belonging to some level $l' = l + 2$. Let C' be the center covering the node P . Then we can do a reassignment step such that P gets assigned to C instead of C' thereby moving C' one level down and C one level up. Thus the max degree can decrease this way removing one clique at a time.

This contradicts our assumption that no more reassignment is possible.

8.3 Counter Examples

8.4 Current Scenario