This is Matt McCutchen’s simplified proof of correctness of the greedy algorithm for $k$-center clustering with outliers. The algorithm achieves a $4$-approximation in general and a $3$-approximation when cluster centers are restricted to input points or when we can enumerate all “useful” center points in the metric space. The $3$-approximation algorithm is described in Section 3 of http://www.cs.umd.edu/~samir/grant/outlier.pdf. To obtain the $4$-approximation variant, increase the radius of the disks $G_i$ to $2r$ and that of the $E_i$ to $4r$.

**Theorem 1.** When the algorithm is invoked with a particular value of $r$, it produces a set of $k$ clusters of radius $4r$ (or $3r$) that covers at least as many input points as the optimal set of $k$ clusters of radius $r$.

**Proof.** Let $E$ be the set of points covered by the algorithm but not the optimal solution, and let $O$ be the set of points covered by the optimal solution but not the algorithm. We need to show $|E| \geq |O|$.

For each $i$, define a “greedy set” $S_i = G_i - \bigcup_{j=1}^{i-1} E_j$; the disk $G_i$ is greedily chosen to maximize $|S_i|$. The sets $S_i$ are disjoint. Furthermore, an optimal cluster that intersects a greedy set $S_i$ is completely covered by $E_i$, so no future $S_i'$ can contain points of $O_j$; consequently, each optimal cluster intersects at most one greedy set.

Without loss of generality, suppose optimal clusters $O_1, \ldots, O_s$ intersect a greedy set while $O_{s+1}, \ldots, O_k$ do not ($0 \leq s \leq k$). The algorithm’s solution completely covers $O_1$ through $O_s$, but $O_{s+1}$ through $O_k$ may contain uncovered points. If $s = k$, then we are done. Otherwise, for $j = s+1, \ldots, k$, let $U_j = O_j - \bigcup_{i=1}^k E_i$ be the set of uncovered points in $O_j$. We have $O = \bigcup_{j=s+1}^k U_j$. Choose $t \in \{s+1, \ldots, k\}$ so that $|U_t|$ is largest; then $|O| \leq (k-s)|U_t|$. 

Observe that, at any stage, the greedy algorithm could have chosen $O_t$, and that greedy set would contain at least the points of $U_t$. But the algorithm never chose $O_t$, so it must have done at least as well at every stage, so $|S_i| \geq |U_t|$ for every $i$. Now, $s$ of the optimal clusters intersect greedy sets, but we showed previously that each is intersected by at most one greedy set. Thus, at most $s$ greedy sets intersect an optimal cluster, leaving at least $k-s$ sets that do not intersect an optimal cluster and thus contain points uncovered by the optimal solution. These sets are disjoint, and each contains at least $|U_t|$ points. Thus, $|E| \geq (k-s)|U_t| \geq |O|$, as desired. 

With this theorem in mind, we just do a binary search on $r$ to find two close-together values $r^-$ and $r^+$ such that the algorithm covers the required number of input points with $r = r^+$ but not with $r = r^-$. The first property gives us a feasible solution of radius $4r^+$ and the second implies $OPT > r^-$, so we essentially have a $4$-approximation (or similarly a $3$-approximation).