Some Lemmas about Dynamic BFS

November 7, 2013

Dutton and Brigham proved the following theorem (stated as a corollary of their main theorem in [1, page 320].

Theorem 1. *Let G be a graph of girth $g \geq 5$ and $d := \lfloor m/n \rfloor \geq 2$. Then $g \leq 2 + 2\log_d(n/4)$.*

Corollary 1. *Let G be a graph of girth $g > 2\lg n$. Then $m \leq 2n$.*

For graphs of girth substantially greater than $\lg n$, Theorem 1 does not apply because $\lg n < g \leq 2 + 2\log_d(n/4)$ implies $d < 2$. The next lemma uses Corollary 1 to bound the number of edges in graphs of girth $g > \lg n$.

Lemma 1. *A graph of girth $g \geq \lg n$ has at most $n\left(1 + \frac{4\lg n}{g}\right)$ edges.*

Proof. We can assume the given graph G is connected because otherwise we can apply the bound independently to each of its connected components. Let T be a spanning tree of G, and let E be the set of edges of G that are not in T. We need to prove that $|E| \leq \frac{n\lg n}{g}$. By applying the Euler tour technique to T, we can divide G into $\frac{2n\lg n}{g}$ clusters such that two vertices in the same cluster have distance at most $\frac{g}{\lg n}$ from each other in G. Now observe that every edge in E connects two vertices in different clusters and that there are no two edges in E whose endpoints belong to the same two clusters. This is true because otherwise G would contain a cycle of length at most $\frac{2g}{\lg n}$, that is, G's girth would be less than g. Now construct a graph G' by replacing each cluster of G with a single vertex and adding an edge xy to G' if G contains an edge with endpoints in the two clusters corresponding to x and y. Since every edge in E has its endpoints in different clusters and there are no two edges in E whose endpoints belong to the same cluster, G' has at least $|E|$ edges. Moreover, G' has girth at least $\lg n$. To see this observe that, if G' had a cycle C of length less than $\lg n$, we would obtain a corresponding cycle of length less than g in G by replacing every vertex in C with a path in its corresponding cluster. Thus, by Corollary 1, G' has at most $\frac{4n\lg n}{g}$ edges because it has $\frac{2n\lg n}{g}$ vertices. This implise that $|E| \leq \frac{4n\lg n}{g}$, that is, G has at most $n - 1 + \frac{4n\lg n}{g} < n\left(1 + \frac{4\lg n}{g}\right)$ edges. $\square$

In the context of the BFS algorithm, E is exactly the set of edges we can add to a connected graph before dropping the girth below g. Thus, we can perform at most $\frac{4n\lg n}{g}$ updates that keep the girth above g.

By applying this result independently to each cluster diameter α, we obtain that, for each α, we can perform at most $\frac{4n\lg n}{\alpha}$ updates "at level α". Since the cost of an update at level α is $\frac{\alpha n}{B}$, this gives that the cost per level is at most $\frac{4n^2\lg n}{B}$; the total cost for all $\lg B$ levels is thus $\frac{4n^2\lg n}{B/\lg B}$. Our next lemma shows that this bound on the number of updates at all levels in the cluster hierarchy is tight.[1]

Lemma 2. *If the initial graph is a path, then there exists a sequence of insertions so that, for each α, the sequence includes $\Omega\left(\frac{n\lg n}{\alpha}\right)$ updates that keep the graph's girth above α.*

Proof. Let us number the vertices from 0 to $n-1$ in the order they appear along the path. Now assume the cluster parameters $\alpha_0, \alpha_1, \ldots, \alpha_k$ satisfy $\alpha_i = bc^i$, for some integer constants b, c, and k. Also assume n is a multiple of α_k, b is a multiple of $\lg n$, $\alpha_k \le \sqrt{n}$, and $n > 3^8$. We divide the path into a hierarchy of clusters. For $0 \le i \le k$, the ith level of the partition divides the path into $\frac{n}{\beta_i}$ subpath of length $\beta_i := \frac{2\alpha_i}{\log_3 n}$. We call such a subpath a *β_i-cluster*. The *center* of such a cluster is its left endpoint.

We use induction on i to show that we can add $\frac{n}{4\beta_i} = \frac{n\log_3 n}{8\alpha_i}$ edges so that (i) the graph's girth remains at least α_i, (ii) the endpoints of every edge we add are centers of β_i-clusters, and (iii) every cluster center is the endpoint of at most one edge. The base case is $i = k$, in which case our input is the initial path and no cluster center has any added edges incident to it yet. We employ the following greedy process: As long as we have not added $\frac{n}{4\beta_k}$ edges yet and there are two β_k-cluster centers x and y that have distance at least α_k from each other and have degree 2 (i.e., have only incident edges that belong to the initial path), we add the edge xy. Since the initial graph, a path, has girth at least α_k and we never create a cycle of length less than α_k, property (i) holds. Our construction also explicitly ensures properties (ii) and (iii). To prove the inductive claim, it remains to show that this process terminates only after adding $\frac{n}{4\beta_k}$ edges. Assume the contrary. Since there are $\frac{n}{\beta_k}$ β_k-clusters and we add fewer than $\frac{n}{4\beta_k}$ edges, there exists a β_k-cluster center x of degree 2. Let G' be the graph obtained from G by contracting all edges not in the original path, i.e., all edges we added, and let T' be a BFS-tree of G' with root x. Observe that only cluster centers can have more than one child in T', no vertex has more than three children in T', and the distance between any pair of cluster centers is a multiple of β_k. Thus, we have at most $\sum_{i=0}^{\lfloor d/\beta_k \rfloor} 3^i < 2 \cdot 3^{\lfloor d/\beta_k \rfloor}$ cluster centers at distance at most d from x. In particular, there are at most $2 \cdot 3^{\lfloor \alpha_k/\beta_k \rfloor} \le 2\sqrt{n}$ cluster centers at distance at most α_k from x. Since we add less than $\frac{n}{4\beta_k}$ edges, there are at least $\frac{n}{2\beta_k} \ge \frac{\sqrt{n}\log_3 n}{4} > 2\sqrt{n}$ cluster centers of degree 2, so there exists a cluster center y of degree 2 that has distance more than α_k from x and we could have added edge xy, a contradiction.

The argument for $i < k$ is identical after we observe that every α_j-cluster center is also an α_i-cluster center for $j \ge i$. Thus, every edge we added for $j = k, k-1, \ldots, i+1$ connects two α_i-cluster centers. Since $\alpha_j \ge \alpha_i$, for $j \ge i$, the addition of each such edge maintains that the girth of the graph is at least α_i. Thus, we can mimic the construction of the base case for α_i by first adding the edges we added for $j = k, k-1, \ldots, i+1$ and then adding further edges until we have added $\frac{n}{4\beta_k}$ edges in total. □

The above analysis using the graph's girth to determine the level of an update approximates the

[1]We do not believe that the cost analysis is tight, but we show that the number of girth-preserving updates at each level can be as high as in the upper bound. We believe, however, that not all these updates can in fact have the right cost to apply to the level α where we count them.

behaviour of the BFS algorithm in two aspects. First, an update at level α actually has to ensure that the BFS levels of the two endpoints are at least α apart, whereas the analysis based on girth ensures only that the endpoints have distance at least α in G. In general, it is possible that two vertices are far apart from each other in G but belong to BFS levels that are very close to each other, possibly even to the same BFS level. Lemma 3 below shows that, for a single cluster diameter α, we do not lose anything using this approximation in the worst case. Second, even an update that connects vertices on BFS levels that are at least α apart may not be expensive or may be too expensive and may thus be treated as an update for the next higher cluster diameter. We do not know how much we lose by ignoring the actual cost of updates in counting updates for each level, but we believe that by taking these costs into account, we can obtain an improved analysis.

Lemma 3. *If the initial graph G is a path and E is a matching such that $G \cup E$ has girth at least g, then there exists an order in which to insert the edges in E such that, for every inserted edge, the endpoints belong to BFS levels that are at least g apart from each other.*

Proof. Consider G to be directed left to right. Let $\langle e_1, e_2, \ldots, e_m \rangle$ be the sequence of edges we obtain by sorting the edges in E by their left endpoints so that e_i's left endpoint is to the left of e_j's endpoint, for $i > j$. We use x_i to denote e_i's left endpoint, and y_i to denote e_i's right endpoint. Let T_i be a BFS-tree of $G_i := G \cup \{e_1, e_2, \ldots, e_{i-1}\}$ rooted in G's left endpoint r. We claim that $\text{dist}_{T_i}(r, y_i) \geq \text{dist}_{T_i}(r, x_i) + g$. Since edges $e_1, e_2, \ldots, e_{i-1}$ have both their endpoints to the right of x_i and y_i is to the right of x_i, the only path from r to y_i in G_i is through x_i. Thus, y_i is a descendant of x_i in T_i, and there exists a path of length $\text{dist}_{T_i}(r, y_i) - \text{dist}_{T_i}(r, x_i)$ from x_i to y_i in $T_i \subseteq G_i \subseteq G \cup E$ that does not include edge e_i. If $\text{dist}_{T_i}(r, y_i) - \text{dist}_{T_i}(r, x_i) < g$, this implies that the addition of edge e_i creates a cycle of length less than g in $G \cup E$, contradicting that its girth is at least g. Thus, we have $\text{dist}_{T_i}(r, y_i) - \text{dist}_{T_i}(r, x_i) \geq g$. □

What the lemmas in this section show is that, in order to obtain an improved analysis, it is necessary to (a) take the interaction of BFS levels for different cluster diameters α into account, (b) take the cost of updates into account in counting updates that are truly pertinent to a particular α, or both. We tried both approaches and encountered serious difficulties. In fact, our attempts to quantify how the proximity of vertices changes as a result of edge insertions in the potential-based analysis led us exactly to the girth-based analysis.

Both approaches, if they are feasible, hold some promise. Lemma 3 establishes the equivalence of girth and BFS levels only for a single α, so Lemma 2 cannot be used to prove that there are many updates for each α such that the endpoints of the inserted edges are at least α BFS levels apart. In other words, we do not know of any lower bound that would show that an improved analysis based on BFS levels is not possible. Taking the cost of updates into account when counting updates seems to be even more promising, which becomes intuitively obvious after trying to construct long update sequences where each update has exactly the right cost, but it also seems to be even more challenging to achieve.

References

[1] R. D. Dutton and R. C. Brigham. Edges in graphs with large girth. *Graphs an Combinatorics*, 7:315–321, 1991.