Bandwidth Approximation of Many-Caterpillars

Yuval Filmus

September 1, 2009

Abstract

Bandwidth is one of the canonical NP-complete problems. It is NP-hard to approximate within any constant factor even on restricted families of trees (Unger).

Feige gave the first polylogarithmic approximation algorithm. Current best algorithm is by Vempala (approximation ratio $\tilde{O}(\log^3 n)$).

Gupta presented a $O(\log^{2.5} n)$-approximation algorithm on trees. We improve on Gupta's analysis for a restricted family of trees.

Bandwidth: Definition

Bandwidth is a canonical NP-complete problem.

$G = (V, E)$ — an undirected graph.

$f \colon V \leftrightarrow \{1, \ldots, n\}$ — an ordering of V.

Bandwidth of an ordering:

$$\begin{aligned} B(f) &= \max_{(x,\, y) \in E} |f(x) - f(y)| \\ &= \max_{x \neq y} \frac{|f(x) - f(y)|}{d(x,\, y)}. \end{aligned}$$

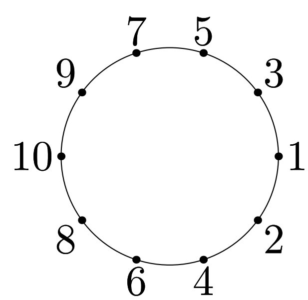
Bandwidth of a graph:

$$B(G) = \min_{f} B(f).$$

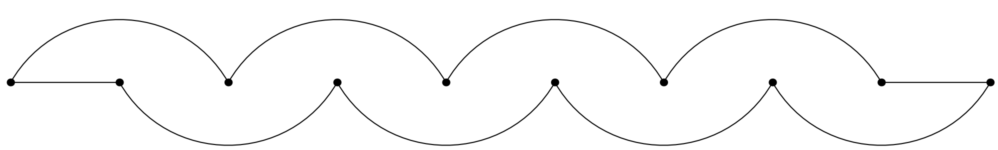
Bandwidth: Examples

Bandwidth of a path is 1:

Bandwidth of a cycle is 2:



Another view:



Bandwidth of the full k-ary tree is $\Theta(n/h)$ (Smithline).

Bandwidth: Complexity Results

Can find whether $B \leq k$ in time $O(n^k)$ (Saxe).

NP-complete on restricted classes of caterpillars (Monien).

NP-hard to approximate within any constant factor on restricted classes of caterpillars (Unger).

First polylogarithmic approximation algorithm — $\tilde{O}(\log^{3.5} n)$ (Feige).

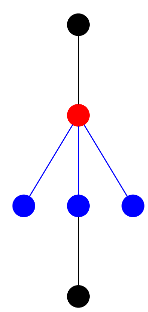
Current best approximation algorithms:

- On general graphs — $\tilde{O}(\log^3 n)$ (Vempala).
- On trees — $O(\log^{2.5} n)$ (Gupta).
- On caterpillars — $O(\log n)$ (HMM, Gupta).

Density

$T = (V, E, r)$ — a rooted tree.

$N(x, \delta)$ — neighborhood of length δ below x.



$N(x, 1)$

Note $x \notin N(x, \delta)$.

Density of G is minimum D such that

$$|N(x, \delta)| \leq \delta D$$

for all $x \in V$ and $\delta \geq 0$.

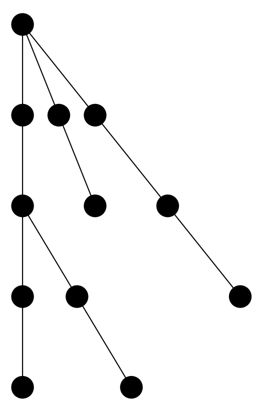
Lemma. $B \geq D/2$.

Caterpillars & Many-caterpillars

Caterpillars — very simple trees.

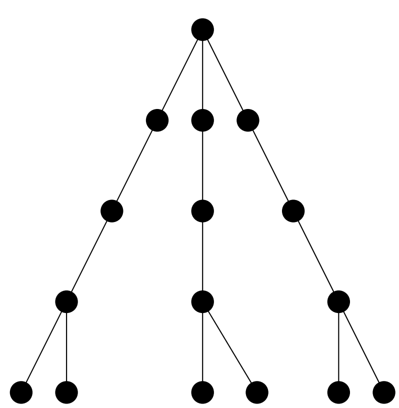
Composed of spine and hairs.

Spine and hairs are paths, hairs emanate from spine.



Many-caterpillars — combinations of caterpillars.

Formed by identifying roots of several caterpillars.



Caterpillar decomposition

Caterpillar decomposition of a tree — a decomposition into edge-disjoint paths.

Decomposition is d-dimensional if every root-to-leaf route is composed of $\leq d$ paths.

Caterpillar dimension — minimum dimension of caterpillar decomposition (denoted κ).

Examples:

- Many-caterpillars are 2-dimensional trees.
- Binary trees have dimension $\log n$.

Dimension is at most $\log n$ (Matoušek).

Gupta's algorithm

$T = (V, E, r)$ — a rooted tree.

1. Choose an optimal caterpillar decomposition.
2. Stretch each path by a factor between 1 and 2.
3. Order the vertices wrt (stretched) distance from r.
4. Output $f(v) = i$ if v is the ith vertex in the ordering.

Gupta showed:

Theorem. $B(f) = O(D \log^2 n \sqrt{\kappa}) = O(D \log^{2.5} n)$.

Theorem. *T a caterpillar $\Rightarrow B(f) = O(D \log n)$.*

Our research

We conjecture that $B(f) = O(D \log n)$.

Proof idea: by induction on caterpillar dimension.

Base: easy (paths).

Step: through construction of $(\kappa + 1)$-trees from κ-trees.

Could only prove step for a certain case.

Also proved base for caterpillars.

Result: conjecture is verified on many-caterpillars.

If proof of step is completed, conjecture will be verified when caterpillar dimension is bounded.

Further refinement could prove conjecture for arbitrary trees.

Approximation algorithm for many-caterpillars

$T = (V, E, r)$ — a many-caterpillar.

1. T is composed of caterpillars $T_1, \ldots, T_d$.

   $h_i =$ height of T_i.

2. For each T_i choose $\sigma_i \in_R \{1, \ldots, h_i\}$.

   Let $s_i = 1 + \sigma_i / h_i$.

   For $v \in T_i$ define $p(v) = s_i d(v, r)$.

3. Order the vertices wrt $p(v)$.

4. Output $f(v) = i$ if v is the ith vertex in the ordering.

Algorithm can be derandomized using method of conditional expectations.

We show:

Theorem. $B(f) = O(D \log h)$.

Algorithm: analysis overview

The analysis of the algorithm is divided into 3 steps:

- Step 1: show that a certain inequality holds for caterpillars.
- Step 2: show that the same inequality holds for many-caterpillars.
- Step 3: deduce that $B(f) = O(D \log h)$.

Algorithm: general analysis

$T = (V, E, r)$ — a many-caterpillar.

$[z, z+1)$ — unit interval (z integral).

X_z = number of $v \in V$ such that

$$p(v) \in [z, z+1)\,.$$

$n^{[z]}$ = number of vertices $v \in V$ such that

$$0 < d(v, r) < z.$$

Theorem. *If for all integers* $1 \leq z \leq 2h$ *and for* $k = \log h$*,*

$$X_z^k \leq \frac{n^{[z]}}{z} c^k D^{k-1} k! \qquad (1)$$

then the output f *of algorithm satisfies*

$$B(f) = O(D \log h).$$

Algorithm: general analysis (2)

Theorem. *If for all integers $1 \le z \le 2h$ and for $k = \log h$,*

$$X_z^k \le \frac{n^{[z]}}{z} c^k D^{k-1} k!$$

then output f of algorithm satisfies

$$B(f) = O(D \log h).$$

Proof. $n^{[z]} \le zD$ by definition of D.

$\Rightarrow X_z^k \le (cDk)^k$.

$\Rightarrow X_z = O(Dk) = O(D \log h)$ for all z.

Choose an edge (x, y).

All vertices v with $f(x) \le f(v) \le f(y)$ satisfy

$$p(v) \in [p(x), p(y)].$$

Since $p(y) \le p(x) + 2$, $[p(x), p(y)]$ spans ≤ 3 unit intervals. $\square$

Derandomization

Need to ensure that for all $1 \le z \le 2h$,

$$X_z^k \le (cDk)^k \,.$$

We will show the inequality holds in expectation, hence

$$E[X_1^k + \cdots + X_{2h}^k] \le 2h(cDk)^k.$$

Using method of conditional expectations, can find an assignment of σ_i so that

$$\begin{aligned} X_1^k + \cdots + X_{2h}^k &\le 2h(cDk)^k \\ \Rightarrow X_z^k &\le 2h(cDk)^k \\ &\le (4cDk)^k, \end{aligned}$$

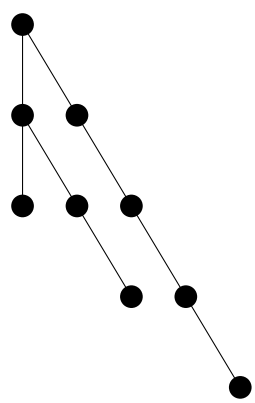
if $k = \log h$.

Regularization of caterpillars: Inventory

$T = (V, E, r)$ — a caterpillar.

Definition. *The inventory* $I(T)$ *is a multiset containing no. of vertices at each depth.*

Inventory of a caterpillar is a multiset:



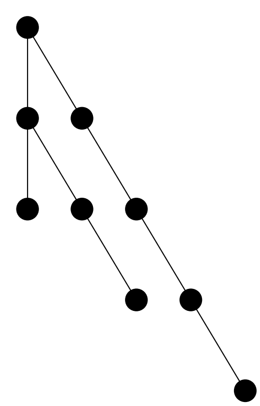
$$I(T) = \{1, 2, 2, 3\}$$

Root is not counted.

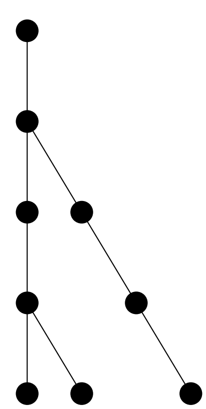
Regularization of caterpillars: Regularity

Definition. *A caterpillar of height h is regular if the spine and all hairs extend to depth h.*

Irregular caterpillar:



Regular caterpillar:



Regularization of caterpillars: Overview

Theorem. *For each caterpillar T there is a (unique) regular caterpillar T^* satisfying:*

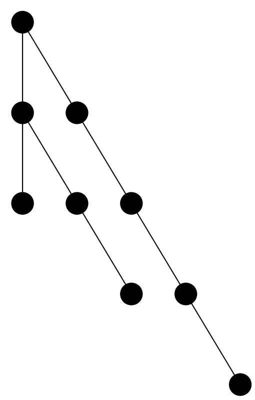
- $I(T^*) = I(T)$.
- $D(T^*) \leq D(T)$.

Overview of process:

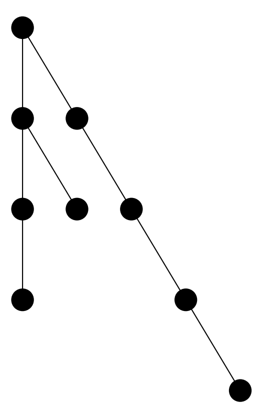
- Spine lengthening is performed until spine extends to maximal depth.
- Normalization and shifting are performed repeatedly until caterpillar is regular.
- All operations preserve the inventory.
- All operations do not increase the density.

Regularization: spine lengthening

Before:



After:

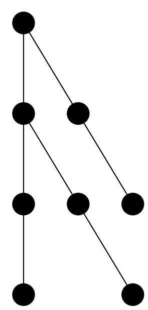


A further step of lengthening is required.

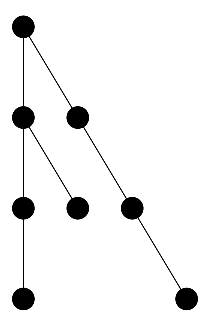
Regularization: normalization

Normalization is performed on pairs of conflicting hairs until no such pairs remain.

Before:



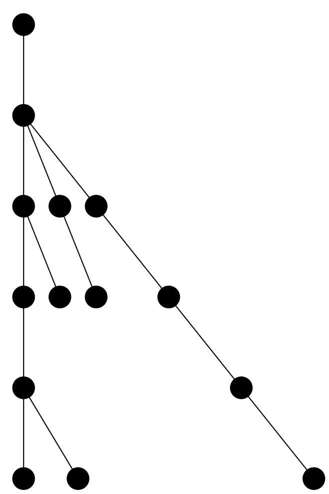
After:



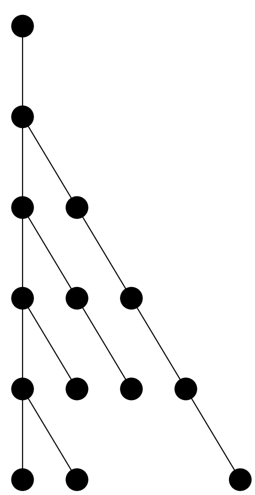
Regularization: shifting

Shifting presupposes normalization.

Before:

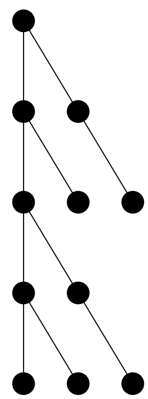


After:

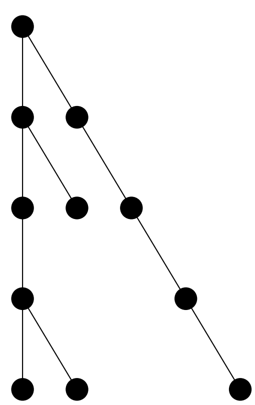


Regularization: example (1)

A non-regular caterpillar (0):

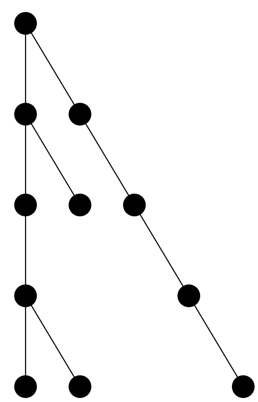


After normalization (1):

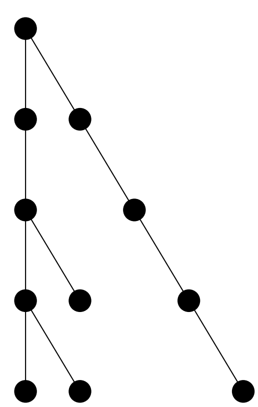


Regularization: example (2)

After normalization (1):

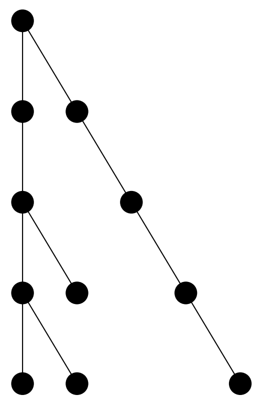


After shifting (2):

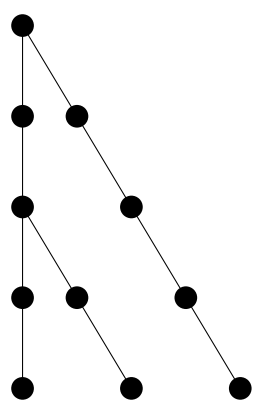


Regularization: example (3)

After shifting (2):

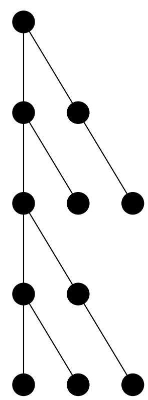


After normalization (3):

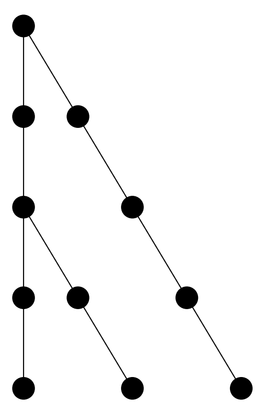


Regularization: example (summary)

A non-regular caterpillar:



Corresponding regular caterpillar:



Both have density 2.5 and inventory

$$\{2,\ 2,\ 3,\ 3\}.$$

Consequences

The inventory of a regular caterpillar is easy to analyze:

Lemma. *The inventory contains at most*

$$\frac{n}{D}2^{-s}$$

elements which are $\geq 2Ds$*.*

Proof. Let $\mathrm{len}(k)$ = length of kth longest hair.

Easy to prove $\mathrm{len}(k) \geq 2\,\mathrm{len}(k+2D)$.

$\Rightarrow \mathrm{len}(2D) \geq 2^{s-1}\,\mathrm{len}(2Ds)$.

$\Rightarrow n \geq 2^{s} D\,\mathrm{len}(2Ds)$. $\square$

Corollary. *For some constant* c *and for all* $k \geq 1$*,*

$$\sum_{x \in I(T)} x^k \leq n c^k D^{k-1} k!\,.$$

Analysis for caterpillars

$T = (V, E, r)$ — a caterpillar.

Theorem.

$$E[X_z^k(T)] \le \frac{8}{z} \sum_{x \in I(T)} x^k.$$

Proof. Recall X_z is no. of vertices v s.t. $\lfloor p(v) \rfloor = z$.

Recall $p(v) = sd(v, r)$ where $s = 1 + \sigma/h$ and

$$\sigma \in_R \{1, \ldots, h\}.$$

$\lfloor p(v) \rfloor = z \Leftrightarrow d(v, r) = \lceil z/s \rceil$.

$\lceil z/s \rceil$ takes each of the values z to $\lceil z/2 \rceil$ w.p. $O(1/z)$. $\square$

Analysis for caterpillars (2)

$T = (V, E, r)$ — a caterpillar.

Recall $n^{[z]}$ = number of vertices $v \in V$ such that

$$0 < d(v, r) < z.$$

Corollary. *For some constant c and for all $k \geq 1$,*

$$E[X_z^k] \leq \frac{n^{[z]}}{z} c^k D^{k-1} k!\,.$$

Proof.

Only vertices of depth $< z$ can be placed inside $[z, z+1)$.

□

Analysis for many-caterpillars

$T = (V, E, r)$ — a many-caterpillar.

Theorem. *For some constant c and for all $k \geq 1$,*

$$E[X_z^k(T)] \leq \frac{n^{[z]}(T)}{z} c^k D^{k-1} k!\,.$$

Proof. Let T be composed of caterpillars $T_1, \ldots, T_d$.

$X_z(T) = X_z(T_1) + \cdots + X_z(T_d)$.

Recall the caterpillar result:

$$E[X_z^p(T_i)] \leq \frac{n^{[z]}(T_i)}{z} c^p D^{p-1} p!, \quad (p \geq 1)$$

where $n^{[z]}(T_1) + \cdots + n^{[z]}(T_d) = n^{[z]}(T)$.

Using the multinomial theorem yields the result. $\square$

Analysis for many-caterpillars (1)

Recall the caterpillar result:

$$E[X_z^p(T_i)] \le \frac{n^{[z]}(T_i)}{z} c^p D^{p-1} p!\,. \quad (p \ge 1)$$

Multinomial theorem gives a sum of terms

$$k! \prod_{s \text{ factors}} E[X_z^{p_i}(T_i)]/p_i!\,,$$

where $\sum p_i = k$.

Worst case is when $n^{[z]}(T_i) = n^{[z]}(T)/d$.

Each term is bounded by

$$k! \left(\frac{n^{[z]}(T)/d}{z}\right)^s c^k D^{k-s}.$$

Analysis for many-caterpillars (2)

A term involving s factors is bounded by

$$k!\left(\frac{n^{[z]}(T)/d}{z}\right)^s c^k D^{k-s}.$$

There are $\leq 2^k(ed/s)^s$ terms involving s factors.

Their sum is at most

$$(e/s)^s k!\left(\frac{n^{[z]}(T)}{z}\right)^s (2c)^k D^{k-s}.$$

Using $n^{[z]}(T)/z \leq D$, this sum is at most

$$(e/s)^s k!\frac{n^{[z]}(T)}{z}(2c)^k D^{k-1}.$$

The result follows since

$$\sum(e/s)^s < \infty.$$