The Maximum Number of Squares in a Tree

Maxime Crochemore, Costas Iliopoulos,
Tomasz Kociumaka, Marcin Kubica,
Jakub Radoszewski, Wojciech Rytter,
Tomasz Waleń, Wojciech Tyczyński

King's College London, University of Warsaw

CPM 2012 Helsinki, July 3, 2012

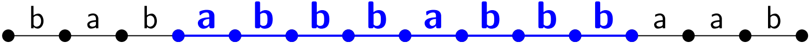
Square

a b b b a b b b

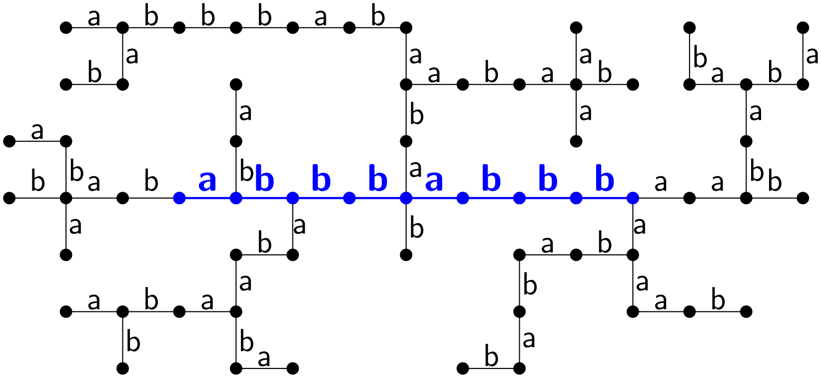
Square in a word

b a b **a b b b a b b b** a a b

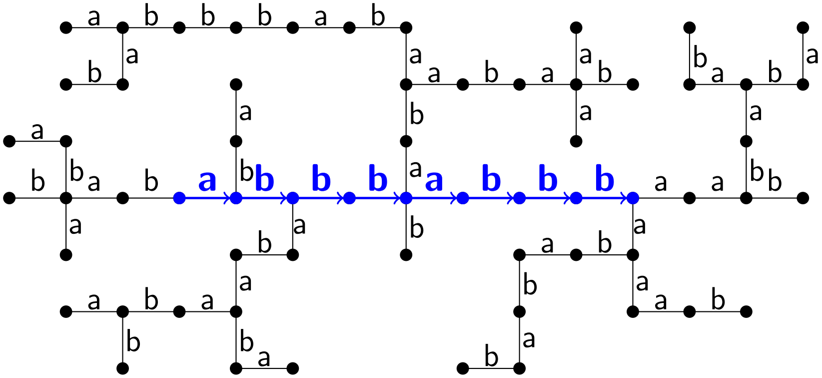
Square in a tree



Square in a tree



Square in a tree

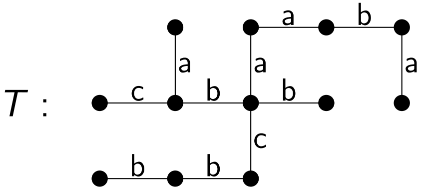


Number of squares in a tree

We consider unrooted, unoriented trees with edges labeled by single letters. Subword of such a tree is a value of a simple path.

Number of squares in a tree

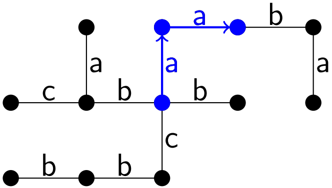
We consider unrooted, unoriented trees with edges labeled by single letters. Subword of such a tree is a value of a simple path.



Number of squares in a tree

We consider unrooted, unoriented trees with edges labeled by single letters. Subword of such a tree is a value of a simple path.

T :

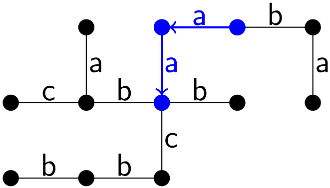


Squares in T: aa

Number of squares in a tree

We consider unrooted, unoriented trees with edges labeled by single letters. Subword of such a tree is a value of a simple path.

T :

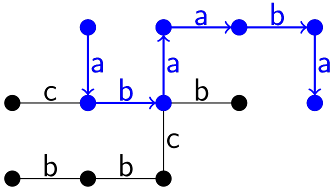


Squares in T: aa

Number of squares in a tree

We consider unrooted, unoriented trees with edges labeled by single letters. Subword of such a tree is a value of a simple path.

T :

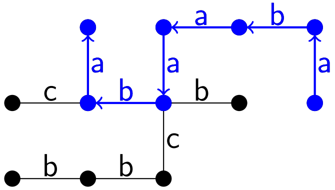


Squares in T: aa, abaaba

Number of squares in a tree

We consider unrooted, unoriented trees with edges labeled by single letters. Subword of such a tree is a value of a simple path.

T :

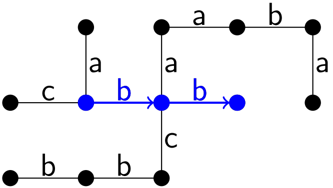


Squares in T: aa, abaaba

Number of squares in a tree

We consider unrooted, unoriented trees with edges labeled by single letters. Subword of such a tree is a value of a simple path.

T :

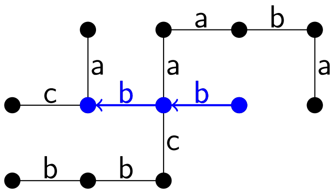


Squares in T: aa, abaaba, bb

Number of squares in a tree

We consider unrooted, unoriented trees with edges labeled by single letters. Subword of such a tree is a value of a simple path.

T :

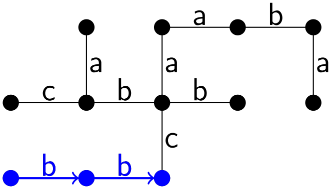


Squares in T: aa, abaaba, bb

Number of squares in a tree

We consider unrooted, unoriented trees with edges labeled by single letters. Subword of such a tree is a value of a simple path.

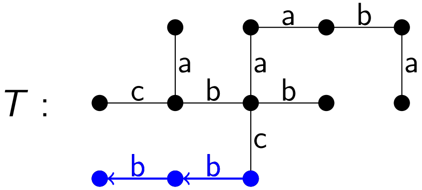
T :



Squares in T: aa, abaaba, bb

Number of squares in a tree

We consider unrooted, unoriented trees with edges labeled by single letters. Subword of such a tree is a value of a simple path.

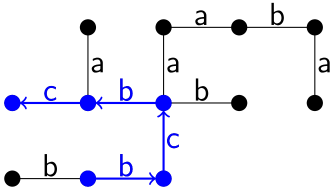


Squares in T: aa, abaaba, bb

Number of squares in a tree

We consider unrooted, unoriented trees with edges labeled by single letters. Subword of such a tree is a value of a simple path.

T :

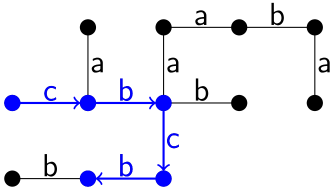


Squares in T: aa, abaaba, bb, bcbc

Number of squares in a tree

We consider unrooted, unoriented trees with edges labeled by single letters. Subword of such a tree is a value of a simple path.

T :

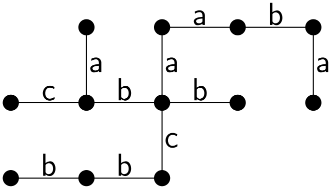


Squares in T: aa, abaaba, bb, bcbc, cbcb

Number of squares in a tree

We consider unrooted, unoriented trees with edges labeled by single letters. Subword of such a tree is a value of a simple path.

T :



Squares in T: aa, abaaba, bb, bcbc, cbcb.
There are 5 distinct squares, i.e. $\mathrm{sq}(T) = 5$.

Maximum number of squares

What is the maximum number of squares a tree of n nodes might contain?

Maximum number of squares

What is the maximum number of squares a tree of n nodes might contain?

Theorem (This paper)

A tree of n nodes contains $O(n^{4/3})$ squares.
This bound is asymptotically tight.

Maximum number of squares

What is the maximum number of squares a tree of n nodes might contain?

Theorem (This paper)

A tree of n nodes contains $O(n^{4/3})$ squares.
This bound is asymptotically tight.

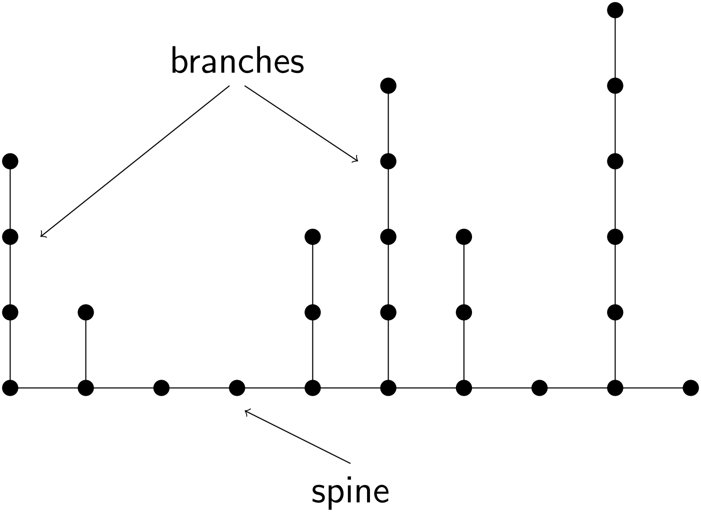
Theorem (Fraenkel & Simpson, 1998)

A word of length n contains at most $2n$ squares.
There is a word of length n with $n - o(n)$ squares.

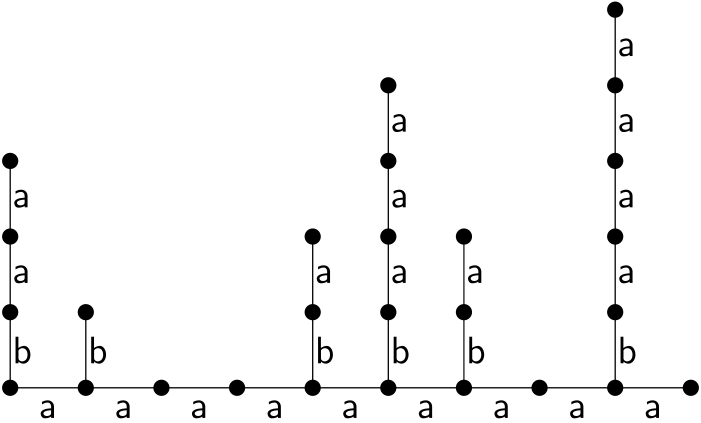
Conjecture

A word of length n contains at most n squares.

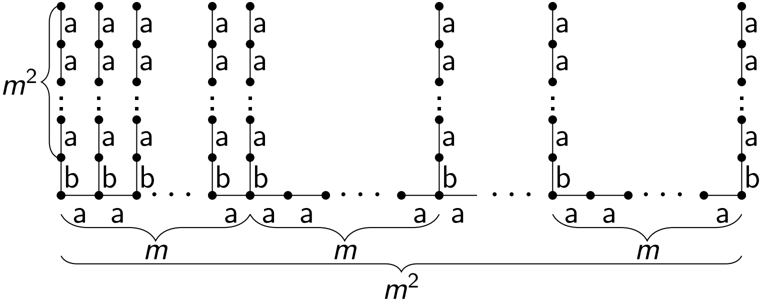
Comb



Standard comb

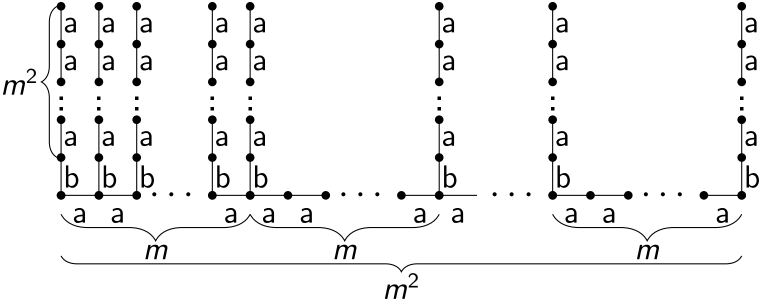


Lower bound



Branches at $\{0, 1, 2, \ldots, m-1, m, 2m, 3m, \ldots, m^2\}$.
$\Theta(m^3)$ nodes,
$\Theta(m^4)$ squares: $\{a^i b a^{i+j} b a^j : 1 \leq i + j \leq m^2\}$.

Lower bound

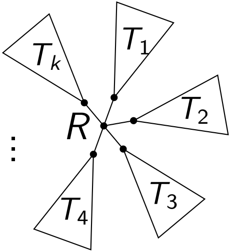


Branches at $\{0, 1, 2, \ldots, m-1, m, 2m, 3m, \ldots, m^2\}$.
$\Theta(m^3)$ nodes,
$\Theta(m^4)$ squares: $\{a^i ba^{i+j} ba^j : 1 \leq i+j \leq m^2\}$.

Theorem

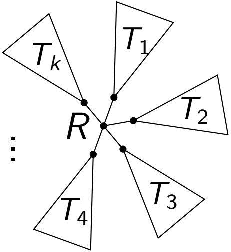
There are trees of n nodes with $\Theta(n^{4/3})$ squares.

Centroid decomposition of T



$$|T_i| \leq \frac{|T|}{2}$$

Centroid decomposition of T

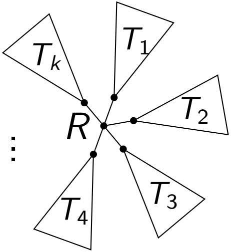


$$|T_i| \leq \frac{|T|}{2}$$

sq(T, R) – number of squares passing through R

$$\mathrm{sq}(T) \leq \mathrm{sq}(T, R) + \Sigma_i \mathrm{sq}(T_i)$$

Centroid decomposition of T



$|T_i| \leq \frac{|T|}{2}$

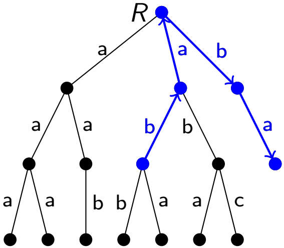
$\mathrm{sq}(T, R)$ – number of squares passing through R

$$\mathrm{sq}(T) \leq \mathrm{sq}(T, R) + \Sigma_i \mathrm{sq}(T_i)$$

Fact

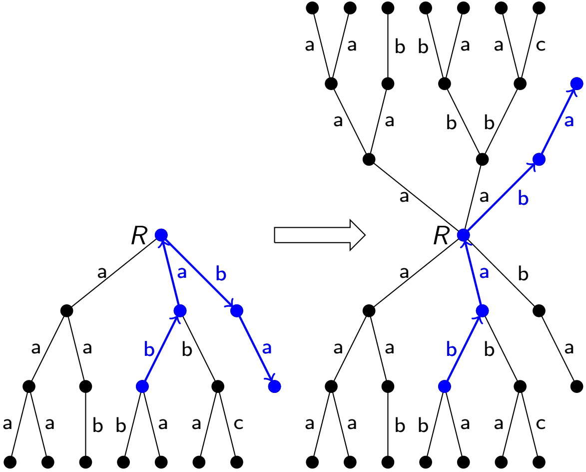
If $sq(T, R) = O(n^{4/3})$ for every tree T of size n, then $sq(T) = O(n^{4/3})$ for every tree T of size n.

D-trees — deterministic double trees



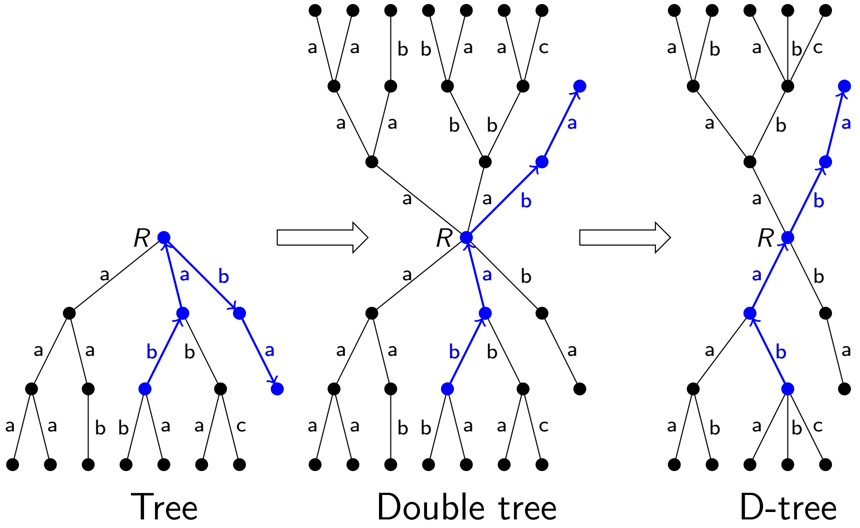
Tree

D-trees — deterministic double trees

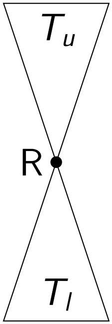


Tree　　　　Double tree

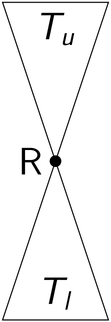
D-trees — deterministic double trees



Squares in D-trees



Squares in D-trees

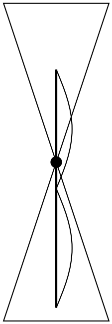


The following lemma implies the main theorem:

Lemma

For any D-tree of size n the number of squares with midpoint in T_l and ending in T_u is $O(n^{4/3})$.

Squares in D-trees

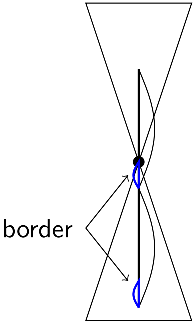


The following lemma implies the main theorem:

Lemma

For any D-tree of size n the number of squares with midpoint in T_l and ending in T_u is $O(n^{4/3})$.

Squares in D-trees



The following lemma implies the main theorem:

Lemma

For any D-tree of size n the number of squares with midpoint in T_l and ending in T_u is $O(n^{4/3})$.

Types of borders

Definition

We say that u is of periodic type (p, q) if $u = (pq)^k p$ for $k \geq 2$, $q \neq \varepsilon$, and pq is primitive.

Types of borders

Definition

We say that u is of periodic type (p, q) if $u = (pq)^k p$ for $k \geq 2$, $q \neq \varepsilon$, and pq is primitive.

Let w be a word of length $\leq n$.

- $O(\log n)$ borders of w have no periodic type.
- Remaining borders are of $O(\log n)$ types.

Types of borders

Definition

We say that u is of periodic type (p, q) if $u = (pq)^k p$ for $k \geq 2$, $q \neq \varepsilon$, and pq is primitive.

Let w be a word of length $\leq n$.

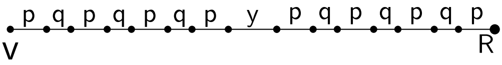
- $O(\log n)$ borders of w have no periodic type.
- Remaining borders are of $O(\log n)$ types.

If w has borders of periodic type (p, q) then it has the following representation:

- $w = (pq)^k p$ (global borders) or
- $w = (pq)^l pyp(qp)^r$ (regular borders).

Squares and borders

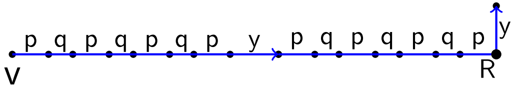
- Global borders – easy, $O(n)$ in total.
- Regular boders of a single type:



Squares and borders

- Global borders – easy, $O(n)$ in total.
- Regular boders of a single type:

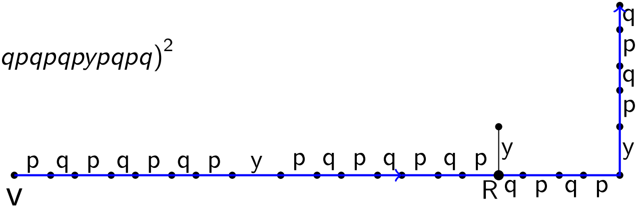
$(pqpqpqpy)^2$



Squares and borders

- Global borders – easy, $O(n)$ in total.
- Regular boders of a single type:

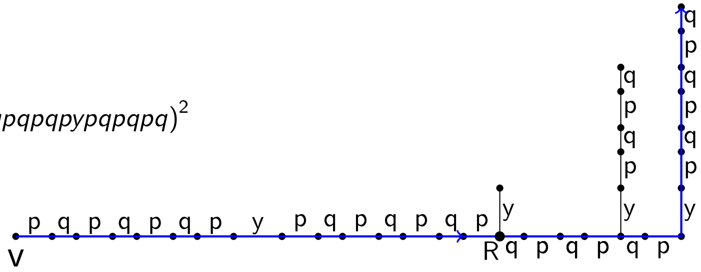
$(pqpqpqpypqpq)^2$



Squares and borders

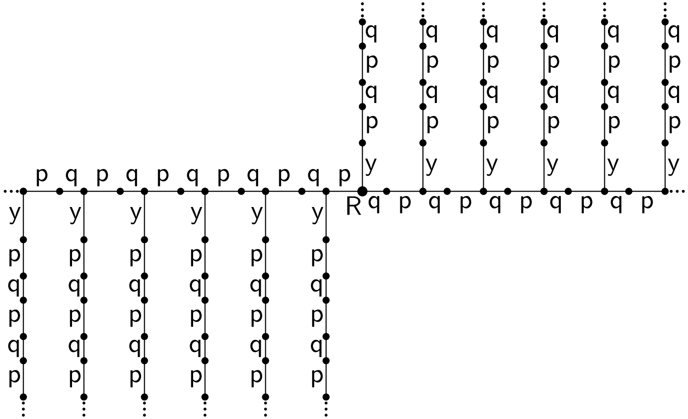
- Global borders – easy, $O(n)$ in total.
- Regular boders of a single type:

$(pqpqpqpypqpqpq)^2$

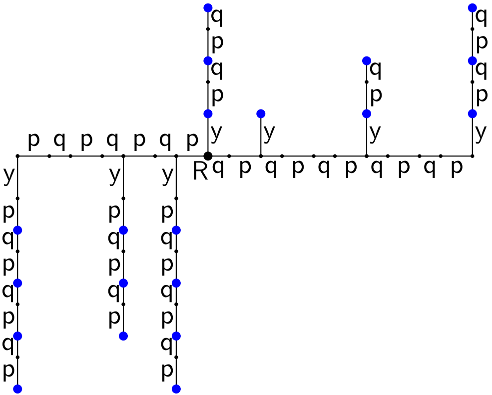


General combs

A (p, q, y)-comb of T is the maximal common subtree of T and the following infinite D-tree:

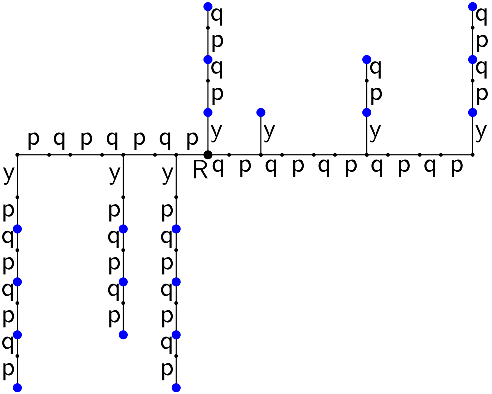


Squares induced by combs



The blue nodes are called *main* nodes of a comb. Squares with both endpoints at these nodes are *induced* by a comb.

Squares induced by combs



The blue nodes are called *main* nodes of a comb. Squares with both endpoints at these nodes are *induced* by a comb.

Size of a comb is the number of main nodes.

Outline of the central proof

1. Just $O(n \log n)$ squares are not induced by combs.
2. Small combs ($\leq n^{0.6}$) induce $o(n^{4/3})$ squares:
   - a comb of size k induces $O(k^{1/2})$ squares starting in a single main node,
   - a single node in T_l can be a main node of $O(\log n)$ combs.
3. Big combs ($> n^{0.6}$) induce $O(n^{4/3})$ squares:
   - combs are almost disjoint in a certain sense: $|Main(\mathcal{C}) \cap Main(\mathcal{C}')| \leq 4$,
   - the total size of big combs is $O(n)$,
   - a comb of size k induces $O(k^{4/3})$ squares.

Thank you

Thank you for your attention!