BIPARTITE GRAPHS OF SMALL READABILITY

Rayan Chikhi, Vladan Jovičić, Stefan Kratsch, Paul Medvedev, Martin Milanič, Sofya Raskhodnikova, **Nithin Varma**

BOSTON UNIVERSITY

COCOON 2018

Appetizer - A Puzzle!

1, 2

1, 3, 2

1, 4, 3, 2

1, 5, 4, 3, 5, 2

1, 6, 5, 4, 3, 5, 2

1, 7, 6, 5, 4, 7, 3, 5, 7, 2

- Rule to form a new line ?
- Number of elements in the i^{th} line ?

Hint: Think prime numbers!

Answer ?

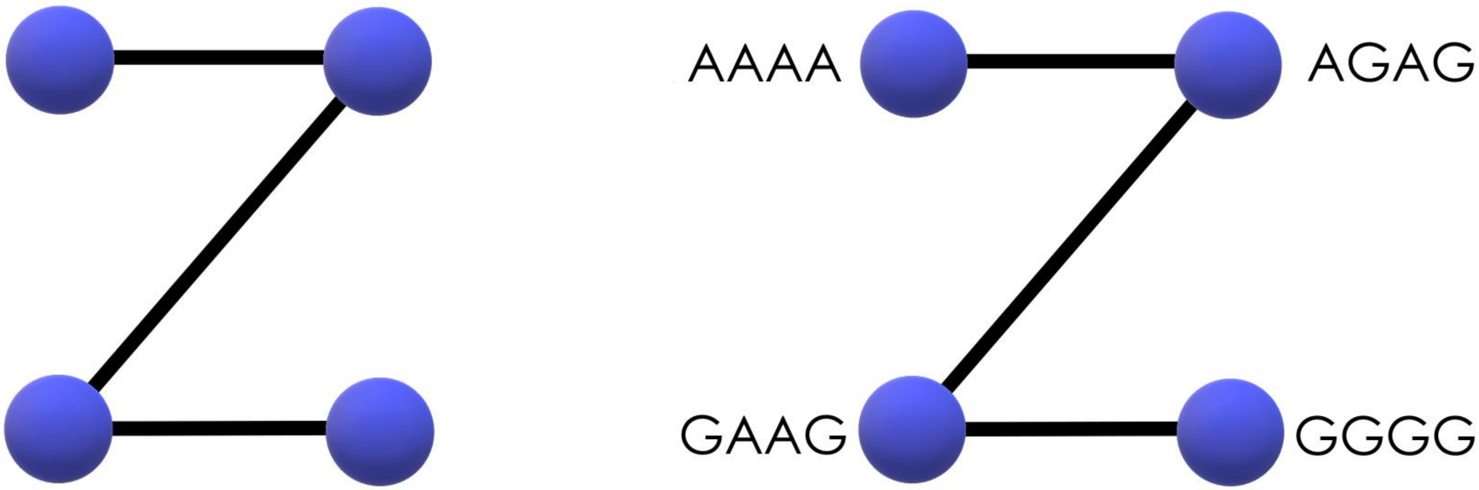
Towards the end of the talk :)

BIPARTITE GRAPHS OF SMALL READABILITY

Bipartite Graphs and Strings

Assign strings to vertices of the bipartite graph

- edges ⇒ overlapping strings
- nonedges ⇒ nonoverlapping strings



Readability

- Parameter of bipartite graphs
  [Chikhi Medvedev Milanič Raskhodnikova '16]
- Captures the assignment of strings to vertices such that string overlaps represent vertex adjacencies
- Arises from the study of overlap graphs in bioinformatics
  - Overlap digraphs (de Bruijn graphs, string graphs) of DNA strings have several applications in the context of genome assembly

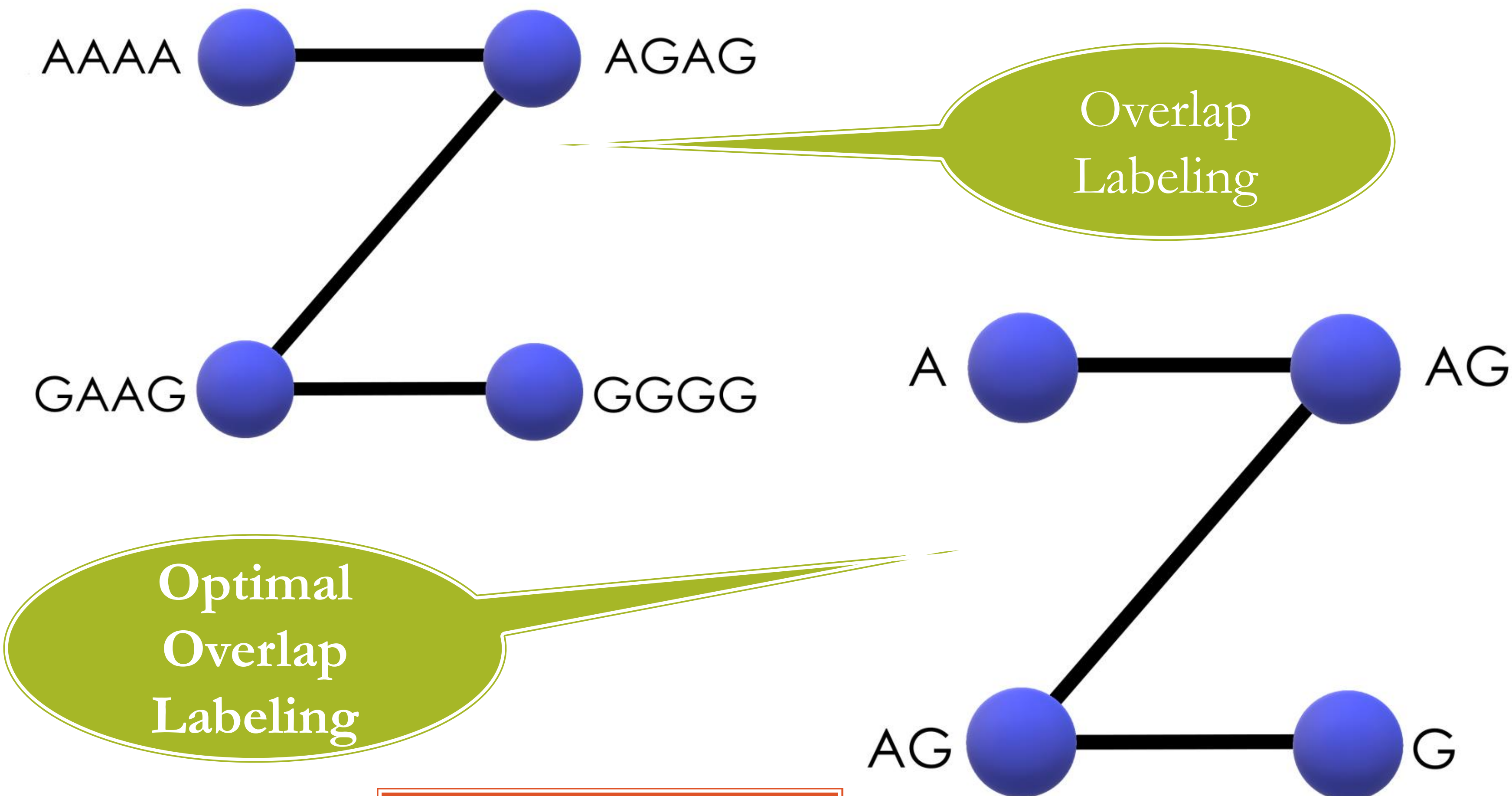
Readability: Definition

No restriction on alphabet

- **Overlapping strings**: Nonempty suffix of first string equals a prefix of second string
- **Overlap labeling:** Assignment of strings to vertices:
  - An **edge** corresponds to **overlapping strings**
  - A **non-edge (across the bipartition)** corresponds to **nonoverlapping** strings
- **Length of an overlap labeling:** Maximum over lengths of all strings in the labeling

Readability : Min. over lengths of all overlap labelings

Overlap labeling : Examples



Readability is 2

Complexity of Computing Readability

Is the following problem in NP?

Given graph G and integer k, does G have readability at most k?

Trivial for k = 1; G has readability 1 iff G is a disjoint union of bicliques (complete bipartite graphs).

Our Results 1: Readability 2 Algorithm

Polynomial-time algorithm to decide whether a graph G has readability at most 2

G has readability at most 2 $\Rightarrow$ Algorithm produces an overlap labeling of length at most 2.

Key idea: We characterize bipartite graphs of readability at most 2 as having a matching with particular properties.

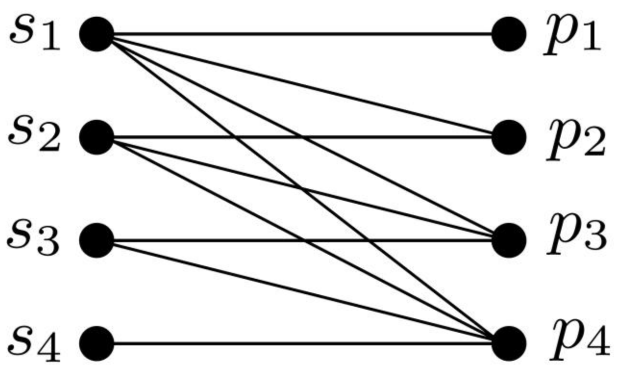
Readability Bounds

- Readability of a graph G is at most $2^{\Delta+1}$, where Δ is the maximum degree of G [Braga Meidanis'02, Chikhi Medvedev Milanič Raskhodnikova '16]
- Readability of almost all bipartite graphs (on n vertices in each part) is $\Omega(n/\log n)$ [CMMR'16]
- Explicit graph family (on n vertices in each part) with readability $\Omega(n)$ [CMMR'16]

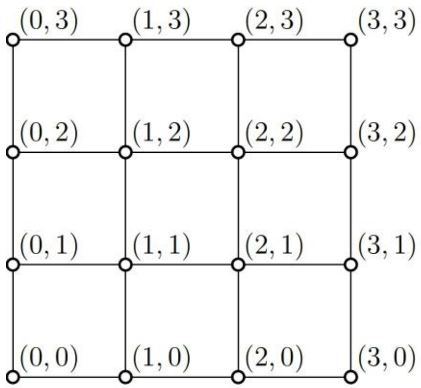
What graph families have small readability?

Our Results 2: Graph Families with Small Readability

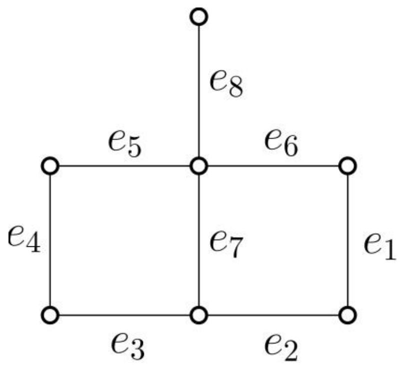
Graph Family	Upper Bound	Lower Bound
Bipartite chain graphs	$O(\sqrt{n})$	$\Omega(\log n)$
Grids and grid graphs	3	3



Bipartite Chain Graph



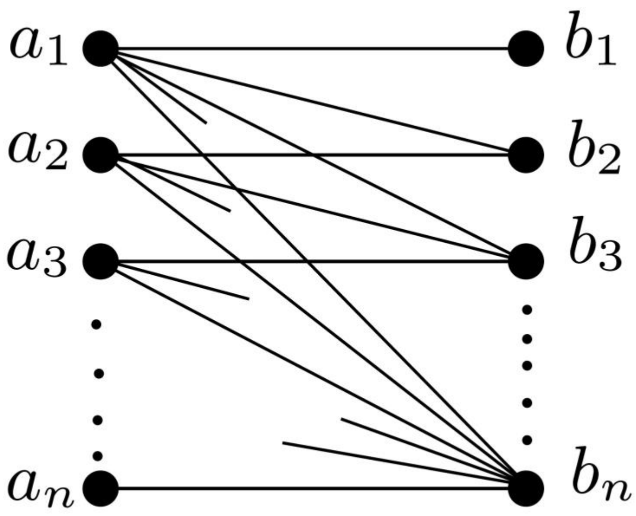
4 × 4 Grid



A Grid Graph

Today: Readability of Bipartite Chain Graphs

Bipartite Chain Graph $C_{n,n}$

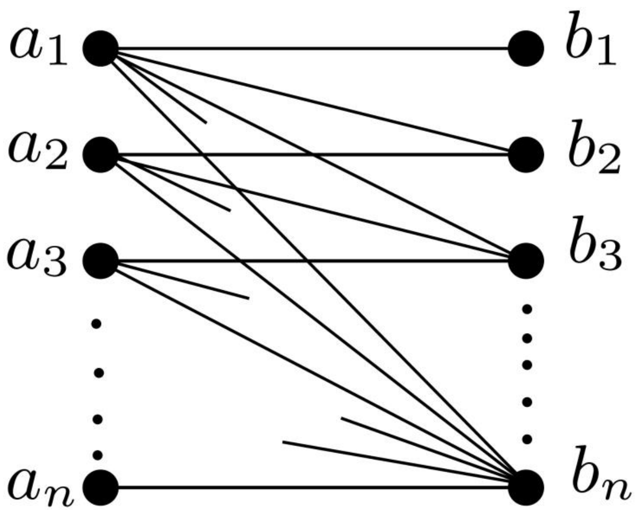


$\forall i \in [n] \quad N(a_i) = \{b_i, b_{i+1}, \ldots, b_n\}$

Readability of $C_{n,n}$ is $O(\sqrt{n})$ using labels of 3 characters.

Forward-Matching Sequences

Bipartite Chain Graph $C_{n,n}$



Sequence of strings $s_1, s_2, \ldots, s_n$ such that:

- **s_i overlaps s_j iff $i < j$**

$\Downarrow$

Overlap Labeling for $C_{n,n}$

For all r, forward-matching sequence of $\Omega(r^2)$ strings each of length at most r

$\Downarrow$

Readability of $C_{n,n}$ is $O(\sqrt{n})$

Constructing Forward-Matching Sequences

20, 0, 01 is a forward-matching sequence

Concatenation preserves forward-matching property.

$$s_1, \ldots s_i \ldots, s_t \Rightarrow s_1, \ldots s_i, s_i \cdot s_{i+1}, s_{i+1} \ldots, s_t$$

Round 2 20, 0, 01

Round 3 20, 200, 0, 001, 01

Round 4 20, 200, 2000, 0, 0001, 001, 01

Round r All strings of length $\leq r$ using concatenation

Lower bound on # strings in round r ?

Number of Strings in the Forward-Matching Sequence (Puzzle Solved!)

Right half of sequences	**Right half of lengths**	
0, 01	1, 2	L_2
0, 001, 01	1, 3, 2	L_3
0, 0001, 001, 01	1, 4, 3, 2	L_4
0, 00001, 0001, 001, 00101, 01	1, 5, 4, 3, 5, 2	L_5

Cool Claim: For $r \geq 2$, $|L_r| = \sum_{k=1}^{r} \varphi(k)/2$.

$\Omega(r^2)$

Euler's totient function

$\varphi(k)$: number of integers relatively prime (co-prime) with k

Proof of Cool Claim

Cool Claim: For $r \geq 2$, $|L_r| = \sum_{k=1}^{r} \varphi(k)/2$.

For $k \leq r$, # of occurrences of k in L_k is $\varphi(k)/2$.

L_{k-1}
$1, \ldots, x, y, \ldots, 2$

L_k
$1, \ldots, x, k, y, \ldots, 2$

# of occurrences of k in L_k
= # neighbors x, y in L_{k-1} such that $x + y = k$

Proof of Cool Claim

Cool Claim: For $r \geq 2$, $|L_r| = \sum_{k=1}^{r} \varphi(k)/2$.

Sub-Claim 1: Any two neighboring elements of L_{k-1} are co-prime (relatively prime). [Proof by Induction]

L_{k-1} → L_k

$1, \ldots, x, y, \ldots, 2$ → $1, \ldots, x, k, y, \ldots, 2$

# of occurrences of k in L_k
= # neighbors x, y in L_{k-1} such that $x + y = k$

Proof of Cool Claim

Cool Claim: For $r \geq 2$, $|L_r| = \sum_{k=1}^{r} \varphi(k)/2$.

Sub-Claim 1: Any two neighboring elements of L_{k-1} are co-prime (relatively prime). [Proof by Induction]

L_{k-1}

$1, \dots, x, y, \dots, 2$

L_k

$1, \dots, x, k, y, \dots, 2$

\# of occurrences of k in L_k

= # neighbors x, y in L_{k-1} such that $x + y = k$

= # neighbors x, y such that x, y are co-prime and $x + y = k$

Proof of Cool Claim

Cool Claim: For $r \geq 2$, $|L_r| = \sum_{k=1}^{r} \varphi(k)/2$.

Sub-Claim 2: Numbers x, y such that $x + y = k$ occur as neighbors exactly once in L_{k-1}. [Proof by strong induction].

L_{k-1}

$1, \dots, x, y, \dots, 2$

L_k

$1, \dots, x, k, y, \dots, 2$

# of occurrences of k in L_k

= # neighbors x, y in L_{k-1} such that $x + y = k$

= # neighbors x, y such that x, y are co-prime and $x + y = k$

Proof of Cool Claim

Cool Claim: For $r \geq 2$, $|L_r| = \sum_{k=1}^{r} \varphi(k)/2$.

Sub-Claim 2: Numbers x, y such that $x + y = k$ occur as neighbors exactly once in L_{k-1}. [Proof by strong induction].

L_{k-1}
$1, \ldots, x, y, \ldots, 2$

⇒

L_k
$1, \ldots, x, k, y, \ldots, 2$

\# of occurrences of k in L_k
= # neighbors x, y in L_{k-1} such that $x + y = k$
= # neighbors x, y such that x, y are co-prime and $x + y = k$
= # co-prime pairs x, y such that $x + y = k$.

Proof of Cool Claim

Cool Claim: For $r \geq 2$, $|L_r| = \sum_{k=1}^{r} \varphi(k)/2$.

Fact: If x and y are co-prime, then x and y are co-prime to $x + y$.

L_{k-1}
$1, \ldots, x, y, \ldots, 2$

L_k
$1, \ldots, x, k, y, \ldots, 2$

\# of occurrences of k in L_k
= # neighbors x, y in L_{k-1} such that $x + y = k$
= # neighbors x, y such that x, y are co-prime and $x + y = k$
= # co-prime pairs x, y such that $x + y = k$

Proof of Cool Claim

Cool Claim: For $r \geq 2$, $|L_r| = \sum_{k=1}^{r} \varphi(k)/2$.

Fact: If x and y are co-prime, then x and y are co-prime to $x + y$.

L_{k-1}

$1, \dots, x, y, \dots, 2$

$\Rightarrow$

L_k

$1, \dots, x, k, y, \dots, 2$

\# of occurrences of k in L_k

= # neighbors x, y in L_{k-1} such that $x + y = k$

= # neighbors x, y such that x, y are co-prime and $x + y = k$

= # co-prime pairs x, y such that $x + y = k$

= # pairs x, y such that $x + y = k$ and x, y co-prime to k

Proof of Cool Claim

Cool Claim: For $r \geq 2$, $|L_r| = \sum_{k=1}^{r} \varphi(k)/2$.

L_{k-1}
$1, \ldots, x, y, \ldots, 2$

⇒

L_k
$1, \ldots, x, k, y, \ldots, 2$

\# of occurrences of k in L_k
= # neighbors x, y in L_{k-1} such that $x + y = k$
= # neighbors x, y such that x, y are co-prime and $x + y = k$
= # co-prime pairs x, y such that $x + y = k$
= # pairs x, y such that $x + y = k$ and x, y co-prime to k
= $\varphi(k)/2$

Proof of Cool Claim

Cool Claim: For $r \geq 2$, $|L_r| = \sum_{k=1}^{r} \varphi(k)/2$.

# of occurrences of k in L_k $= \varphi(k)/2$

$$\Rightarrow |L_r| = \sum_{k=1}^{r} \varphi(k)/2$$

$\Omega(r^2)$

Cool Claim proved. □

Forward-Matching sequence of $\Omega(r^2)$ strings each of length at most r

Proof of Cool Claim

Cool Claim: For $r \geq 2$, $|L_r| = \sum_{k=1}^{r} \varphi(k)/2$.

# of occurrences of r in $L_r = \varphi(r)/2$

$$\Rightarrow \quad |L_r| = \sum_{k=1}^{r} \varphi(k)/2$$

$\Omega(r^2)$

Cool Claim proved. □

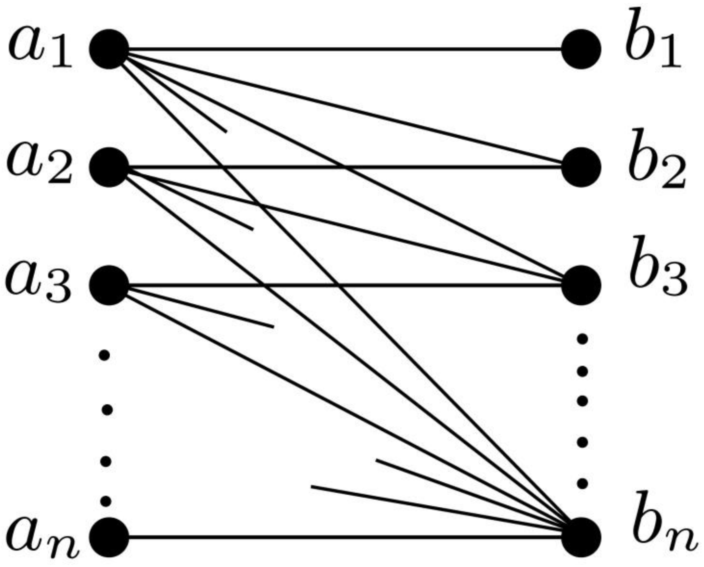
Forward-Matching sequence of $\Omega(r^2)$ strings each of length at most r

$\Downarrow$

Readability of $C_{n,n}$ is $O(\sqrt{n})$.

What we showed

Bipartite Chain Graph $C_{n,n}$



$\forall i \in [n] \quad N(a_i) = \{b_i, b_{i+1}, \ldots, b_n\}$

Readability of $C_{n,n}$ is $O(\sqrt{n})$ using labels of 3 characters.

Open Questions

- Other graph families of low readability ?
- Complexity of checking if the readability is at most k, for larger values of $k > 2$.
- Can readability be used as a parameter ? Are there hard problems that become easy on low readability graphs ?
- Close the gap between the lower bound and upper bound on the readability of bipartite chain graphs.

Thank you!