Implicit Representations of Graphs & Randomized Communication

Viktor Zamaraev (University of Liverpool)

joint work with

Nathan Harms
(EPFL)

Sebastian Wild
(University of Liverpool)

LoGAlg 2023, Warsaw, Poland,
November 15-17, 2023

My motivation

Speed of hereditary graph classes

- Class of graphs is hereditary if it is closed under vertex deletion
- If $\mathscr{X}$ is a class of labeled graphs, then $\mathscr{X}_n$ is the set of graphs from $\mathscr{X}$ with vertex set $[n] := \{1,2,\ldots,n\}$
- The speed of $\mathscr{X}$ is the function that maps $n \mapsto |\mathscr{X}_n|$

Example

Let $\mathcal{P}$ be the class of all graphs.

$$|\mathcal{P}_n| = 2^{\binom{n}{2}} = 2^{n(n-1)/2}$$

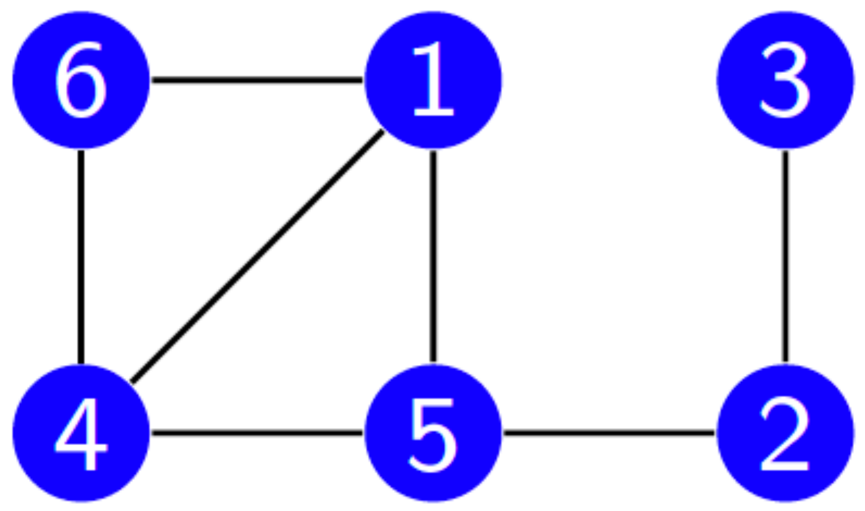
$$\log_2 |\mathcal{P}_n| = \Theta(n^2)$$

Graph coding

Why are we interested in $\log_2 |\mathcal{X}_n|$?

A graph coding is a representation of the graph by a word in a finite alphabet.

Graph G



Adjacency matrix

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Binary word
(*canonical code* of G)

$$\underbrace{0\ 0\ 1\ 1\ 1\ 1\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ 1\ 0}_{n(n-1)/2 \text{ bits}}$$

Graph coding

Why are we interested in $\log_2 |\mathcal{X}_n|$?

If we have no a priori information, then in the worst case we need $\log_2 |\mathcal{P}_n| = n(n-1)/2$ bits to represent an n-vertex graph G.

If $G \in \mathcal{X}_n$ and we know something about $\mathcal{X}$ it may help to represent G with less than $\binom{n}{2}$ bits.

On the other hand, in the worst case we need $\log_2 |\mathcal{X}_n|$ bits to represent an n-vertex graph from $\mathcal{X}$.

$$\frac{\log_2 |\mathcal{X}_n|}{\binom{n}{2}}$$

is the best possible coefficient of compressibility for representing graphs in $\mathcal{X}_n$.

Speed of **hereditary** graph classes

Alekseev V.E. (1982) showed that for every hereditary class $\mathcal{X}$ the limit $\lim\limits_{n\to\infty} \log_2 |\mathcal{X}_n| / \binom{n}{2}$ exists.

Alekseev V.E. (1992), and Bollobás B. & Thomason A. (1994):
$$\lim_{n\to\infty} \log_2 |\mathcal{X}_n| / \binom{n}{2} \in \left\{1 - \frac{1}{k} \mid k \in \mathbb{N}\right\}$$

Speed of **hereditary** graph classes

Theorem (Alekseev V.E., 1992; Bollobás B. & Thomason A., 1994)

For every infinite proper hereditary class $\mathcal{X}$:

$$\log_2 |\mathcal{X}_n| = \left(1 - \frac{1}{k(\mathcal{X})}\right) \frac{n^2}{2} + o(n^2),$$

where $k(\mathcal{X}) \in \mathbb{N}$ is the index of class $\mathcal{X}$.

(i) For $k(\mathcal{X}) > 1$, $\log_2 |\mathcal{X}_n| = \Theta(n^2)$

(ii) For $k(\mathcal{X}) = 1$, $\log_2 |\mathcal{X}_n| = o(n^2)$

Jumps in the speed of hereditary graph classes

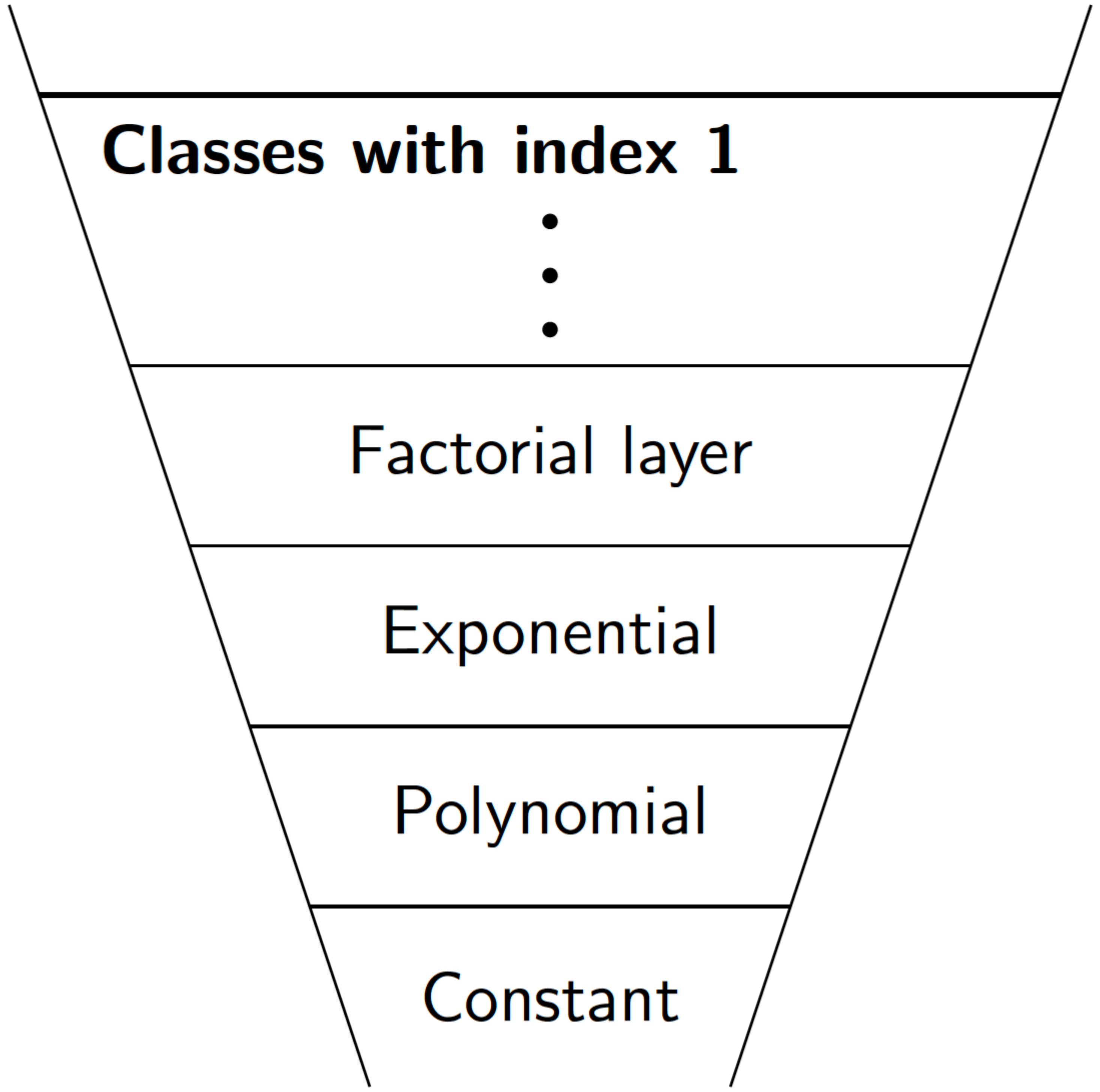
Let $k(\mathcal{X}) = 1$

Question

What are possible rates of growth of the function $\log_2 |\mathcal{X}_n|$*?*

Scheinerman E.R. & Zito J. (1994)

- Constant classes: $\log_2 |\mathcal{X}_n| = \Theta(1)$.
- Polynomial classes: $\log_2 |\mathcal{X}_n| = \Theta(\log n)$.
- Exponential classes: $\log_2 |\mathcal{X}_n| = \Theta(n)$.
- Factorial classes: $\log_2 |\mathcal{X}_n| = \Theta(n \log n)$.
- All other classes are superfactorial.

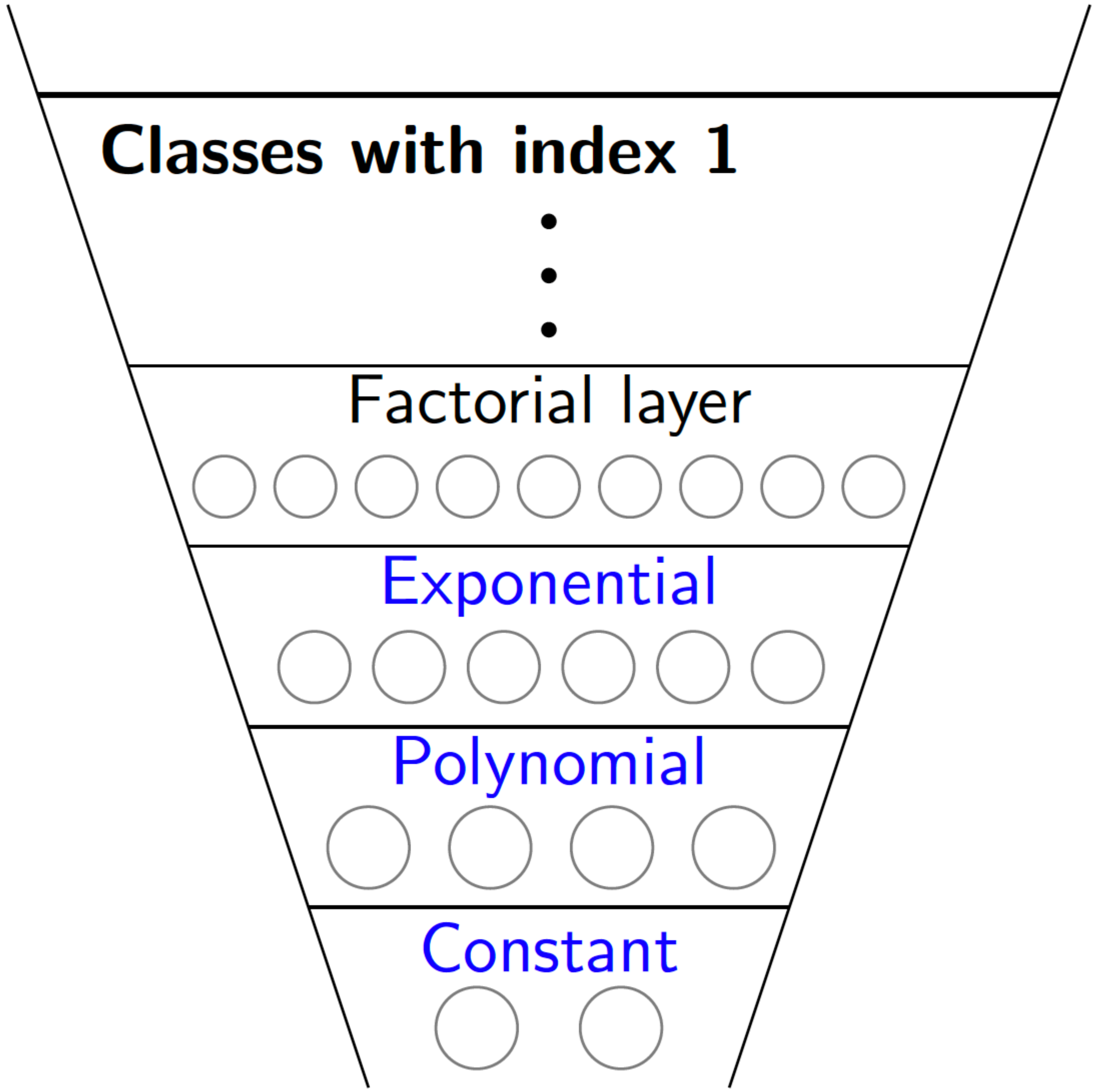


Structure of subfactorial classes

Alekseev V.E. (1997), and Balogh J., Bollobás B. & Weinreich D. (2000)

- Constant classes: $\log_2 |\mathcal{X}_n| = \Theta(1)$.
- Polynomial classes: $\log_2 |\mathcal{X}_n| = \Theta(\log n)$.
- Exponential classes: $\log_2 |\mathcal{X}_n| = \Theta(n)$.
- Factorial classes: $\log_2 |\mathcal{X}_n| = \Theta(n \log n)$.
- All other classes are superfactorial.

1. Structural characterizations of the first three layers.
2. All minimal classes in each of the layers.



Structure of **factorial** classes

Challenge: find a structural characterisation of the factorial layer

Except the definition, nothing common to **all** factorial classes is known

However, it was conjectured that **every factorial hereditary class** admits an **implicit representation** (or adjacency labels of size $O(\log n)$)

Implicit representation

Given a class $\mathcal{X}$ find an algorithm $\mathcal{A}$
such that for every n-vertex graph in $\mathcal{X}$ there is a labeling

- $v \mapsto \ell(v)$; and
- $v \sim w \iff \mathcal{A}[\ell(v), \ell(w)] = 1$; and
- labels are "short" ($O(\log n)$ bits).

Implicit representation and **universal graphs**

Given a class $\mathcal{X}$ find an algorithm $\mathcal{A}$
such that for every n-vertex graph in $\mathcal{X}$ there is a labeling

- $v \mapsto \ell(v)$; and
- $v \sim w \iff \mathcal{A}[\ell(v), \ell(w)] = 1$; and
- labels are "short" ($O(\log n)$ bits).

Universal Graph (sequence), for a graph class $\mathscr{X}$, of size $m(n)$ is a sequence $U = (U_n)_{n\in\mathbb{N}}$ of graphs with $|U_n| = m(n)$ such that, for all $n \in \mathbb{N}$, every graph $G \in \mathscr{X}_n$ is an induced subgraph of U_n.

Theorem (S. Kannan, M. Naor, S. Rudich, 1992) A class $\mathscr{X}$ admits an implicit representation if and only if it has a universal graph of size poly(n).

Implicit Graph Conjecture

Implicit Graph Conjecture

Problem (S. Kannan, M. Naor, S. Rudich, 1992)

Is it true that every hereditary class $\mathcal{X}$ with $|\mathcal{X}_n| = 2^{O(n \log n)}$ admits an implicit representation?

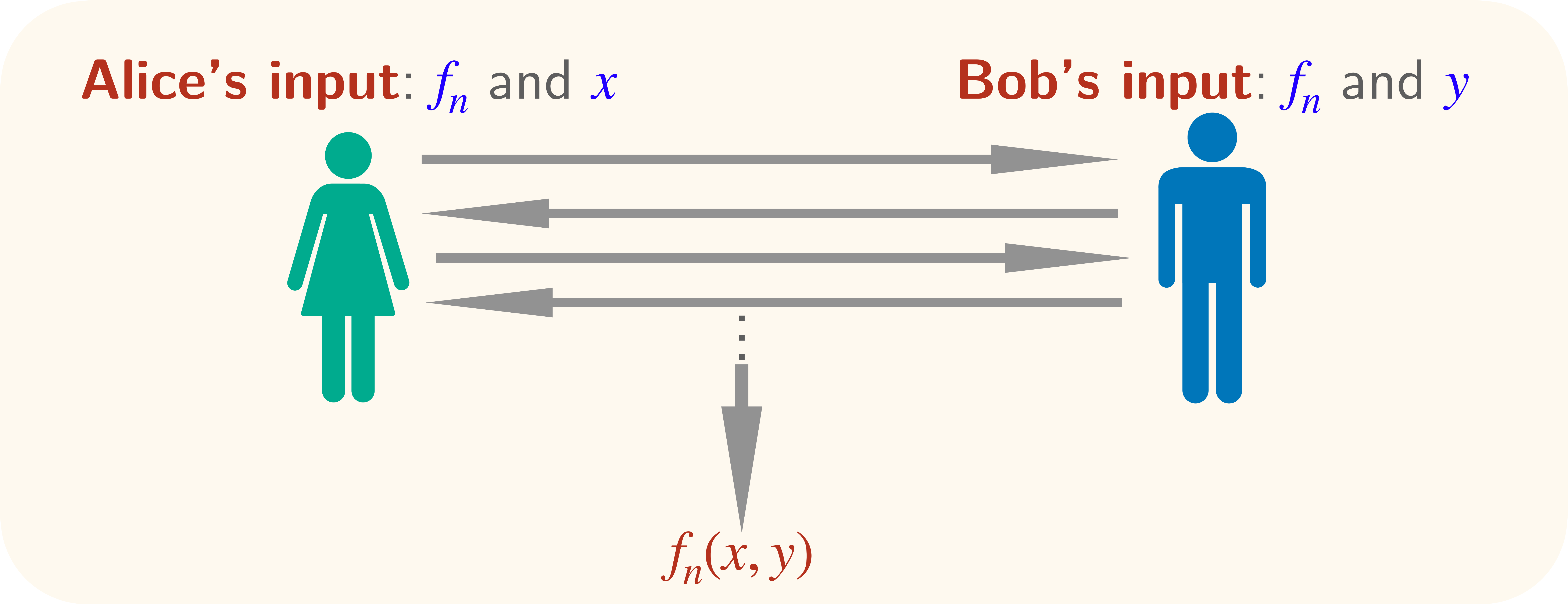
Implicit Graph Representation Conjecture (J. Spinrad, 2003)

Every hereditary class $\mathcal{X}$ with $|\mathcal{X}_n| = 2^{O(n \log n)}$ admits an implicit representation.

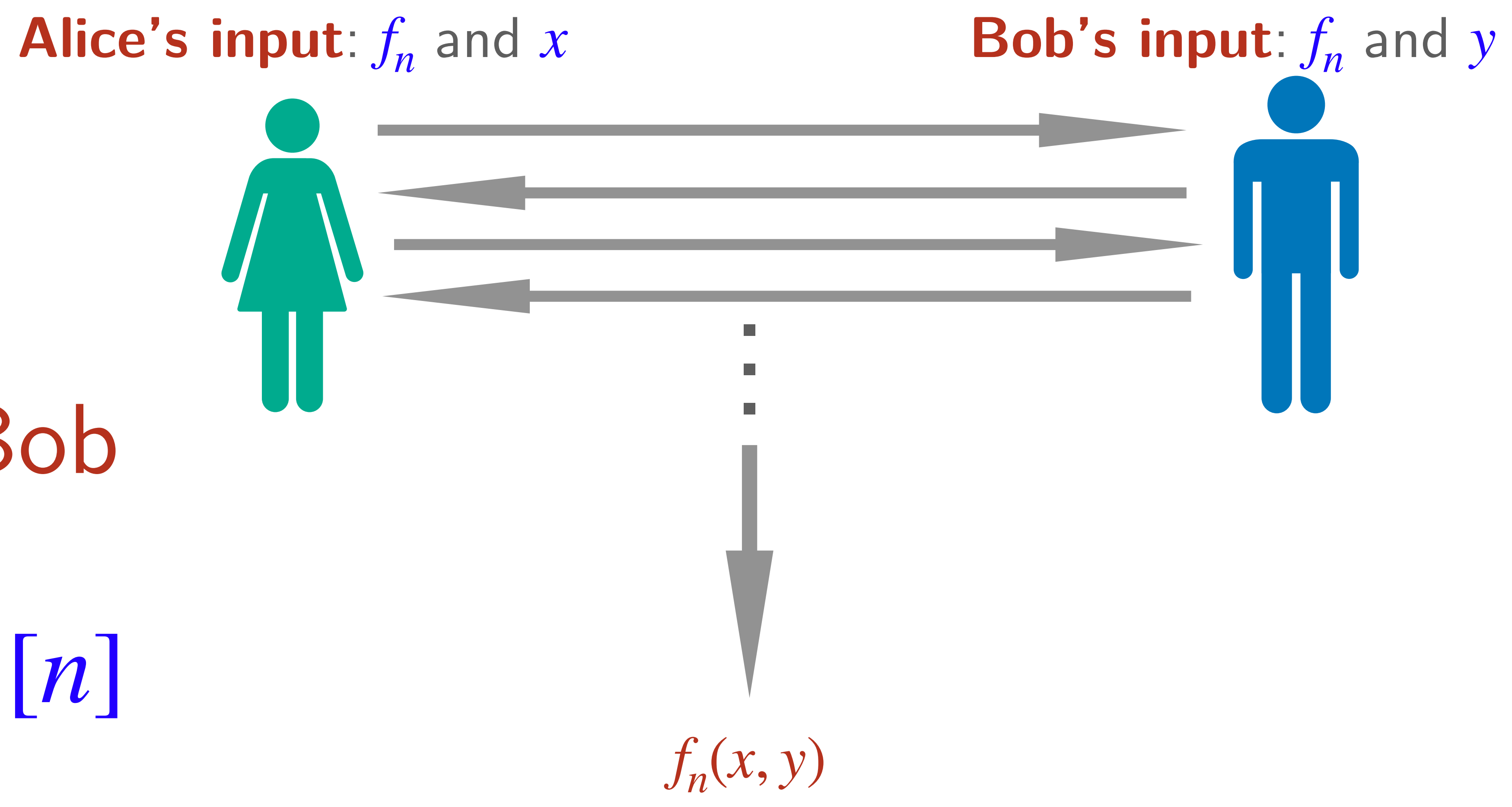
Communication Complexity Problems

Communication Complexity problems

- 2 parties: Alice and Bob
- Target function $f_n : [n] \times [n] \to \{0,1\}$ is known by Alice and Bob
- Alice receives an input $x \in [n]$ and Bob receives an input $y \in [n]$
- Alice and Bob exchange (single bit) messages in turn in order to find $f_n(x, y)$



Communication Complexity problems



- 2 parties: Alice and Bob
- Target function $f_n : [n] \times [n] \to \{0,1\}$ is known by Alice and Bob
- Alice receives an input $x \in [n]$ and Bob receives an input $y \in [n]$
- Alice and Bob exchange (single bit) messages in turn in order to find $f_n(x, y)$
- The total size (*in bits*) of exchanged messages is the cost of the communication protocol
- The communication complexity (or communication cost) of f_n, denoted $CC(f_n)$, is the minimum cost of a communication protocol that computes f_n
- A communication problem is a sequence $F = (f_n)_{n \in \mathbb{N}}$
- A communication cost of F is the function $CC(F) : n \mapsto CC(f_n)$

Examples

- **Equality problem**
  - $\text{Equality}_n : [n] \times [n] \rightarrow \{0,1\}$, where $\text{Equality}_n(x, y) = 1$ if an only if $x = y$
  - Communication complexity of Equality: $\lceil \log n \rceil + 1$
- **Greater-Than problem**
  - $\text{GT}_n : [n] \times [n] \rightarrow \{0,1\}$, where $\text{GT}_n(x, y) = 1$ if an only if $x \leq y$
  - Communication complexity of GT: $\lceil \log n \rceil + 1$

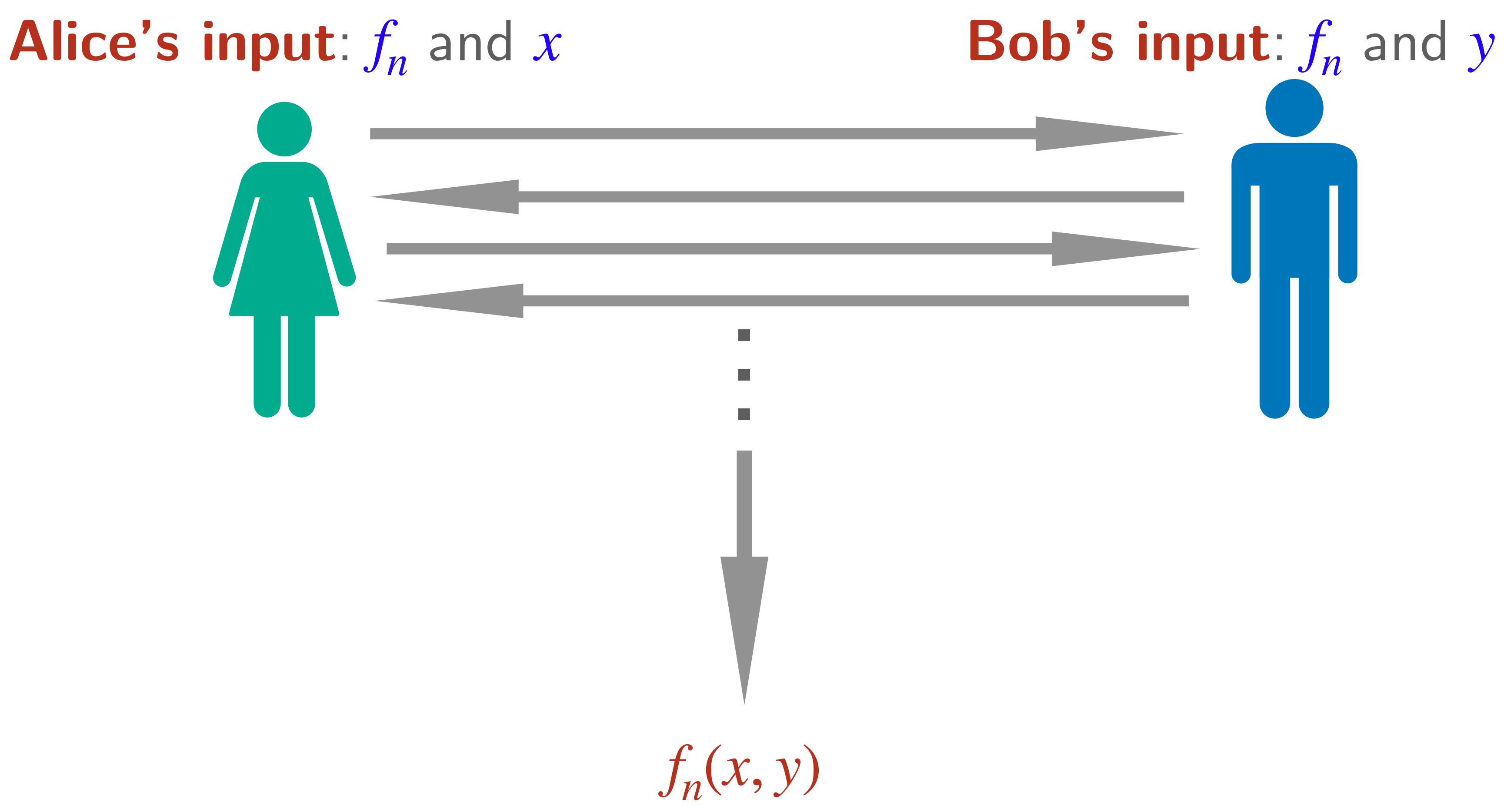
From **Communication Complexity** to **Adjacency Labelling**

From a **Communication Complexity** problem to **Adjacency Labelling**

We can think of f_n

as a bipartite graph $G_n = ([n], [n], E)$,

where $E = \{(x, y) \in [n] \times [n] \mid f(x, y) = 1\}$



Alice and Bob compute, in an interactive way, adjacency of two given vertices x and y

One can use
- messages sent by Alice (Alice's protocol) as labels for vertices in the left part
- messages sent by Bob (Bob's protocol) as labels for vertices in the right part

Given labels of two vertices from different parts the decoder executes protocol to decide adjacency of the vertices

From a **Communication Complexity** problem to **Adjacency Labelling**

Alice and Bob compute, in an interactive way, adjacency of two given vertices x and y

One can use
- messages sent by Alice (Alice's protocol) as labels for vertices in the left part
- messages sent by Bob (Bob's protocol) as labels for vertices in the right part

Given labels of two vertices from different parts the decoder executes protocol to decide adjacency of the vertices

Because the communication between the parties is interactive (e.g. next message of Bob depends on all previous messages by Alice and Bob) we need to encode all possible "conversations" in the label.

If the communication cost of a protocol is c, then it can be stored as a binary tree with 2^c nodes.

A communication protocol of cost c gives adjacency labels of size $O(2^c)$.

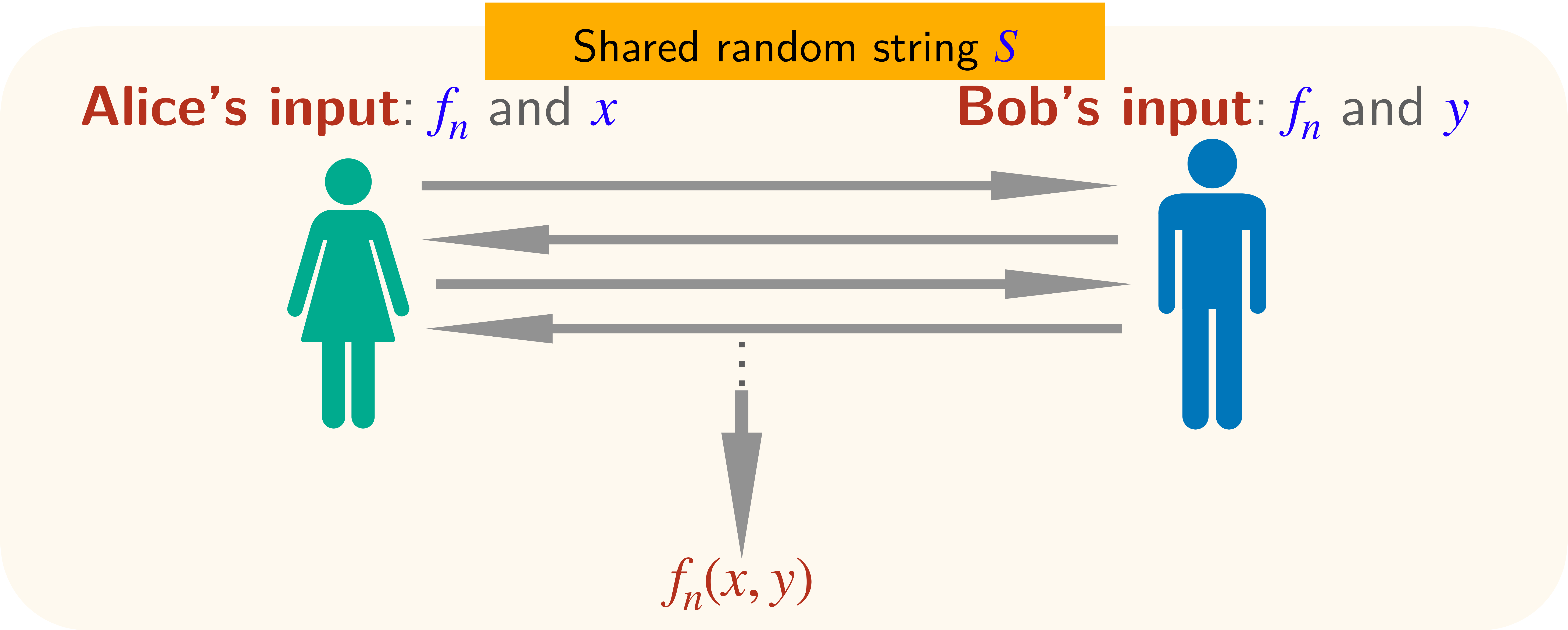
Examples

- **Equality problem**
  - $\mathsf{Equality}_n : [n] \times [n] \to \{0,1\}$, where $\mathsf{Equality}_n(x, y) = 1$ if an only if $x = y$
  - Corresponds to a matching graph: nK_2
- **Greater-Than problem**
  - $\mathsf{GT}_n : [n] \times [n] \to \{0,1\}$, where $\mathsf{GT}_n(x, y) = 1$ if an only if $x \leq y$
  - Corresponds to a half graph

Randomized Communication Complexity Problems

Randomized Communication Complexity problems

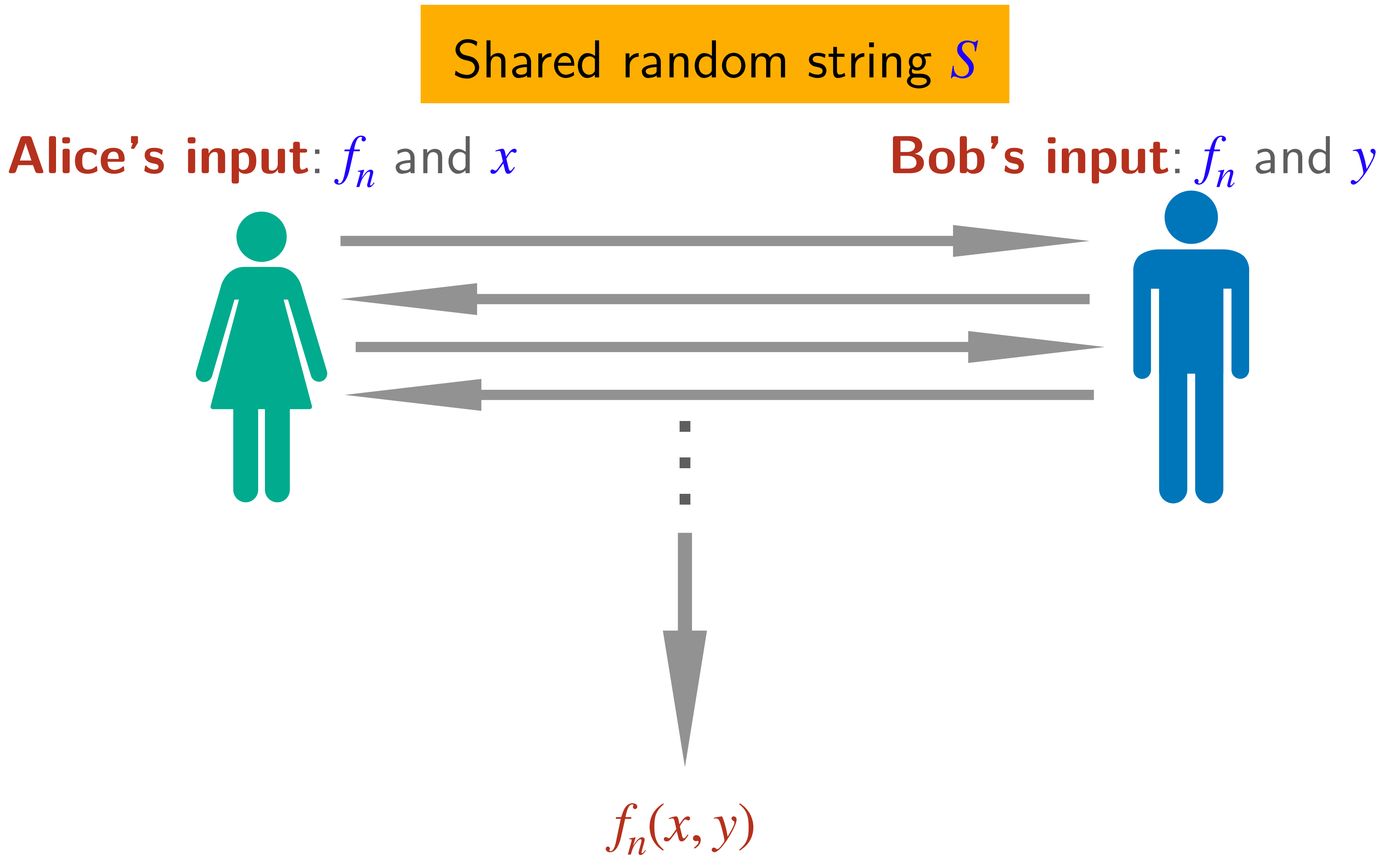
- 2 parties: Alice and Bob
- Target function $f_n : [n] \times [n] \to \{0,1\}$ is known by Alice and Bob
- Alice receives an input $x \in [n]$ and Bob receives an input $y \in [n]$
- Alice and Bob exchange (single bit) messages in turn in order to find $f_n(x, y)$
- Alice and Bob have access to a random string S



Randomized Communication Complexity problems

- A **randomised protocol** π is a distribution over deterministic protocols such that for $\forall x, y \in [n]$

$$\mathbb{P}\left[\pi(x,y) = f_n(x,y)\right] \geq 2/3$$



- The maximum total size (*in bits*) of exchanged messages is the cost of the randomised protocol π

- The randomised communication complexity of f_n, denoted $CC^R(f_n)$, is the minimum cost of a randomised communication protocol that computes f_n

- A communication cost of $F = (f_n)_{n\in\mathbb{N}}$ is the function $CC^R(F) : n \mapsto CC^R(f_n)$

Examples

- **Equality problem**
  - $\text{Equality}_n : [n] \times [n] \to \{0,1\}$, where $\text{Equality}_n(x, y) = 1$ if an only if $x = y$
  - **Randomized** Communication complexity of Equality: $O(1)$

- **Greater-Than problem**
  - $\text{GT}_n : [n] \times [n] \to \{0,1\}$, where $\text{GT}_n(x, y) = 1$ if an only if $x \leq y$
  - **Randomized** Communication complexity of GT: $\Omega(\log \log n)$

Constant-cost randomized communication problems

Open problem: Characterise communication problems that admit a constant-cost randomized communication protocol

From **Randomized Communication Complexity** to
Randomized Adjacency Labelling (or **Probabilistic Universal Graphs**)

From **Randomized Communication Complexity** to **Probabilistic Universal Graphs (PUGs)**

Probabilistic Universal Graph (sequence), for a graph family $\mathcal{X}$, of size $m(n)$ is a sequence $U = (U_n)_{n \in \mathbb{N}}$ of graphs with $|U_n| = m(n)$ such that, for all $n \in \mathbb{N}$ and all $G \in \mathcal{X}_n$ the following holds: there exists a probability distribution over the mappings $\phi : V(G) \to V(U_n)$ such that

$$\forall u, v \in V(G) \qquad \mathbb{P}_{\phi}\left[(\phi(u), \phi(v)) \in E(U_n) \iff (u, v) \in E(G)\right] \geq 2/3$$

Nathan Harms.
"Universal Communication, Universal Graphs, and Graph Labeling." *(*ITCS 2020*)*

Pierre Fraigniaud, Amos Korman.
"On randomized representations of graphs using short labels." (SPAA 2009)

Correspondence between

Communication Problems and **Adjacency Labelling for classes of graphs**

Communication Problems vs Adjacency Labelling for hereditary graph classes

1. Let $F = (f_n)_{n \in \mathbb{N}}$ be communication problem:

   1. G_i is the bipartite graph corresponding to f_i

   2. $\mathscr{Y}(F)$ is the hereditary closure of $\{G_1, G_2, \dots\}$

2. Let $\mathscr{X}$ be a hereditary class:

   1. $\mathrm{Adj}_{\mathscr{X}} = (f_n)_{n \in \mathbb{N}}$ is a communication problem such that f_n is a "hardest" function corresponding to a graph in $\mathscr{X}_n$

Constant cost problems vs constant-size PUGs

Theorem 1. For any communication problem $F = (f_n)_{n \in \mathbb{N}}$ and hereditary graph class $\mathscr{X}$:

1. F has constant randomized communication complexity if and only if $\mathscr{Y}(F)$ has a constant-size PUG
2. $\mathscr{X}$ has a constant-size PUG if and only if $\mathsf{Adj}_{\mathscr{X}}$ has constant randomized communication complexity

Open problem: Characterise communication problems that admit a constant-cost randomized communication protocol

Equivalent open problem: Characterise hereditary graph classes that admit a constant-size PUG

Constant-size PUGs

Theorem 2. If a class $\mathscr{X}$ has a constant-size PUG then it admits an adjacency labelling scheme with labels of size $O(\log n)$.

Corollary. The classes that have a constant-size PUG is a subset of the classes satisfying the Implicit Graph Conjecture.

Thus by characerizing classes that admit a constant-size PUG we:

1. characterise communication problems with a constant randomized communication complexity

2. make progress towards the Implicit Graph Conjecture

Necessary condition

Lemma. If a class of bipartite graphs $\mathscr{X}$ has a constant-size PUG then it excludes a half graph as an induced subgraph.

Proof sketch.

If $\mathscr{X}$ contains all half graphs, then the corresponding communication problem $\mathsf{Adj}_{\mathscr{X}}$ is at least as hard as the **Greater-Than** problem (which has complexity $\Omega(\log \log n)$), and

thus it does not have a constant-cost randomized protocol.

Therefore $\mathscr{X}$ cannot have a constant-size PUG.

A class of bipartite graphs that excludes a half graph is called **edge-stable**.

Many factorial edge-stable classes of graphs have constant-size PUG

- Graph of bounded degeneracy
- All edge-stable $\{H\}$-free bipartite graphs
- All edge-stable classes of bounded twin-width

  Jakub Gajarský, Michał Pilipczuk, Szymon Toruńczyk
  "Stable graphs of bounded twin-width" (LICS 2022)

- All edge-stable classes of permutation graphs
- All edge-stable classes of interval graphs
- All edge-stable classes of unit disk graphs
- ...

Probabilistic Implicit Graph Conjecture

Probabilistic Implicit Graph Conjecture

Probabilistic Implicit Graph Conjecture

A hereditary class of bipartite graphs has a constant-size PUG if and only if it is factorial and edge-stable two weeks later...

Probabilistic Implicit Graph Conjecture

is false

Probabilistic Implicit Graph Conjecture is False

Lianna Hambardzumyan, **Hamed Hatami**, **Pooya Hatami**
studied (independently and concurrently to our work) communication problems with constant randomized complexity

Lianna Hambardzumyan, **Hamed Hatami**, **Pooya Hatami**
"Dimension-free Bounds and Structural Results in Communication Complexity"
Israel Journal of Mathematics 253(2) (2023): 555-616.

Lianna Hambardzumyan, **Hamed Hatami**, **Pooya Hatami**
"A counter-example to the probabilistic universal graph conjecture via randomized communication complexity"
Discrete Applied Mathematics 322 (2022): 117-122.

Probabilistic Implicit Graph Conjecture is False

Construction:

Sequence of functions (bipartite graphs) $M = (M_n)_{n \in \mathbb{N}}$ such that

1. Randomized communication complexity of M is unbounded (i.e. $\omega(1)$)
2. Every $a \times b$ submatrix of M_n with $a, b \leq \sqrt{n}$ contains a row or a column with at most four 1's

In the graph-theoretical language (2) means that every subgraph of M_n with at most $\sqrt{n}$ vertices in each of the parts is 4-degenerate.

It implies that the hereditary closure $\mathcal{X}$ of $\{M_1, M_2, \dots\}$ is

1. **Edge-stable**, i.e. excludes some half graph
2. **Factorial**

another week later...

The Implicit Graph Conjecture
is false!

Hamed Hatami, **Pooya Hatami**
"The Implicit Graph Conjecture is False."
FOCS (2022)

The Implicit Graph Conjecture is False

Proof sketch:

1. A bipartite graph is $G = ([n], [n], E)$ is **good** if
   1. $|E| = \lfloor n^{2-\epsilon} \rfloor$ (where ϵ is some fixed constant)
   2. Every induced subgraph of G with at most $\sqrt{n}$ vertices in each of the parts is c-degenerate (where c is some fixed constant)

2. Let $\mathscr{G}$ be the family of all good graphs

The Implicit Graph Conjecture is False

Proof sketch (2):

Claim: For every n, let $\mathcal{M}_n \subseteq \mathcal{G}_n$ be any subset with $|\mathcal{M}_n| \leq 2^{\sqrt{n}}$.

Then the hereditary closure of $\bigcup_{n \in \mathbb{N}} \mathcal{M}_n$ is at most **factorial**.

Counting: For every large n

- there are **a lot** of sets $\mathcal{M}_n \subseteq \mathcal{G}_n$ with $|\mathcal{M}_n| = 2^{\sqrt{n}}$.

- **so many**, that there exists such a set $\mathcal{M}'_n \subseteq \mathcal{G}_n$ that **cannot be represented** by a universal graph of polynomial size $2^{O(\log n)}$, in fact, of size smaller than $2^{n^{0.5-\delta}}$ for some constant δ

The Implicit Graph Conjecture is False

Proof sketch (3):

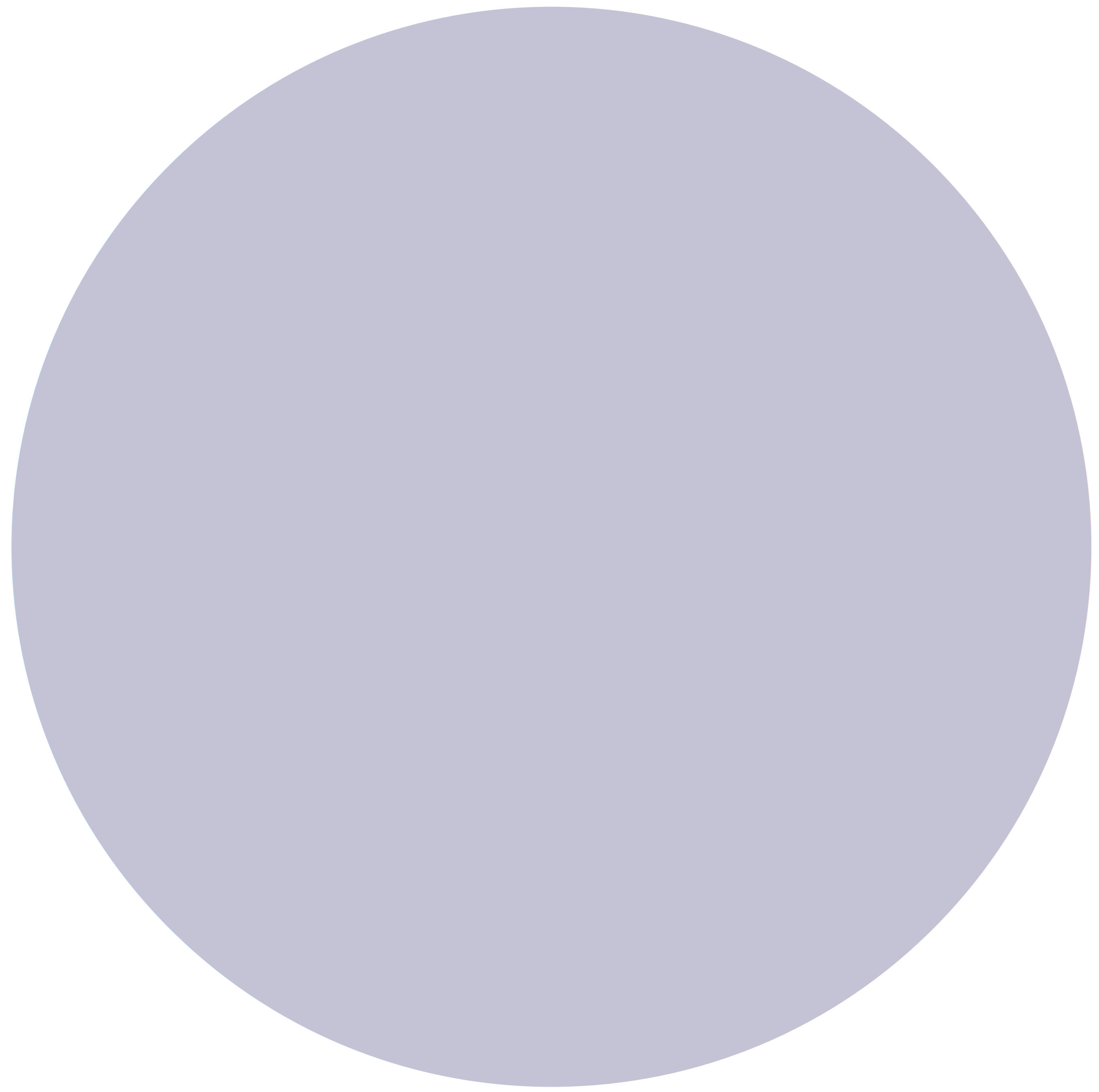
Then the hereditary closure of $\bigcup_{n \in \mathbb{N}} \mathscr{M}'_n$ is most factorial, but does not admit a universal graph sequence of size smaller than $2^{n^{0.5-\delta}}$.

Conclusion

The Implicit Graph Conjecture

Factorial classes

Classes with $O(\log n)$ labelling scheme

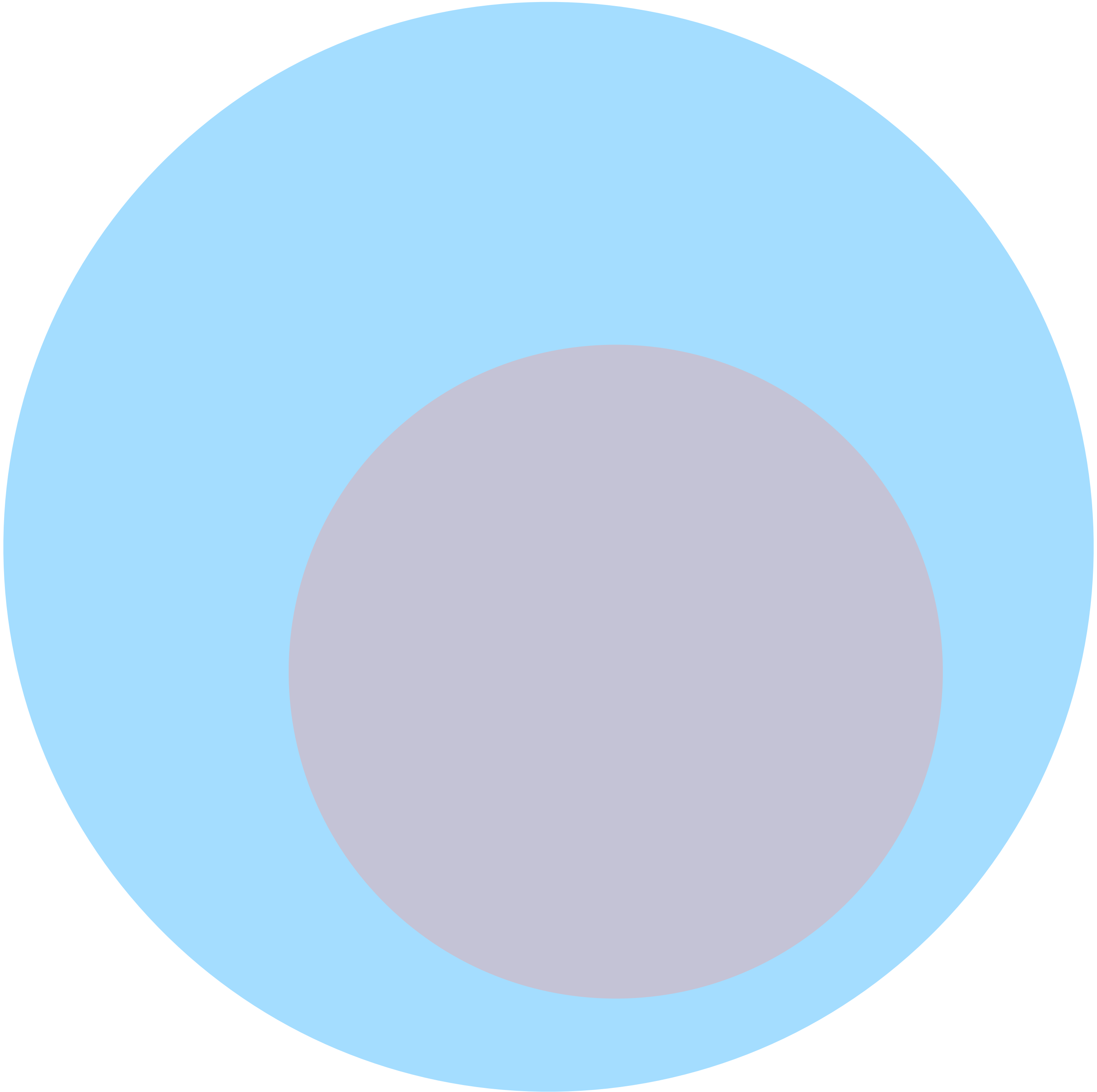


Conclusion

Reality

Factorial classes

Classes with $O(\log n)$ labelling scheme



Conclusion

Bad news for characterization of the factorial classes

However, opens up a new perspective for labelling schemes:

1. What are the classes of graphs that admit a $O(\log n)$ labelling scheme?
2. What are the edge-stable classes that admit a $O(\log n)$ labelling scheme?
3. What are the classes that admit a constant-size probabilistic universal graph?

Thank you!

Thank you!

Thank you to Jakub, Michał, Szymon!