

A “Boosting” Proof of the Weak Regularity Lemma

Luca Trevisan

U.C. Berkeley

*Ongoing joint work with Omer Reingold,
Madhur Tulsiani, Salil Vadhan*

Weak Regularity Lemma

- Szemerédi,
Frieze-Kannan



Dense Model Theorem

- Green-Tao,
Tao-Ziegler

Hard-Core Set Lemma

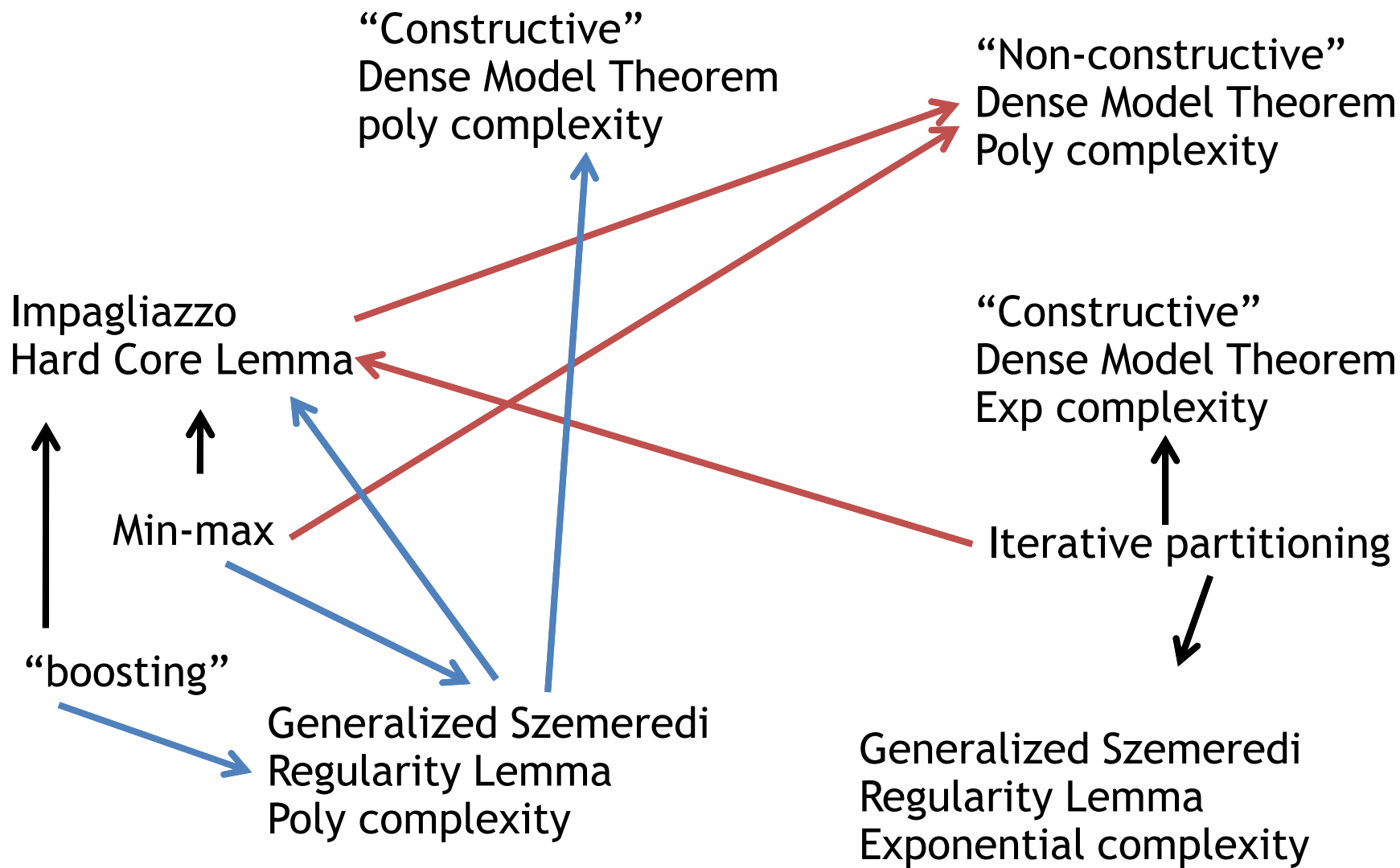
- Impagliazzo



Boosting

- Freund, Shapire

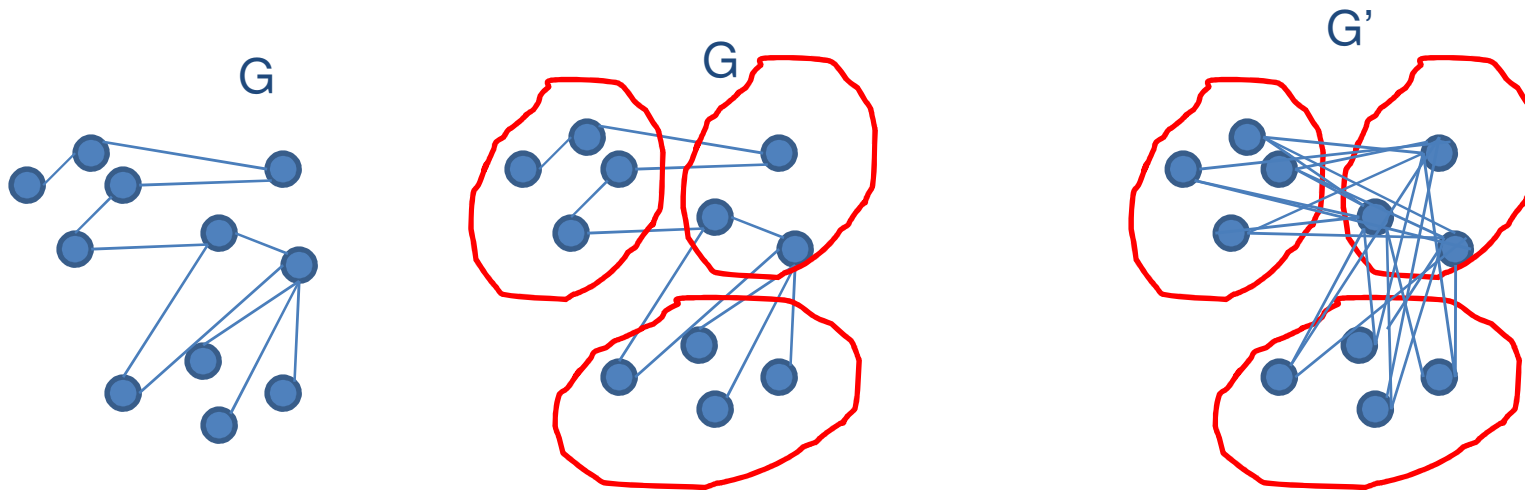




Regularity Lemma

Every dense graph is ε -approximated by a graph of “complexity” $O_\varepsilon(1)$

Weak Regularity Lemma

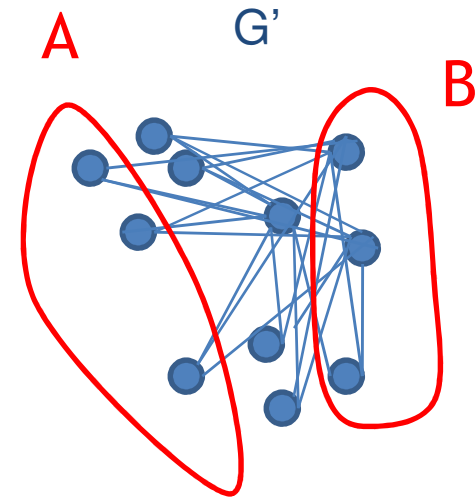
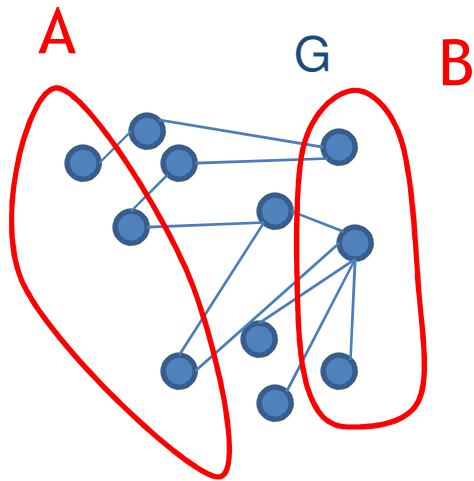


Given: dense graph G , approximation ε

There is partition of vertices into $\exp((1/\varepsilon)^{O(1)})$ blocks

S.t. G is “ ε –indistinguishable by cuts” from graph G' that has weighted complete bipartite subgraph between blocks

Weak Regularity Lemma



G is “ ϵ –indistinguishable by cuts” from G' iff

For every A, B

$$| \text{edges}_G(A, B) - \text{edges}_{G'}(A, B) | \leq \epsilon n^2$$

Relative Weak Regularity Lemma

Suppose $G_R = (V, E_R)$ is a “pseudorandom graph”

Then for every subgraph $G = (V, E)$

There is partition of G into $\exp(\varepsilon^{-O(1)})$ blocks
defining G' such that for all vertex sets A, B

$$| \text{edges}_G(A, B) - \text{edges}_{G'}(A, B) | \leq \varepsilon |E_R|$$

Proof

1. Start from a trivial partition
 2. If current partition does not work
 - Use counterexample sets A,B to refine
 - Go to 2
- Every step:
 - Increases # blocks by constant factor
 - Increases energy function by $\varepsilon^{O(1)}$
 - Energy function is ≤ 1

Dense Model Theorem

- If R pseudorandom in X
and D subset of R has size $\geq \delta |R|$
- Then there is M of size $\geq \delta |X|$ that is
indistinguishable from D

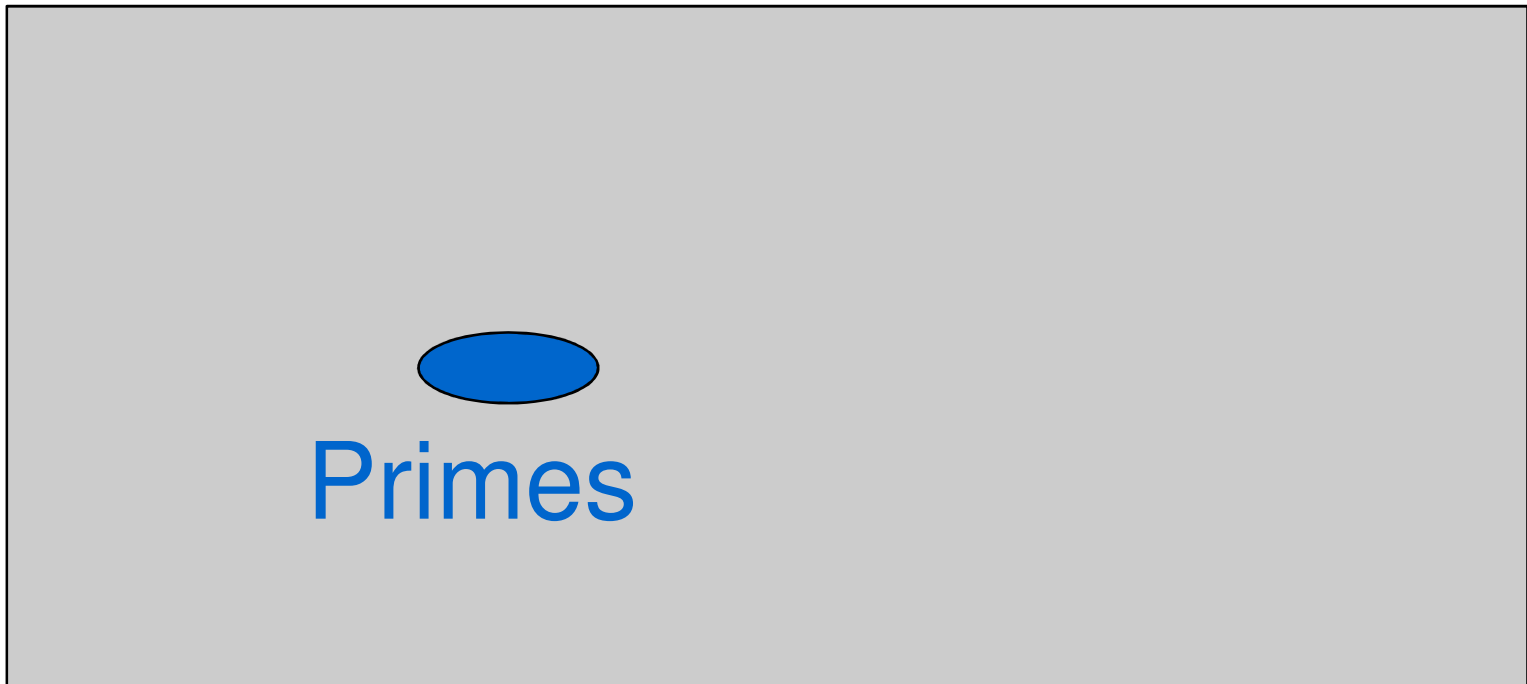
Green-Tao Theorem

The primes contain arbitrarily long arithmetic progressions

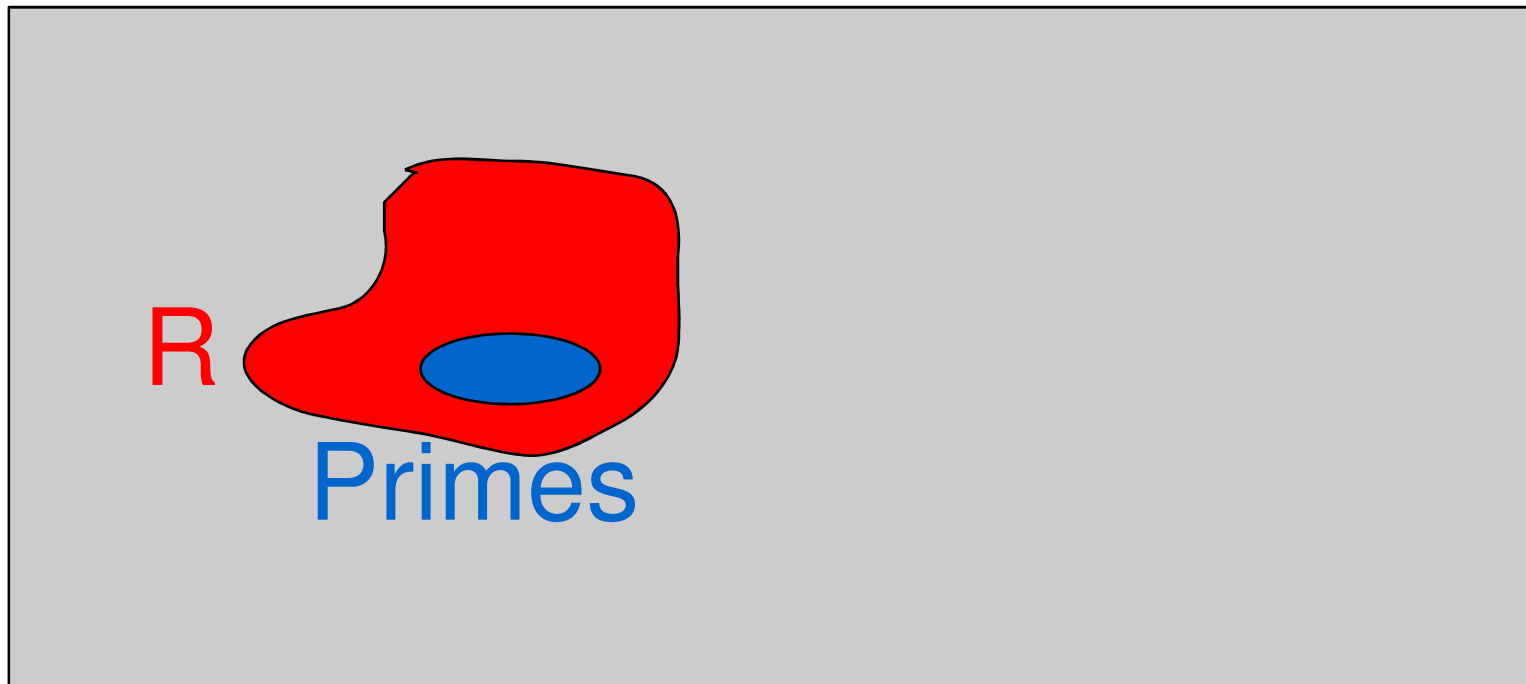
Proof:

1. Thm: primes have const density in almost primes
2. Thm: almost primes are pseudorandom
3. Thm: [Main]
if R is pseudorandom
and D has constant density in R ,
then there is M of constant density
indistinguishable from D
4. Thm: [Szemerédi] M has long APs
5. so do the primes

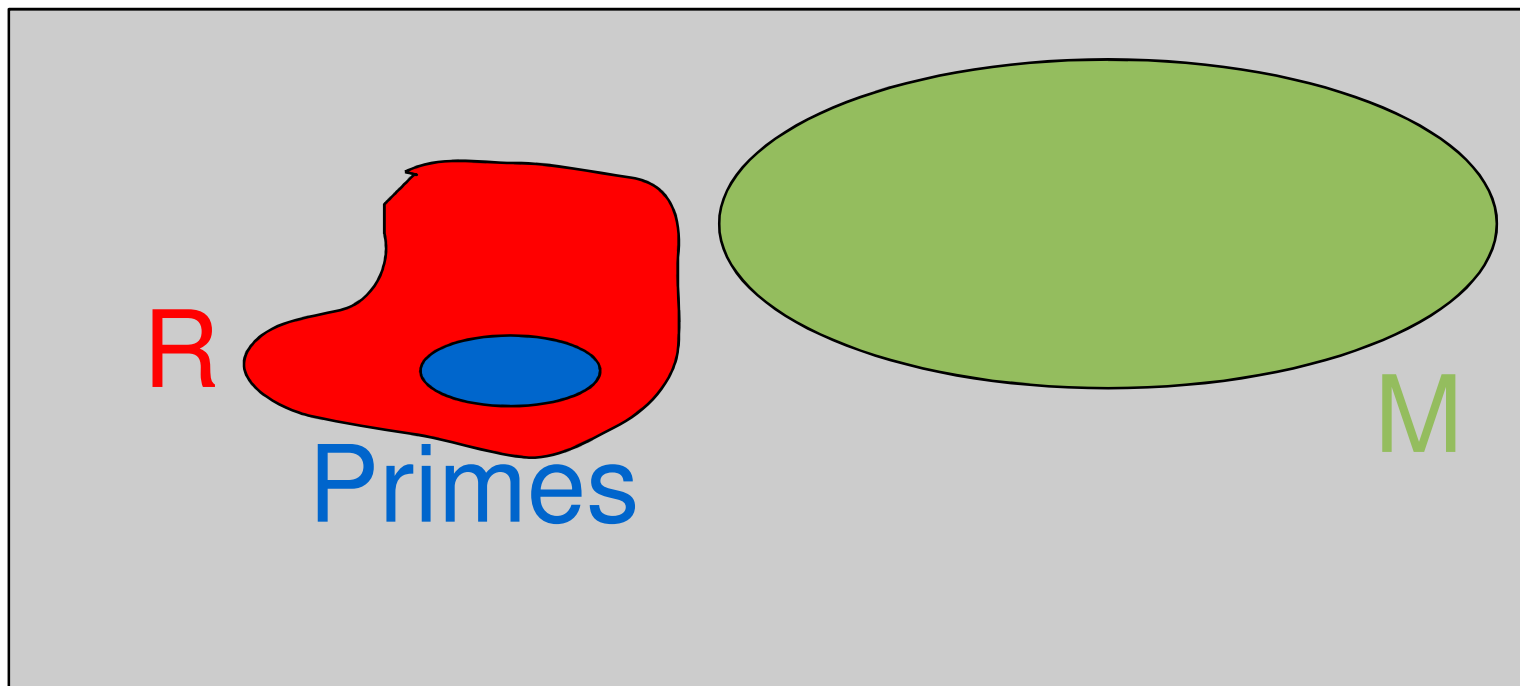
$\{1, \dots, N\}$



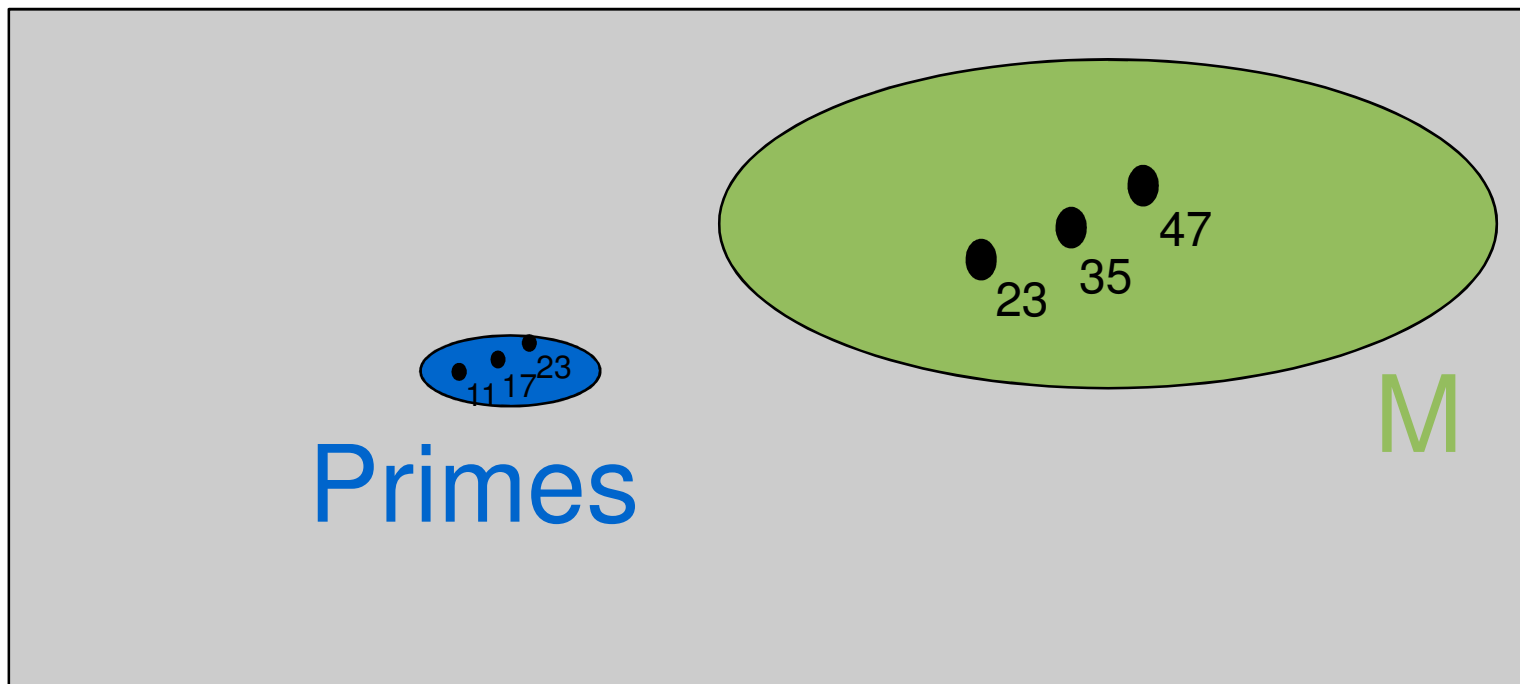
$\{1, \dots, N\}$



$\{1, \dots, N\}$



$\{1, \dots, N\}$



Key step

Dense Model Theorem:

- If R pseudorandom in X
and D subset of R has size $\geq \delta |R|$
- Then there is M of size $\geq \delta |X|$ that is
indistinguishable from D

Dense Model Theorem

Given:

- universe X ,
- collection F of functions $f: X \rightarrow [0,1]$
- parameter ε
- **pseudorandom** (depending on F, ε) subset R of X ,
- D subset of R of size $\geq \delta |R|$

There is: set M that is

- **Dense**: $|M| > \delta |X|$
- **Indistinguishable** from D : for all f in F ,
 $| \mathbf{E}_{x \sim D} f(x) - \mathbf{E}_{x \sim M} f(x) | < \varepsilon$

Dense Model Theorem

Given:

- universe X ,
- collection F of functions $f: X \rightarrow [0,1]$
- parameter ε
- **pseudorandom** (depending on F, ε) measure $\nu: X \rightarrow \mathbf{R}$
- measure $g: X \rightarrow \mathbf{R}$, $0 < g(x) < \nu(x)$

There is: bounded function $h: X \rightarrow [0,1]$ that is

- **Dense**: $\mathbf{E}_x g(x) = \mathbf{E}_x h(x)$
- **Indistinguishable** from g : for all f in F ,
$$| \mathbf{E}_x g(x)f(x) - \mathbf{E}_x h(x)f(x) | < \varepsilon$$

Constructive Dense Model Theorem

Given:

- X, F, ε , pseudorandom measure $v: X \rightarrow \mathbf{R}$,
measure $g: X \rightarrow \mathbf{R}$, $0 < g(x) < v(x)$

There is: bounded function $h: X \rightarrow [0, 1]$ that is

- **Dense**: $\mathbf{E}_x g(x) = \mathbf{E}_x h(x)$
- **Constructive**: there are f_1, \dots, f_k in F , $k = \varepsilon^{-O(1)}$ s.t.
 $h(x) = H(f_1(x), \dots, f_k(x))$
- **Indistinguishable** from g : for all f in F ,
 $|\mathbf{E}_x g(x)f(x) - \mathbf{E}_x h(x)f(x)| < \varepsilon$

Constructive Dense Model Theorem

Special case $v=1$:

- Given $X, F, \varepsilon, g: X \rightarrow [0, 1]$

There is: bounded function $h: X \rightarrow [0, 1]$ that is

- **Constructive**: there are f_1, \dots, f_k in F , $k = \varepsilon^{-O(1)}$ s.t.
$$h(x) = H(f_1(x), \dots, f_k(x))$$
- **Indistinguishable** from g : for all f in F ,
$$| \mathbf{E}_x g(x)f(x) - \mathbf{E}_x h(x)f(x) | < \varepsilon$$

Constructive Dense Model Theorem => Weak Regularity Lemma

Given $X, F, \varepsilon, g: X \rightarrow [0, 1]$

There is: bounded function $h: X \rightarrow [0, 1]$ that is

- **Constructive:** there are f_1, \dots, f_k in F , $k = \varepsilon^{-O(1)}$ s.t.
 $h(x) = H(f_1(x), \dots, f_k(x))$
- **Indistinguishable** from g : for all f in F ,
 $| \mathbb{E}_x g(x)f(x) - \mathbb{E}_x h(x)f(x) | < \varepsilon$

E.g.:

- X edges of complete graph on vertices V
- $F = \{ f_{S,T} \}$, $f_{S,T}(u,v) := (1 \text{ if } u \text{ in } S \text{ and } v \text{ in } T; 0 \text{ otherwise})$
- $g: X \rightarrow \{0, 1\}$ is a graph
- h is an approximation satisfying Weak Regularity Lemma

Proof of Constr Dense Mod Thm

Similar to proof of Relative Weak Regularity Lemma

Start from trivial partition of X

1. Define $h(x) :=$ average of g on block of x
2. If h not indistinguishable from g
 - Use distinguishing function f to refine partition
 - Go to 1

Each step increases energy function

Energy function is bounded

[need g to be bounded by 1 or by a p.r. measure]

Impagliazzo Hard-Core Set Lemma

If a computational problem is hard-on-average in a weak sense

There is a dense subset of inputs on which it is hard-on-average in a much stronger sense

Impagliazzo Hard-Core Set Lemma

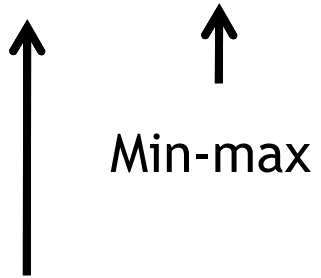
$g: X \rightarrow \{0,1\}$ computational problem

F family of algorithms, $f: X \rightarrow \{0,1\}$

- If
 - $\mathbf{P}_x[f(x) = g(x)] < 1 - \delta$ for all f in F^*
- Then there is set S of size $> 2 \delta |X|$ such that
 - $\mathbf{P}_{x \sim S} [f(x) = g(x)] < \frac{1}{2} + \epsilon$ for all f in F

$F^* :=$ functions of “complexity” $(\epsilon\delta)^{-O(1)}$ relative to F

Impagliazzo
Hard Core Lemma



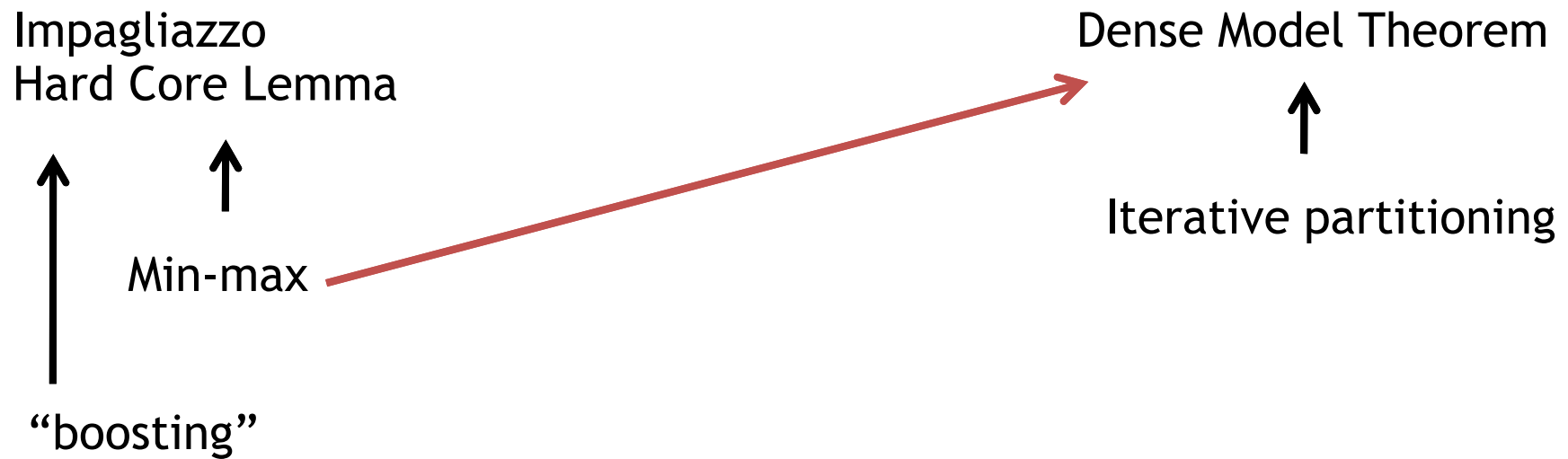
“boosting”

Dense Model Theorem



Iterative partitioning

New proof of D.M.T.



Gowers, Reingold-T-Tulsiani-Vadhan

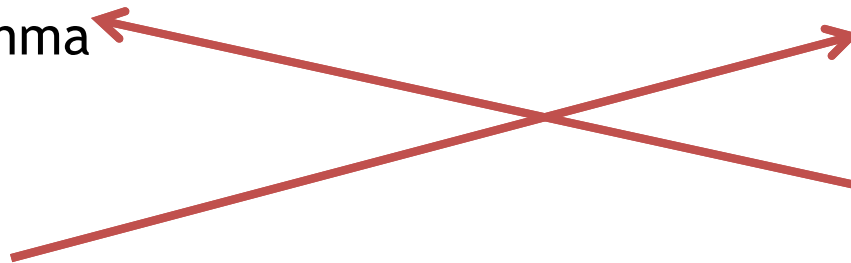
Impagliazzo
Hard Core Lemma

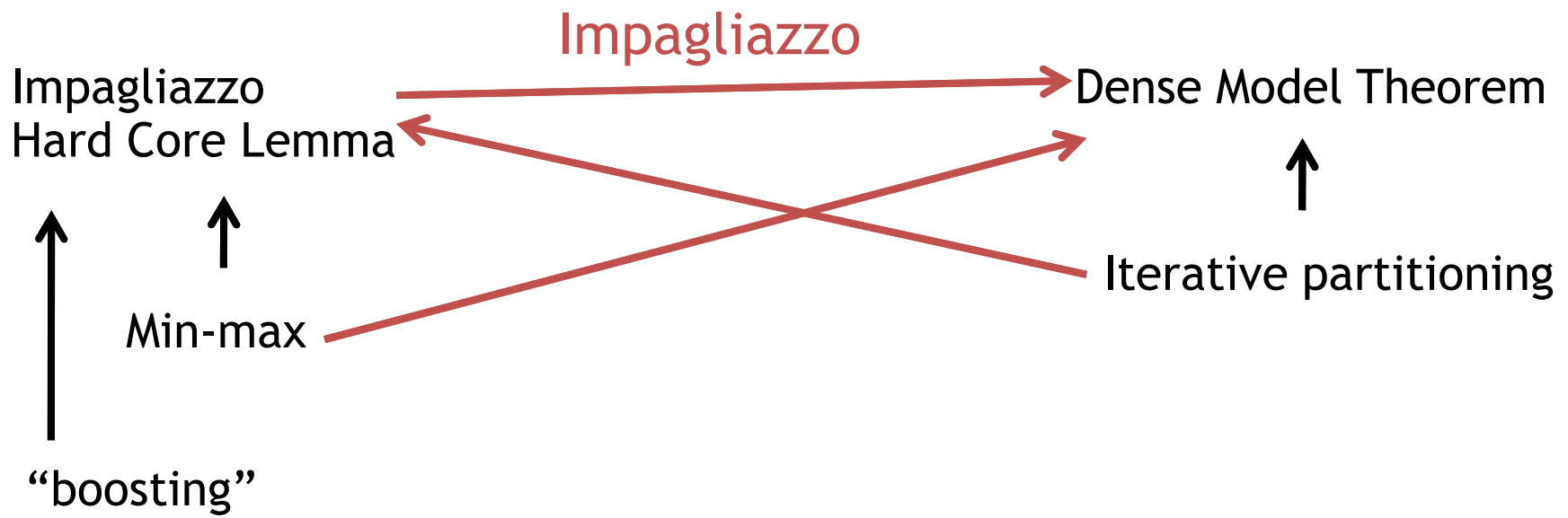
Dense Model Theorem

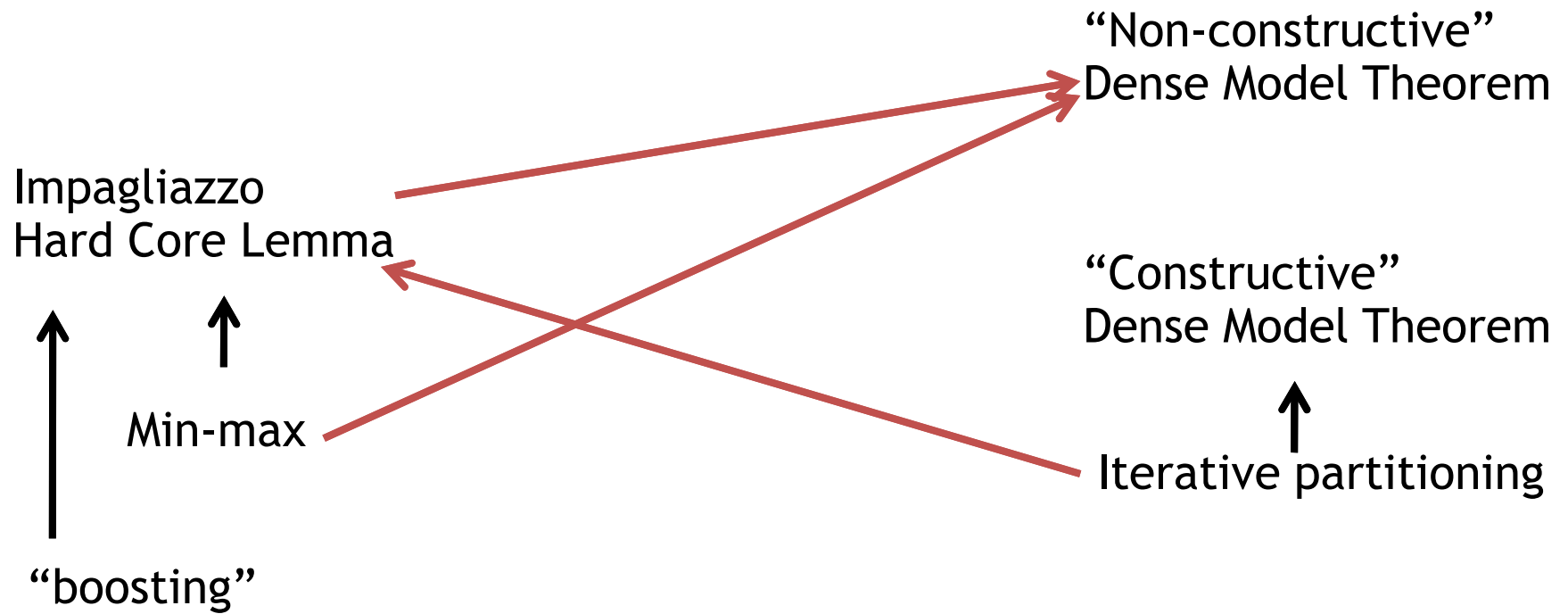
Min-max

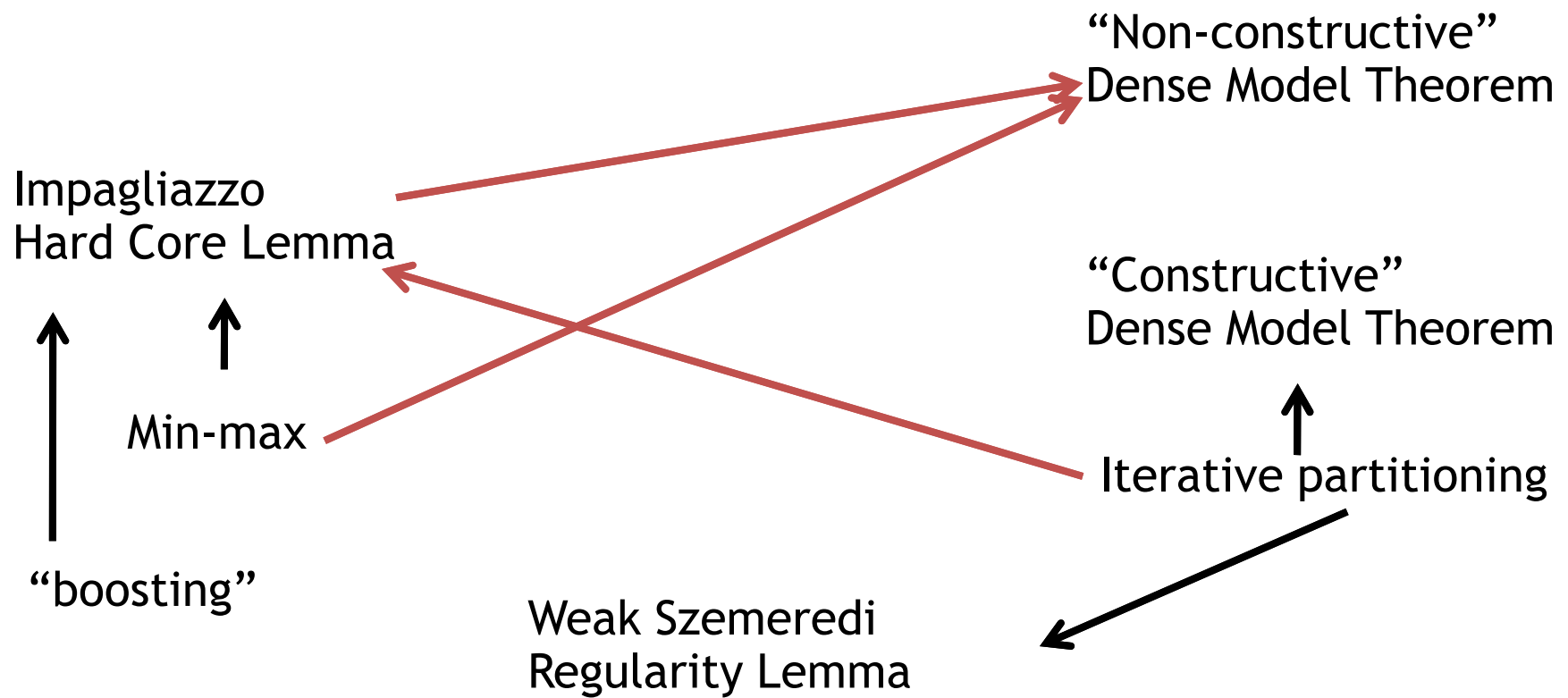
Iterative partitioning

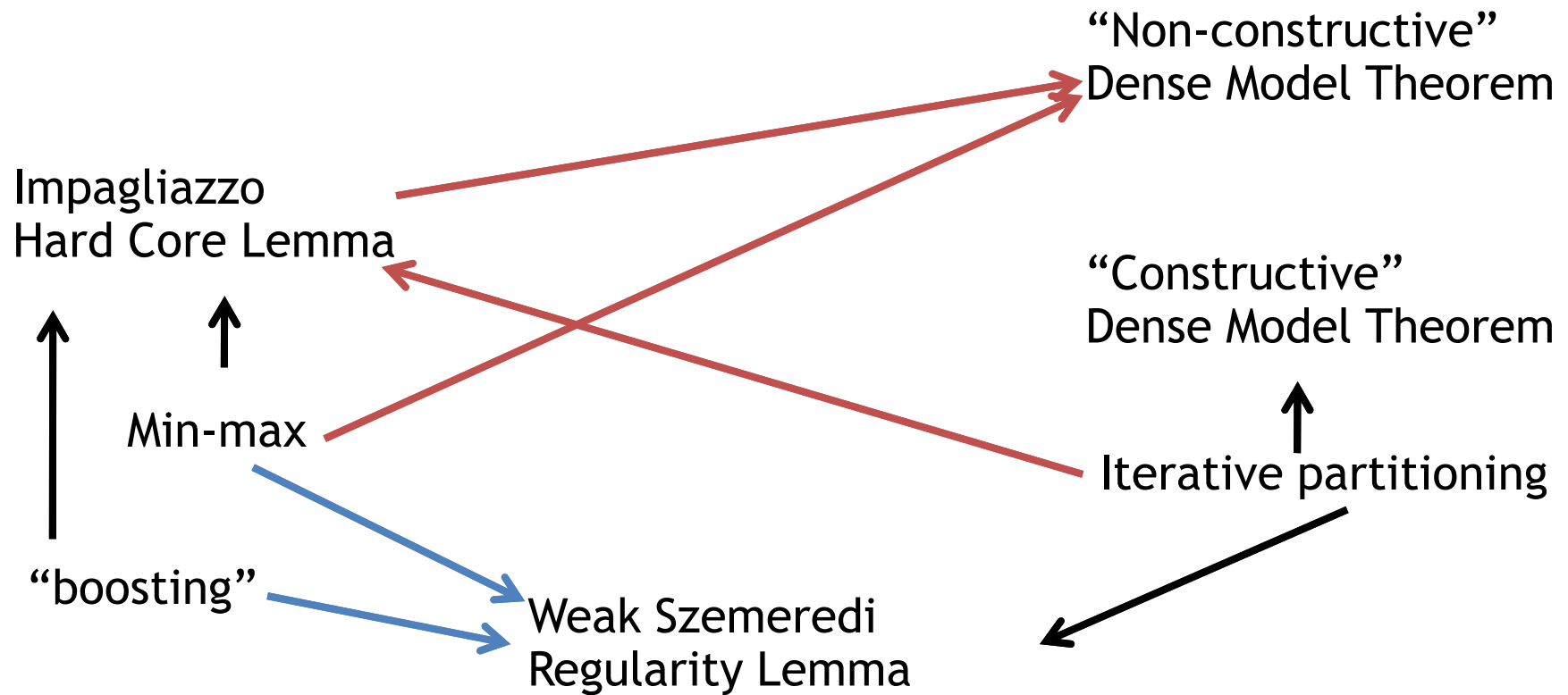
“boosting”

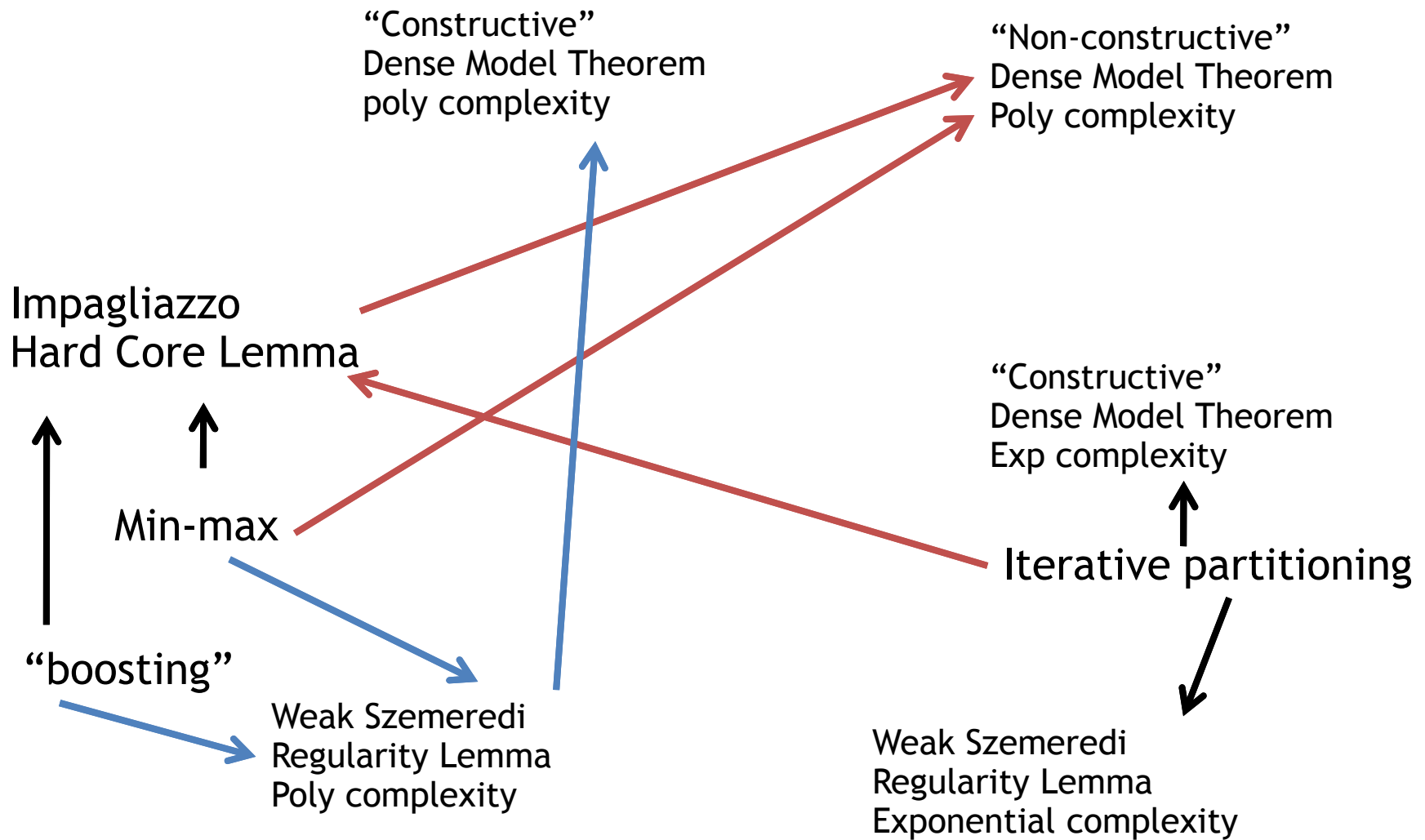


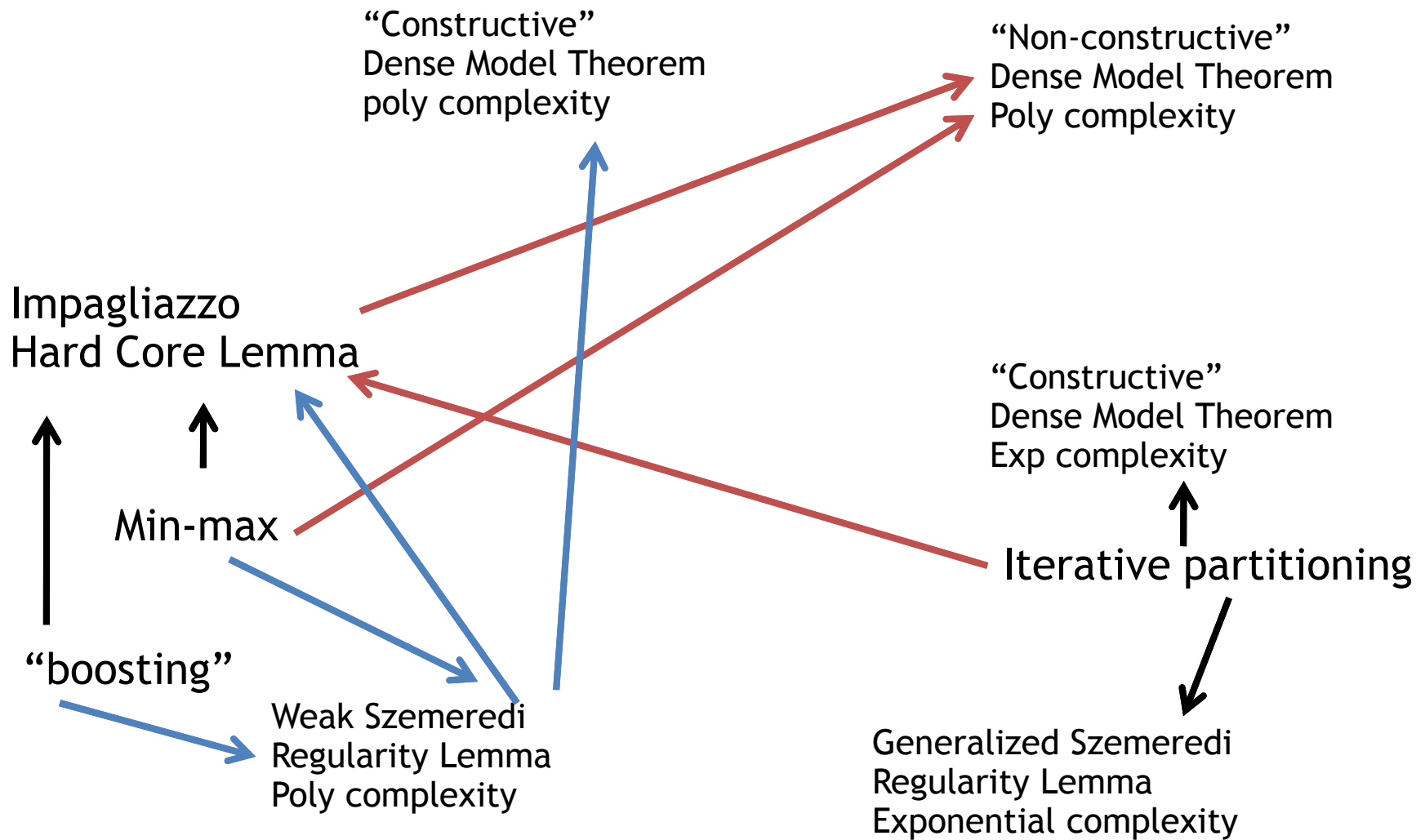












Constr. D.M.T. / Weak Regularity

- Finite space X [edges of complete graph]
- $g: X \rightarrow [0,1]$ [graph]
- ε
- $F = \{ f \}, f: X \rightarrow [0,1]$ [char functions of cuts]
- Can find approximating function $h: X \rightarrow [0,1]$
 $h(x) = H(f_1(x), \dots, f_k(x)), k = \text{poly}(1/\varepsilon), f_i() \text{ in } F$
- Such that for all f in F
 $| \mathbf{E} f(x)g(x) - \mathbf{E} f(x)h(x) | \leq \varepsilon$

- Finite space X [edges of complete graph]
- $g: X \rightarrow [0,1]$ [graph]
- ε
- $F = \{ f \}$, $f: X \rightarrow [0,1]$ [char functions of cuts]
- Can find approximating function $h: X \rightarrow [0,1]$
 $h(x) = H(f_1(x), \dots, f_k(x))$, $k = \text{poly}(1/\varepsilon)$, $f_i() \in F$
in standard proof, H has $\exp(1/\varepsilon)$ complexity
- Such that for all f in F
 $| \mathbf{E} f(x)g(x) - \mathbf{E} f(x)h(x) | \leq \varepsilon$

- Finite space X [edges of complete graph]
- $g: X \rightarrow [0,1]$ [graph]
- ε
- $F = \{ f \}$, $f: X \rightarrow [0,1]$ [char functions of cuts]
- Can find approximating function $h: X \rightarrow [0,1]$
 $h(x) = H(f_1(x), \dots, f_k(x))$, $k = \text{poly}(1/\varepsilon)$, $f_i() \in F$
in new “CS proofs”, H has $\text{poly}(1/\varepsilon)$ complexity
- Such that for all f in F
 $| \mathbf{E} f(x)g(x) - \mathbf{E} f(x)h(x) | \leq \varepsilon$

The “Boosting” Proof

- Finite space X , $g: X \rightarrow [-1,1]$, ε , $F=\{ f \}$, $f: X \rightarrow [-1,1]$

$h(x) := 0$; $S := \emptyset$

1. If there is f in F such that

$$\mathbf{E} f(x)g(x) - \mathbf{E} f(x)h(x) > \varepsilon$$

Then

– $S := S \cup \{ f \}$

– $h(x) := \begin{cases} \frac{\sum_{f \in S} f(x)}{B} & \text{if } -B \leq \sum_{f \in S} f(x) \leq B \\ -1 & \text{if } \sum_{f \in S} f(x) < -B \\ 1 & \text{if } \sum_{f \in S} f(x) > B \end{cases}$

– Go to 1

Analysis

- Call f_t distinguishing function at step t (if any)
- h_t approximating function at beginning of step t

1. For every x , after T steps

$$\sum_t f_t(x) * (g(x) - h_t(x)) \leq 4B + T/2B$$

[non-trivial part of analysis]

2. For every t ,

$$\mathbf{E}_x f_t(x) * (g(x) - h_t(x)) \geq \varepsilon$$

Observe: $4B + T/2B \geq \varepsilon T$

Set $B := 1/\varepsilon$,

then $T \leq 8/\varepsilon^2$

Impagliazzo Hard-Core Set Lemma

$g: X \rightarrow \{0,1\}$ computational problem

F family of algorithms, $f: X \rightarrow \{0,1\}$

- If
 - $\mathbf{P}_x[f(x) = g(x)] < 1 - \delta$ for all f in F^*
- Then there is set S of size $> 2 \delta |X|$ such that
 - $\mathbf{P}_{x \sim S} [f(x) = g(x)] < \frac{1}{2} + \epsilon$ for all f in F

$F^* :=$ functions of “complexity” $(\epsilon\delta)^{-O(1)}$ relative to F

Impagliazzo Hard-Core Set Lemma

$g: X \rightarrow \{-1, 1\}$ computational problem

F family of algorithms, $f: X \rightarrow \{-1, 1\}$

- If
 - $\sum_x |f(x) - g(x)| > \delta |X|$ for all f in F^*
- Then there is $s: X \rightarrow [0, 1]$, $\sum_x s(x) > \delta |X|$ such that
 - $\mathbf{E}_x [s(x)f(x)g(x)] < \varepsilon$ for all f in F

$F^* :=$ functions of “complexity” $(\varepsilon\delta)^{-O(1)}$ relative to F

Deriving H.C.L. from Regularity

$g: X \rightarrow \{-1, 1\}$ computational problem

F family of algorithms, $f: X \rightarrow \{-1, 1\}$ s.t.

$$\sum_x |f(x) - g(x)| > 2\delta |X| \quad \text{for all } f \text{ in } F^*$$

- Find $h: X \rightarrow \{-1, 1\}$, h in F^* such that
 $|E f(x)g(x) - E f(x)h(x)| \leq \varepsilon$ for all f in F
- Define $s := |h(x) - g(x)| / 2$
- Then
 - $\sum_x s(x) > \delta |X|$
 - $E_x [s(x)f(x)g(x)] =$
 $\frac{1}{2} E_x [|h(x) - g(x)| f(x)g(x)] =$
 $\frac{1}{2} E_x [h(x)f(x) - g(x)f(x)] < \varepsilon/2$

Returning to the big picture

- Additive combinatorics
 - Graph theory
 - regularity lemma
 - Analysis
 - Gowers Uniformity
 - Ergodic Theory
 - ???
 - Finitary Ergodic Theory
 - ??
 - Computer Science
 - Graph property testing
 - Pseudorandomness
 - Direct Product Thms
 - PCP
 - Average-case complexity, Pseudorandomness
-
- The diagram consists of blue arrows connecting specific items in the 'Additive combinatorics' list to items in the 'Computer Science' list. An arrow points from 'regularity lemma' to 'Graph property testing'. Three arrows originate from 'Gowers Uniformity', pointing to 'Pseudorandomness', 'Direct Product Thms', and 'PCP'. A double-headed arrow connects 'Finitary Ergodic Theory' and 'Average-case complexity, Pseudorandomness'.