Improved Convergence for ℓ_∞ and ℓ_1 Regression via Iteratively Reweighted Least Squares

Alina Ene, Adrian Vladu

BOSTON UNIVERSITY

IRLS Method

Basic primitive:

$$\min \sum r_i x_i^2$$

$$Ax = b$$

IRLS Method

Basic primitive:

$$\min \sum r_i x_i^2$$
$$Ax = b$$

solution given by one linear system solve

$$x = R^{-1}A^T(A^TR^{-1}A)^{-1}Ab$$

* R = diag(r)

IRLS Method

Basic primitive:

$$\min \sum r_i x_i^2$$
$$Ax = b$$

solution given by one linear system solve

$$x = R^{-1}A^T(A^TR^{-1}A)^{-1}Ab \quad *$$

"Hard" problem:

$$\min |x|_p \quad **$$
$$Ax = b$$

* R = diag(r)

** p = {1, ∞}

IRLS Method

Basic primitive:

$$\min \sum r_i x_i^2$$
$$Ax = b$$

solution given by one linear system solve

$$x = R^{-1}A^T(A^TR^{-1}A)^{-1}Ab$$ *

"Hard" problem:

$$\min |x|_p$$ **
$$Ax = b$$

equivalent to linear programming

* R = diag(r)

** p = {1, ∞}

IRLS Method

Basic primitive:

$$\min \sum r_i x_i^2$$
$$Ax = b$$

solution given by **one** linear system solve

$$x = R^{-1}A^T(A^TR^{-1}A)^{-1}Ab \quad *$$

"Hard" problem:

$$\min \ |x|_p \quad **$$
$$Ax = b$$

equivalent to linear programming

* $R = \text{diag}(r)$

** $p = \{1, \infty\}$

IRLS Method

Basic primitive:

$$\min \sum r_i x_i^2$$
$$Ax = b$$

solution given by one linear system solve

$$x = R^{-1}A^T(A^TR^{-1}A)^{-1}Ab \quad *$$

"Hard" problem:

$$\min\ |x|_p \quad **$$
$$Ax = b$$

equivalent to linear programming

* $R = \mathrm{diag}(r)$

** $p = \{1, \infty\}$

IRLS Method

Basic primitive:

$$\min \sum r_i x_i^2$$
$$Ax = b$$

solution given by one linear system solve

$$x = R^{-1}A^T(A^TR^{-1}A)^{-1}Ab \quad ^{*}$$

"Hard" problem:

$$\min \; |x|_p \quad ^{**}$$
$$Ax = b$$

equivalent to linear programming

* $R = \text{diag}(r)$

** $p = \{1, \infty\}$

IRLS Method

Basic primitive:

$$\min \sum r_i x_i^2$$
$$Ax = b$$

solution given by **one** linear system solve

$$x = R^{-1}A^T(A^TR^{-1}A)^{-1}Ab$$ *

"Hard" problem:

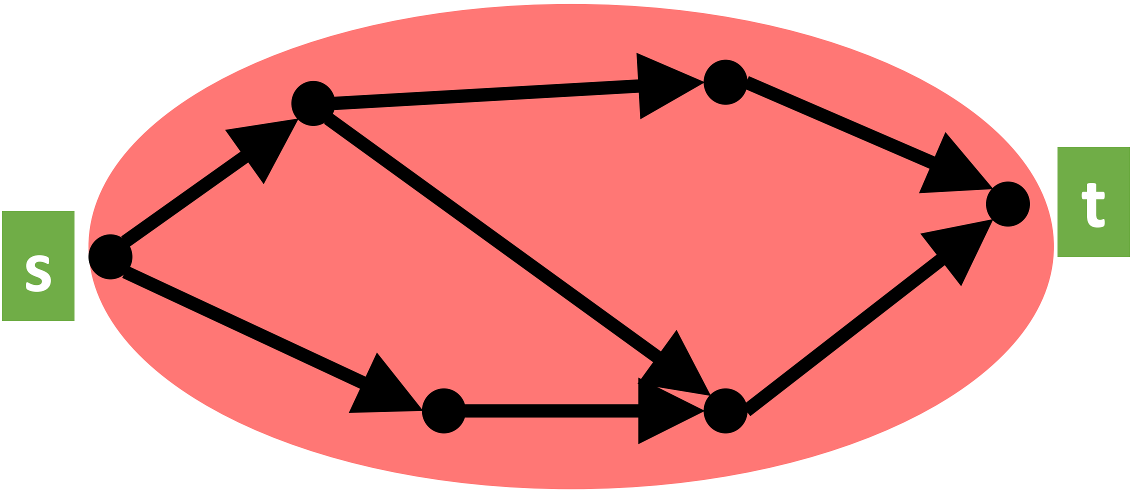
$$\min |x|_p$$ **
$$Ax = b$$

equivalent to linear programming

* R = diag(r)

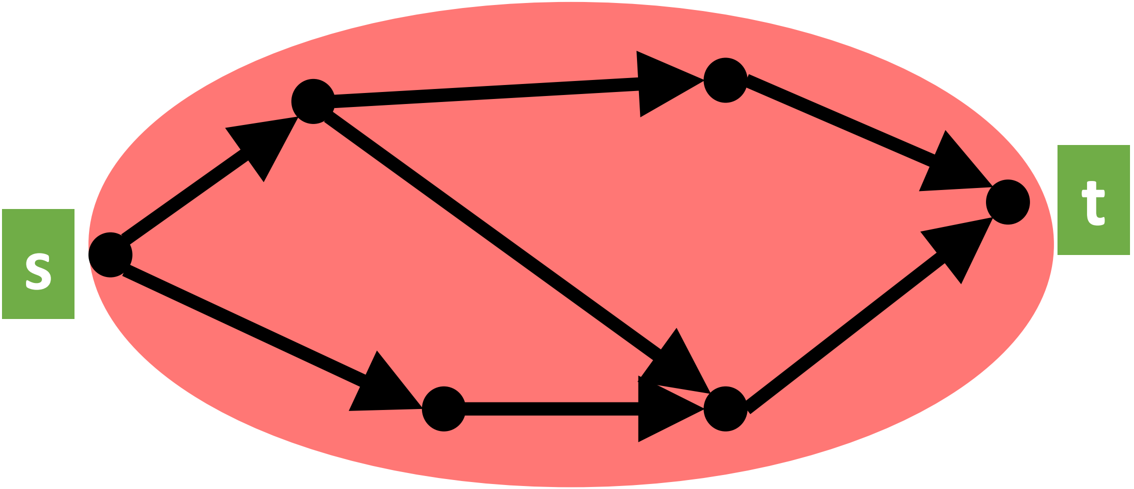
** p = {1, ∞}

Benchmark: Optimization on Graphs



$$\min |\mathbf{x}|_\infty$$
$$A\mathbf{x} = b$$

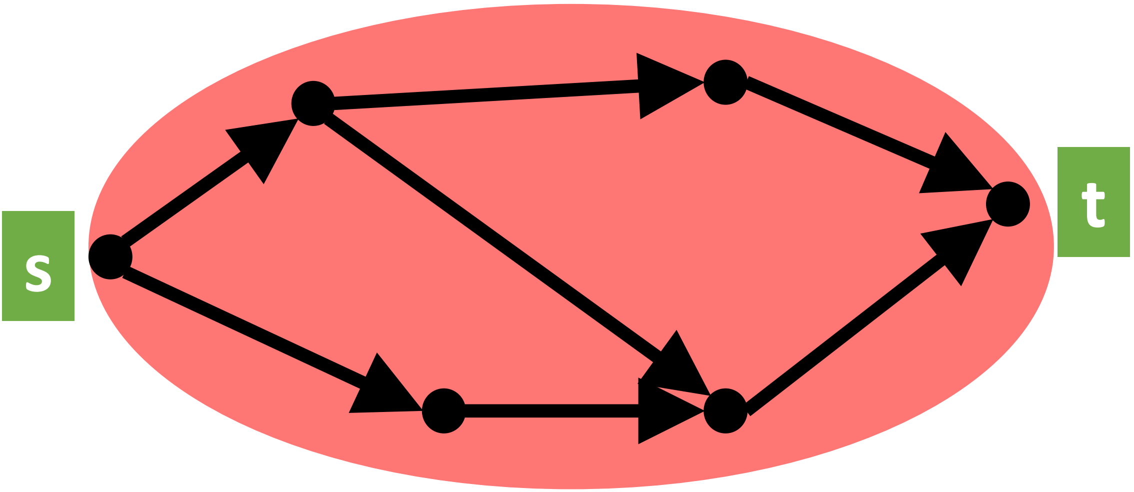
Benchmark: Optimization on Graphs



minimize congestion of flow **x**

$$\min \|x\|_\infty$$
$$Ax = b$$

Benchmark: Optimization on Graphs

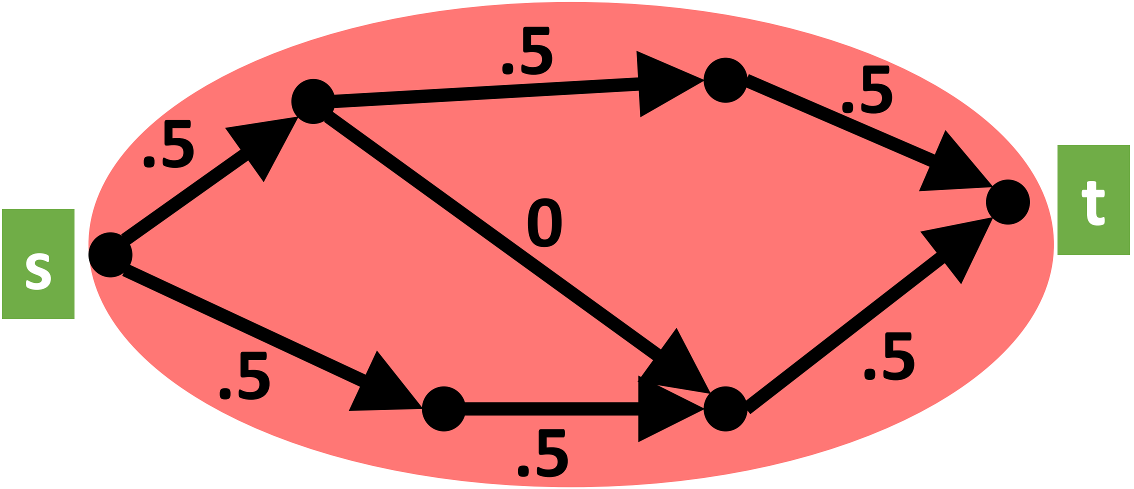


minimize congestion of flow **x**

$$\min \|x\|_\infty$$
$$Ax = b$$

boundary condition: **x** routes demand from s to t

Benchmark: Optimization on Graphs



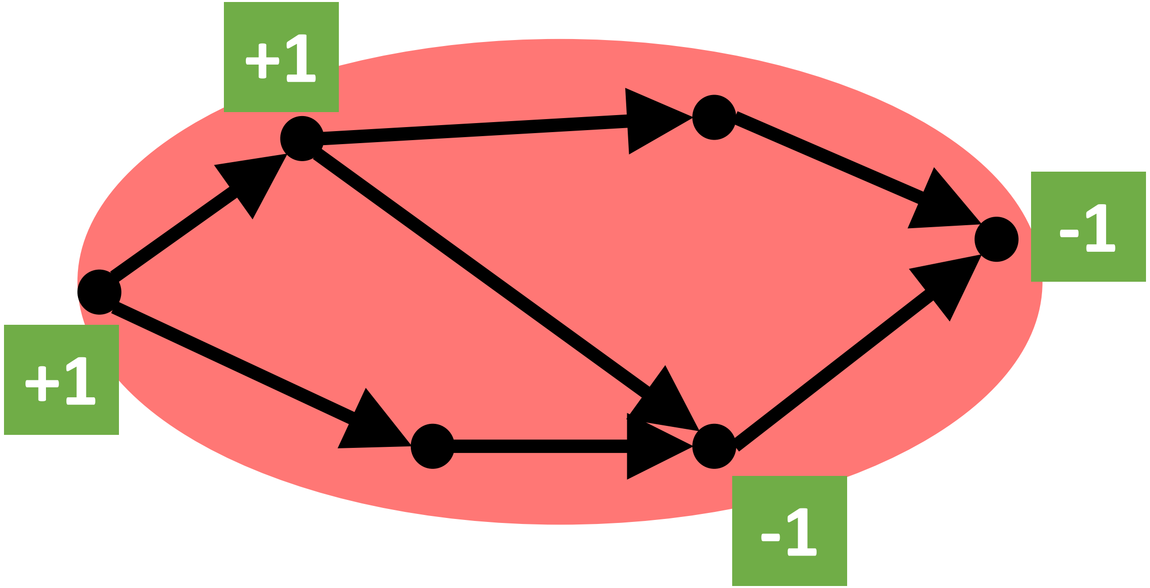
Maximum flow

minimize congestion of flow x

$$\min |x|_\infty$$
$$Ax = b$$

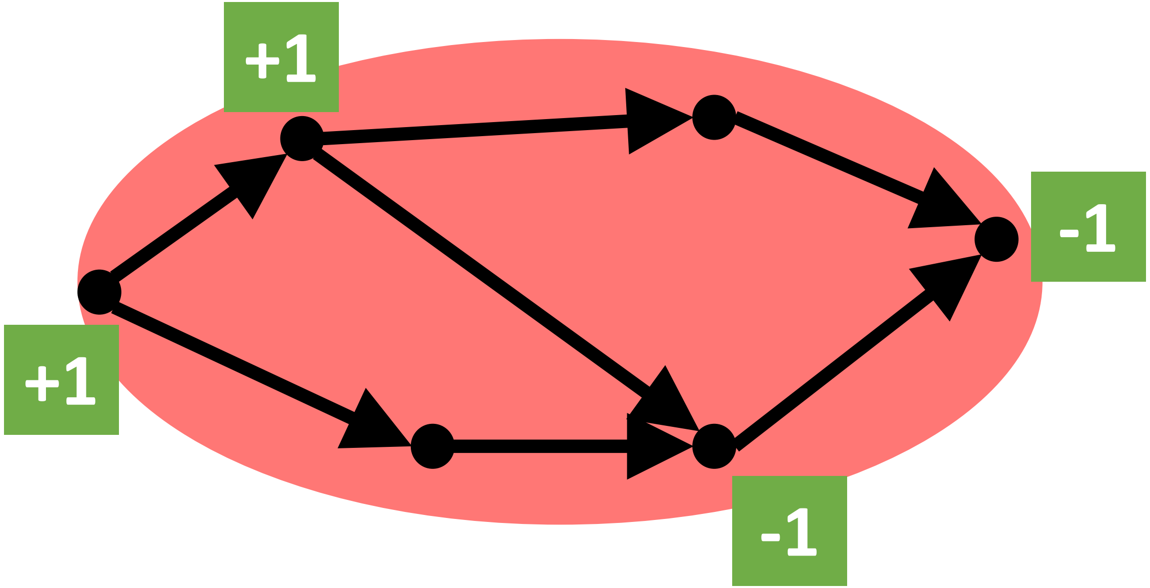
boundary condition: x routes demand from s to t

Benchmark: Optimization on Graphs



$$\min \ |\mathbf{x}|_1$$
$$A\mathbf{x} = \mathbf{b}$$

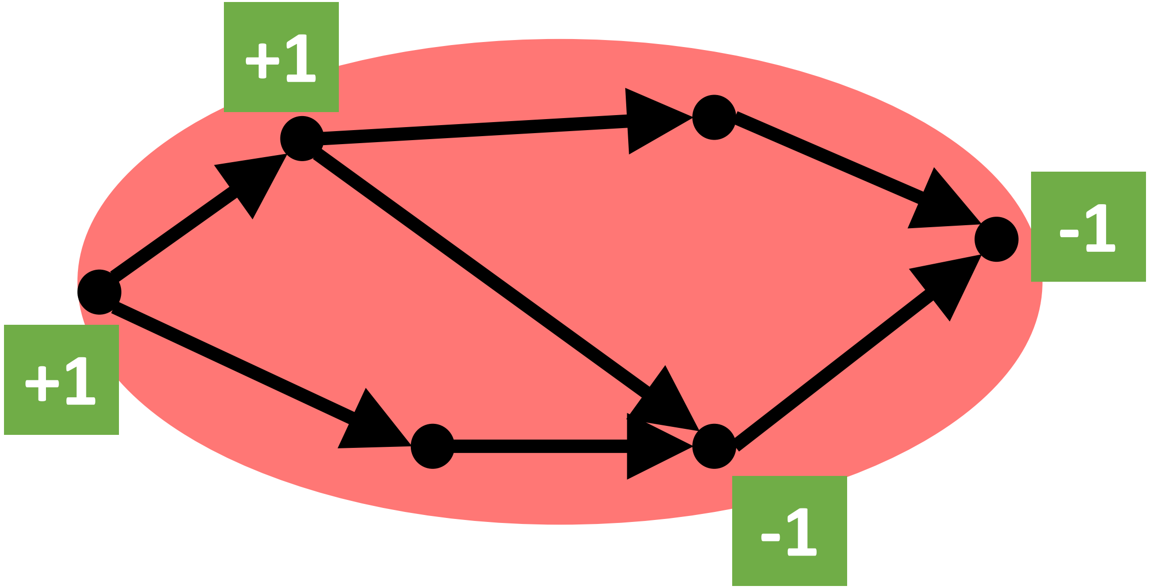
Benchmark: Optimization on Graphs



minimize
cost of
flow **x**

$$\min \|\mathbf{x}\|_1$$
$$A\mathbf{x} = b$$

Benchmark: Optimization on Graphs

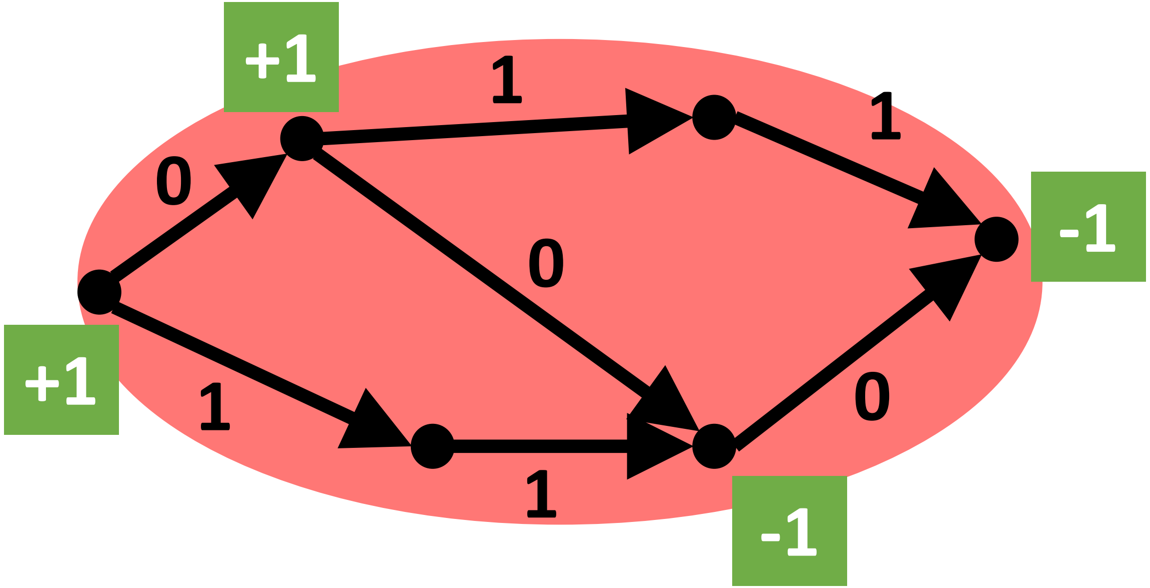


minimize
cost of
flow $\mathbf{x}$

$$\min \, |\mathbf{x}|_1$$
$$A\mathbf{x} = \mathbf{b}$$

boundary condition:
$\mathbf{x}$ routes demand
from +1 to -1

Benchmark: Optimization on Graphs



minimize
cost of
flow x

$$\min \|x\|_1$$
$$Ax = b$$

boundary condition:
x routes demand
from +1 to -1

Minimum cost flow

Benchmark: Optimization on Graphs

$$\min |x|_{\infty}$$
$$Ax = b$$

max flow

$$\min |x|_1$$
$$Ax = b$$

min cost flow

Benchmark: Optimization on Graphs

$$\min |x|_\infty$$
$$Ax = b$$

max flow

Q: ***Are these problems really that hard?***

$$\min |x|_1$$
$$Ax = b$$

min cost flow

Benchmark: Optimization on Graphs

min $|x|_\infty$
$Ax = b$

max flow

Q: *Are these problems really that hard?*

min $|x|_1$
$Ax = b$

min cost flow

First order methods (gradient descent)

➜ running time strongly depends on matrix structure

➜ in general, takes time at least **$\Omega(m^{1.5}/\mathrm{poly}(\varepsilon))$**

Benchmark: Optimization on Graphs

$\min |\mathbf{x}|_\infty$

$A\mathbf{x} = b$

max flow

Q: *Are these problems really that hard?*

$\min |\mathbf{x}|_1$

$A\mathbf{x} = b$

min cost flow

First order methods (gradient descent)

➜ running time strongly depends on matrix structure

➜ in general, takes time at least $\Omega(m^{1.5}/\text{poly}(\varepsilon))$

Second order methods (Newton method, IRLS)

➜ interior point method: $\tilde{O}(m^{1/2})$ linear system solves

➜ can be made $\tilde{O}(n^{1/2})$ with a lot of work **[Lee-Sidford ’14]**

Benchmark: Optimization on Graphs

$\min \|x\|_\infty$

$Ax = b$

max flow

Q: *Are these problems really that hard?*

$\min \|x\|_1$

$Ax = b$

min cost flow

First order methods (gradient descent)

➜ running time strongly depends on matrix structure

➜ in general, takes time at least $\Omega(m^{1.5}/\text{poly}(\varepsilon))$

Second order methods (Newton method, IRLS)

➜ interior point method: $\tilde{O}(m^{1/2})$ linear system solves

➜ can be made $\tilde{O}(n^{1/2})$ with a lot of work [Lee-Sidford '14]

"Hybrid" method

➜ [Christiano-Kelner-Madry-Spielman-Teng '11] $\tilde{O}(m^{1/3}/\varepsilon^{11/3})$ linear system solves

➜ ~30 pages of description and proofs for complicated method

This work

Natural IRLS method runs in $\tilde{O}(m^{1/3}/\varepsilon^{2/3}+1/\varepsilon^2)$ iterations

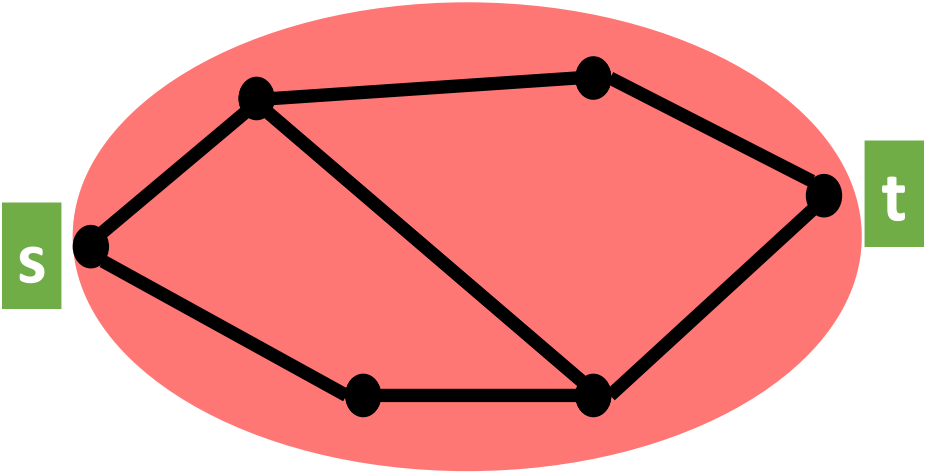
This work

Natural IRLS method runs in $\tilde{O}(m^{1/3}/\varepsilon^{2/3}+1/\varepsilon^2)$ iterations

*** no matter what the structure of the underlying matrix is**

This work

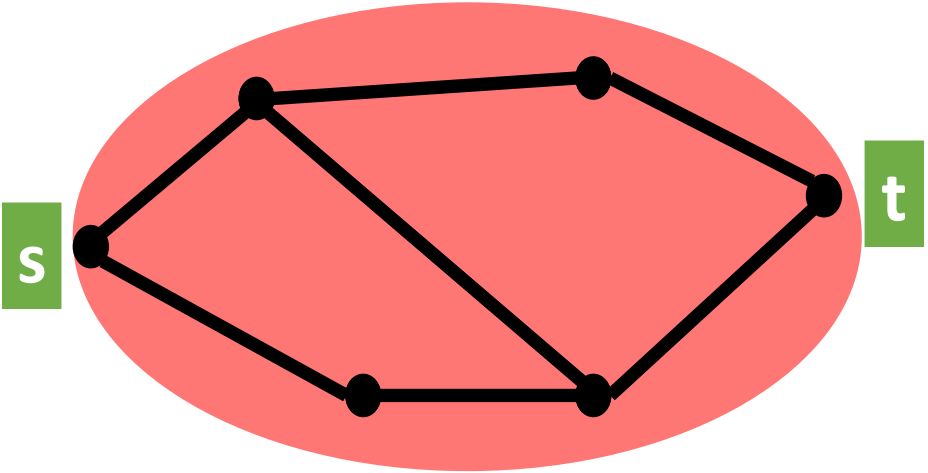
Natural IRLS method runs in $\tilde{O}(m^{1/3}/\varepsilon^{2/3}+1/\varepsilon^2)$ iterations



$$\min \|x\|_\infty \quad Ax = b \quad \leq OPT$$

This work

Natural IRLS method runs in $\tilde{O}(m^{1/3}/\varepsilon^{2/3}+1/\varepsilon^{2})$ iterations



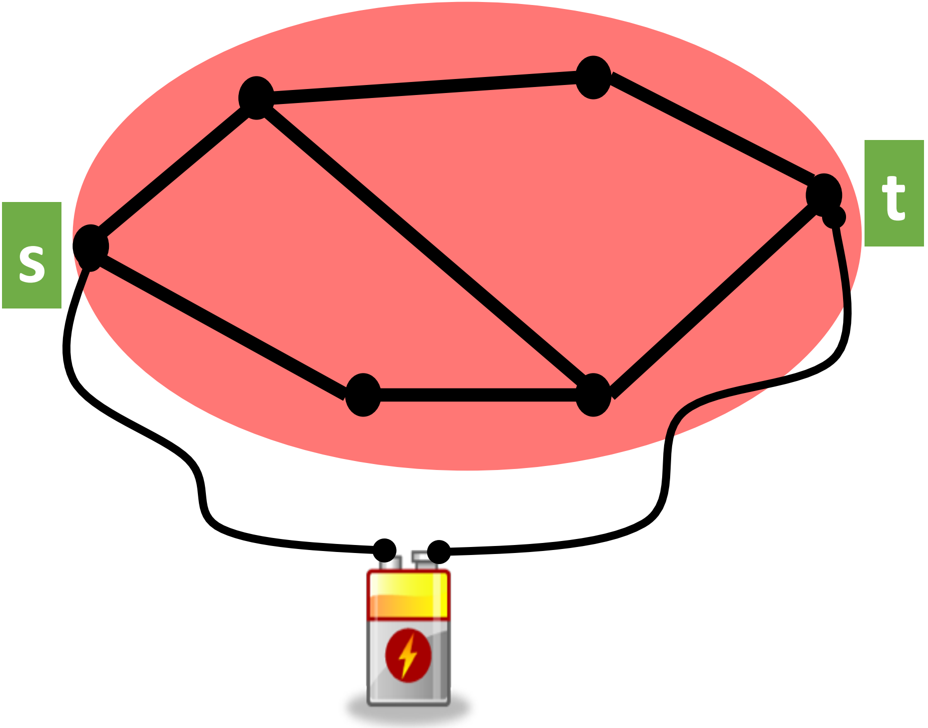
Guess OPT value

$$\min \ |x|_\infty \quad Ax = b$$

$\leq$ OPT (.5)

This work

Natural IRLS method runs in $\tilde{O}(m^{1/3}/\varepsilon^{2/3}+1/\varepsilon^2)$ iterations



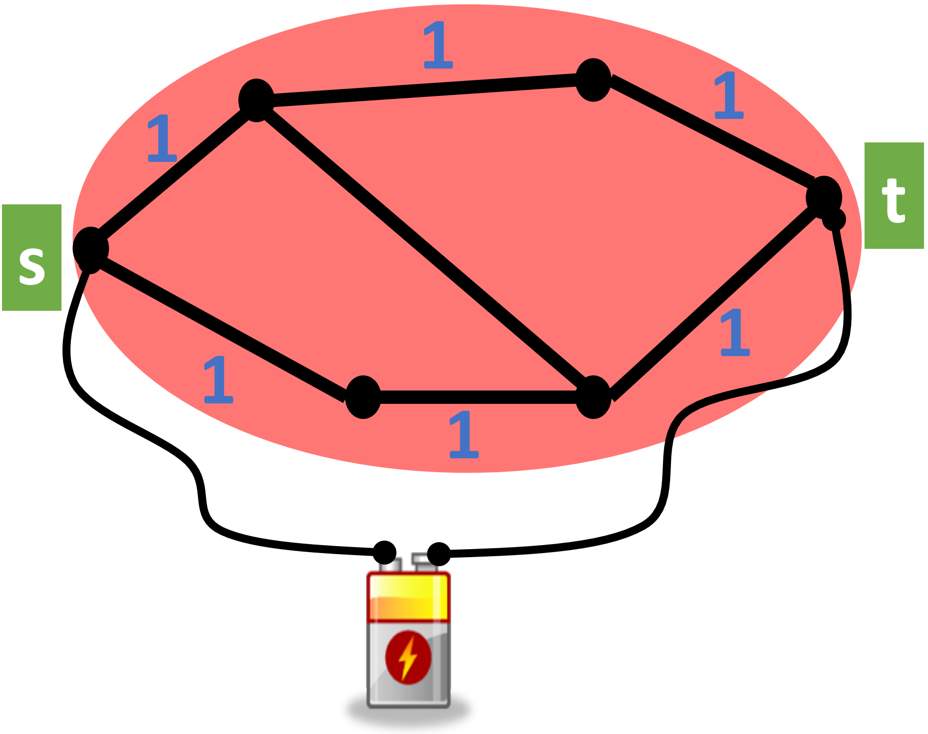
Guess OPT value

$$\min \|x\|_\infty \quad Ax = b$$

$\leq$ OPT (.5)

This work

Natural IRLS method runs in $\tilde{O}(m^{1/3}/\varepsilon^{2/3}+1/\varepsilon^2)$ iterations



Guess OPT value

$$\min \|x\|_\infty \quad Ax = b$$

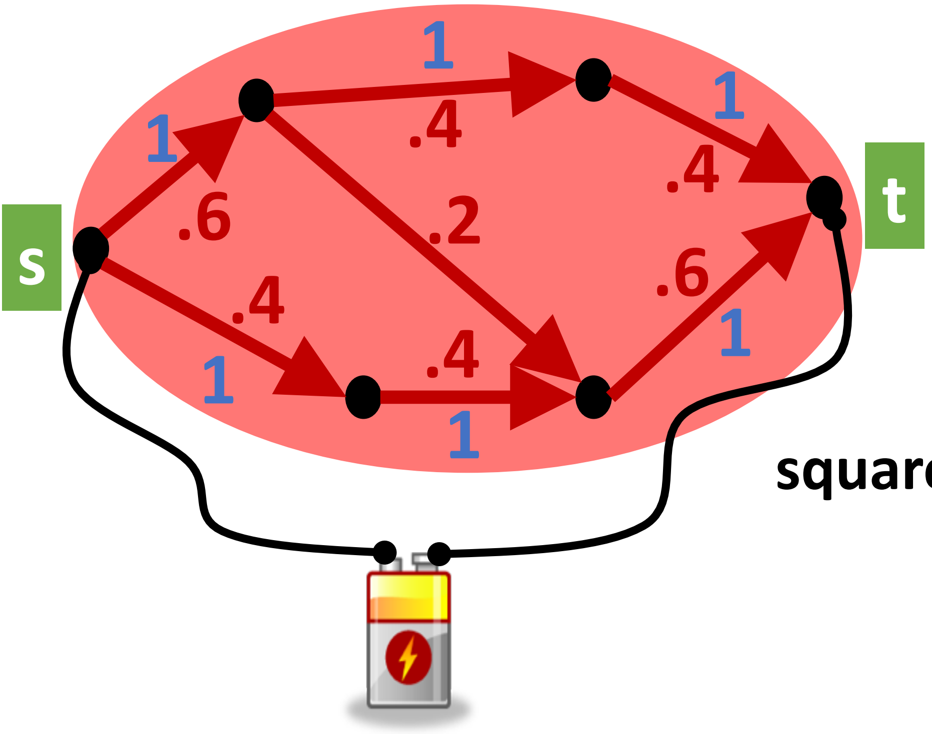
$\leq$ OPT (.5)

Initialize

$r = 1$

This work

Natural IRLS method runs in $\tilde{O}(m^{1/3}/\varepsilon^{2/3}+1/\varepsilon^{2})$ iterations



Guess OPT value

$$\min \ |x|_\infty \quad Ax = b$$

$\leq$ OPT (.5)

Initialize

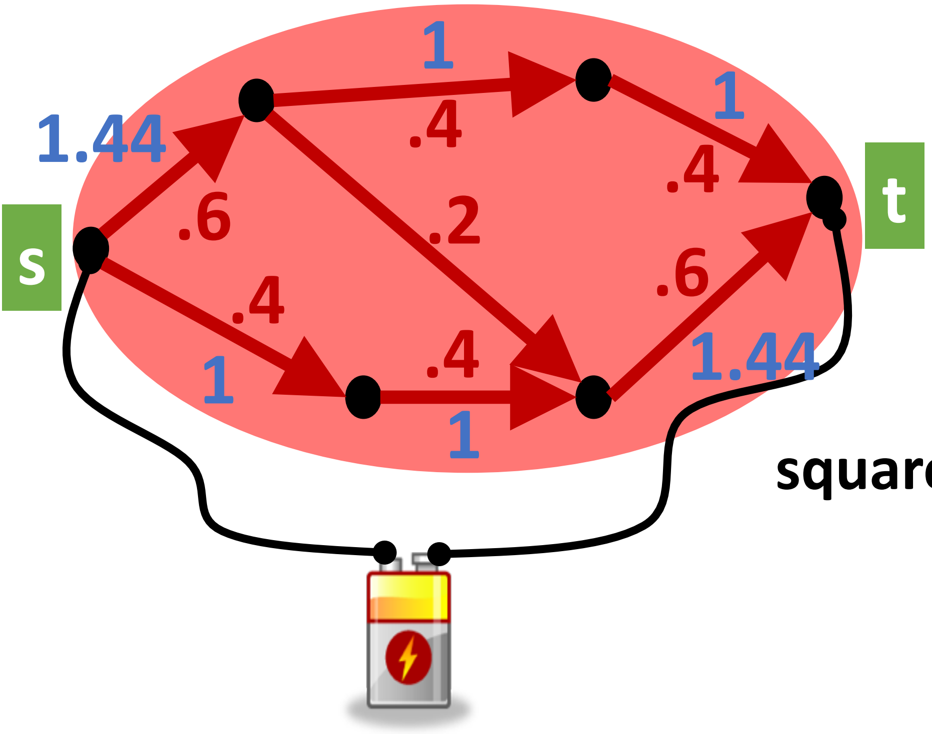
$$r = 1$$

Solve least squares problem

$$\min \sum r_i x_i^2 \quad Ax = b$$

This work

Natural IRLS method runs in $\tilde{O}(m^{1/3}/\varepsilon^{2/3}+1/\varepsilon^2)$ iterations



Guess OPT value: $\min |x|_\infty$, $Ax = b$ $\leq$ OPT (.5)

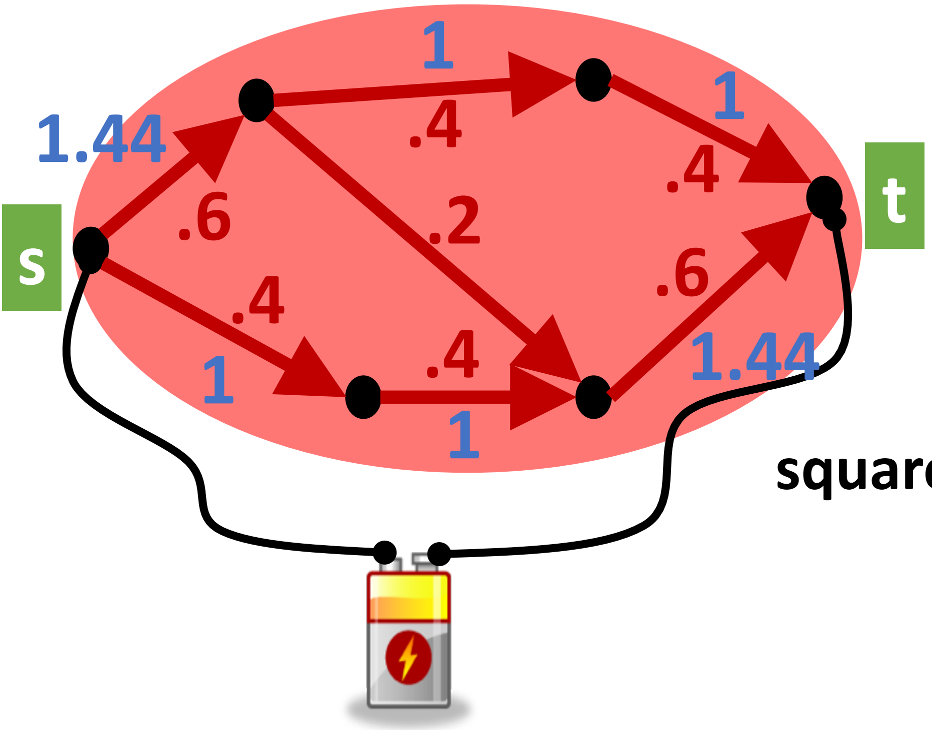
Initialize: $r = 1$

Solve least squares problem: $\min \sum r_i x_i^2$, $Ax = b$

Update r: $r_i \leftarrow r_i * \max\{(x_i/OPT)^2, 1\}$

This work

Natural IRLS method runs in $\tilde{O}(m^{1/3}/\varepsilon^{2/3}+1/\varepsilon^2)$ iterations



Guess OPT value: $\min |x|_\infty$, $Ax = b$ $\leq$ OPT (.5)

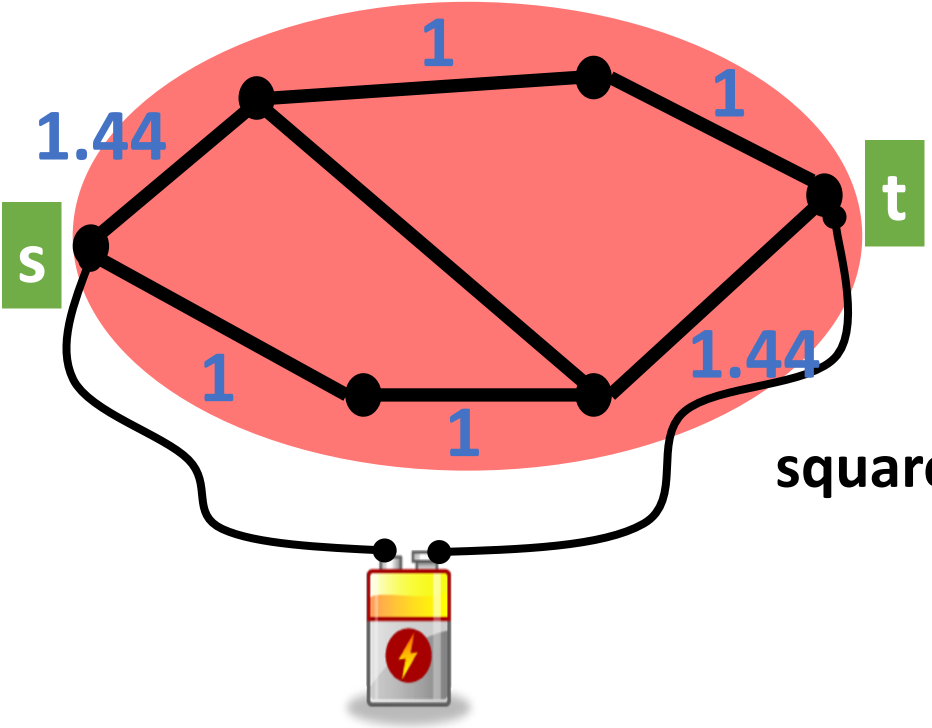
Initialize: $r = 1$

Solve least squares problem: $\min \sum r_i x_i^2$, $Ax = b$

Update r: $r_i \leftarrow r_i * \max\{(x_i/\mathrm{OPT})^2, 1\}$

This work

Natural IRLS method runs in $\tilde{O}(m^{1/3}/\varepsilon^{2/3}+1/\varepsilon^2)$ iterations



Guess OPT value

$$\min |x|_\infty \quad Ax = b$$

$\leq$ OPT (.5)

Initialize

$$r = 1$$

Solve least squares problem

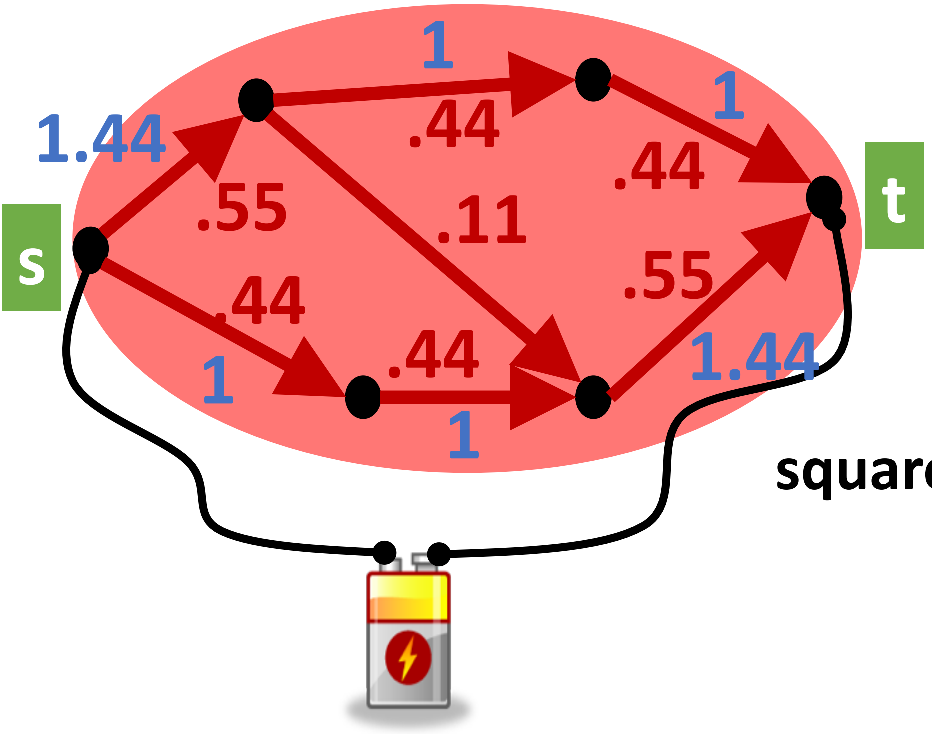
$$\min \sum r_i x_i^2 \quad Ax = b$$

Update r

$$r_i \leftarrow r_i * \max\{(x_i/OPT)^2, 1\}$$

This work

Natural IRLS method runs in $\tilde{O}(m^{1/3}/\varepsilon^{2/3}+1/\varepsilon^2)$ iterations



Guess OPT value

$$\min \|x\|_\infty \quad Ax = b$$

$\leq$ OPT (.5)

Initialize

$$r = 1$$

Solve least squares problem

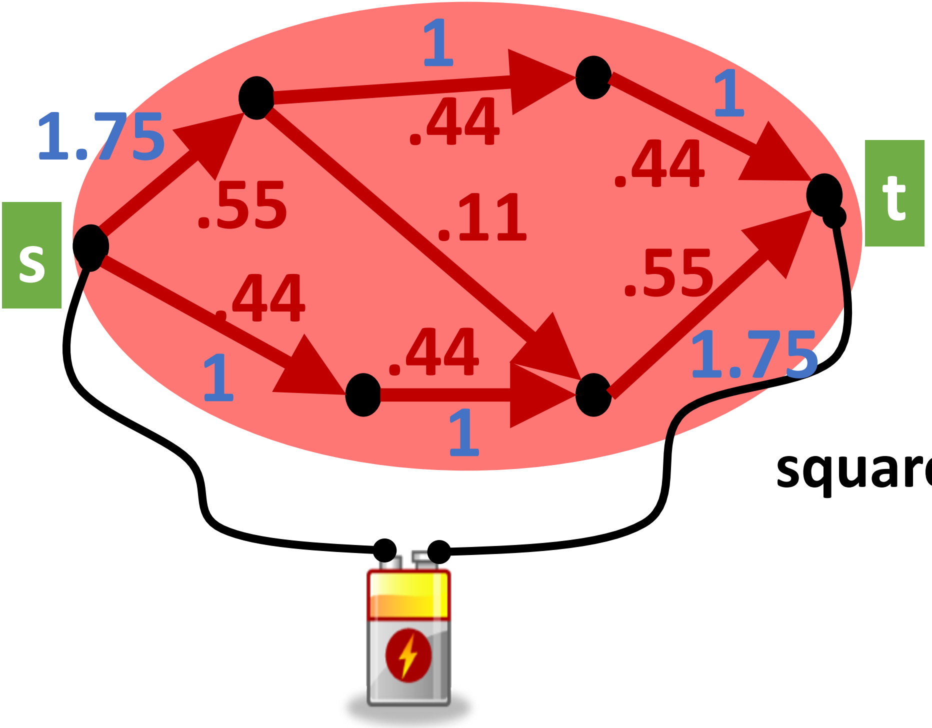
$$\min \sum r_i x_i^2 \quad Ax = b$$

Update r

$$r_i \leftarrow r_i * \max\{(x_i/\mathrm{OPT})^2, 1\}$$

This work

Natural IRLS method runs in $\tilde{O}(m^{1/3}/\varepsilon^{2/3}+1/\varepsilon^2)$ iterations



Guess OPT value

$$\min \ |x|_\infty$$
$$Ax = b$$

$\leq$ OPT (.5)

Initialize

$$r = 1$$

Solve least squares problem

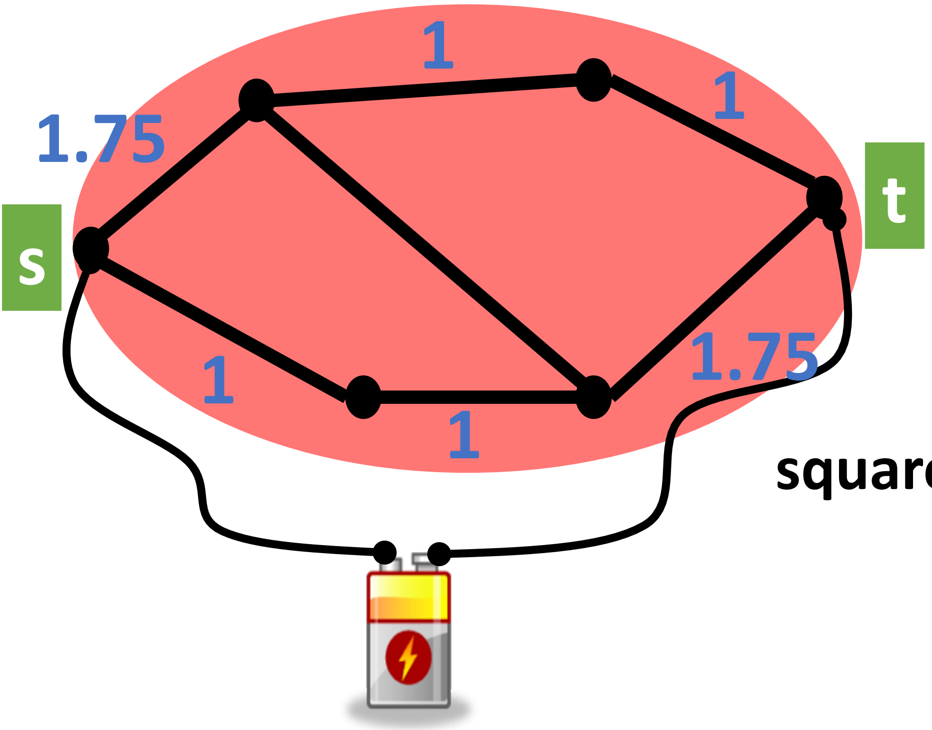
$$\min \sum r_i x_i^2$$
$$Ax = b$$

Update r

$$r_i \leftarrow r_i * \max\{(x_i/\mathrm{OPT})^2, 1\}$$

This work

Natural IRLS method runs in $\tilde{O}(m^{1/3}/\varepsilon^{2/3}+1/\varepsilon^2)$ iterations



Guess OPT value: $\min \|x\|_\infty$, $Ax = b$ $\leq$ OPT (.5)

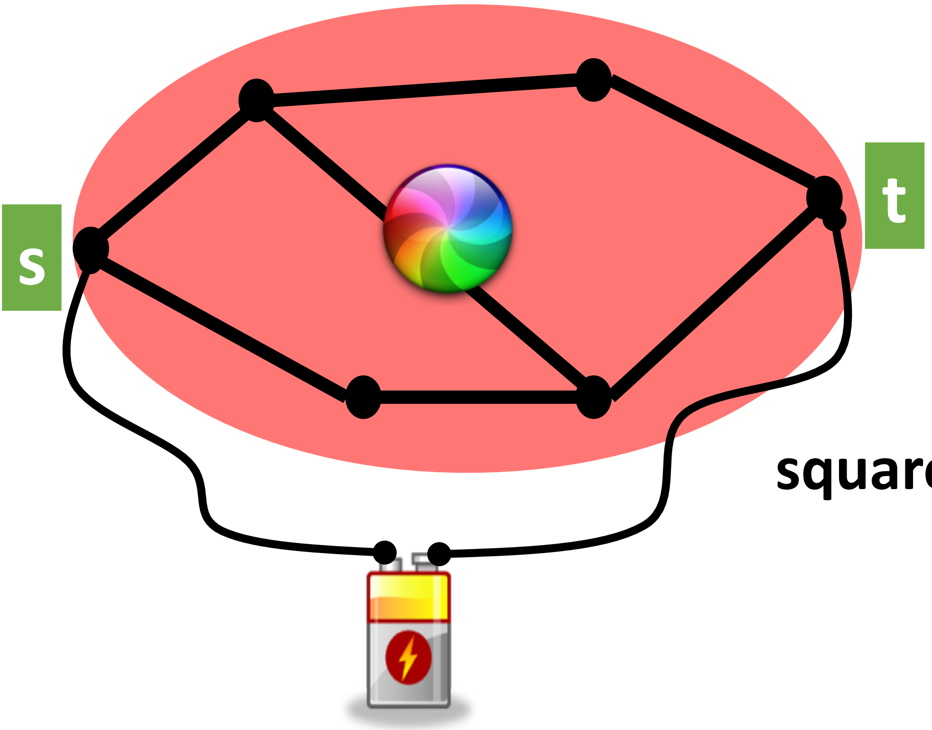
Initialize: $r = 1$

Solve least squares problem: $\min \sum r_i x_i^2$, $Ax = b$

Update r: $r_i \leftarrow r_i * \max\{(x_i/\mathrm{OPT})^2, 1\}$

This work

Natural IRLS method runs in $\tilde{O}(m^{1/3}/\varepsilon^{2/3}+1/\varepsilon^2)$ iterations



Guess OPT value: $\min |x|_\infty$, $Ax = b$ $\leq$ OPT (.5)

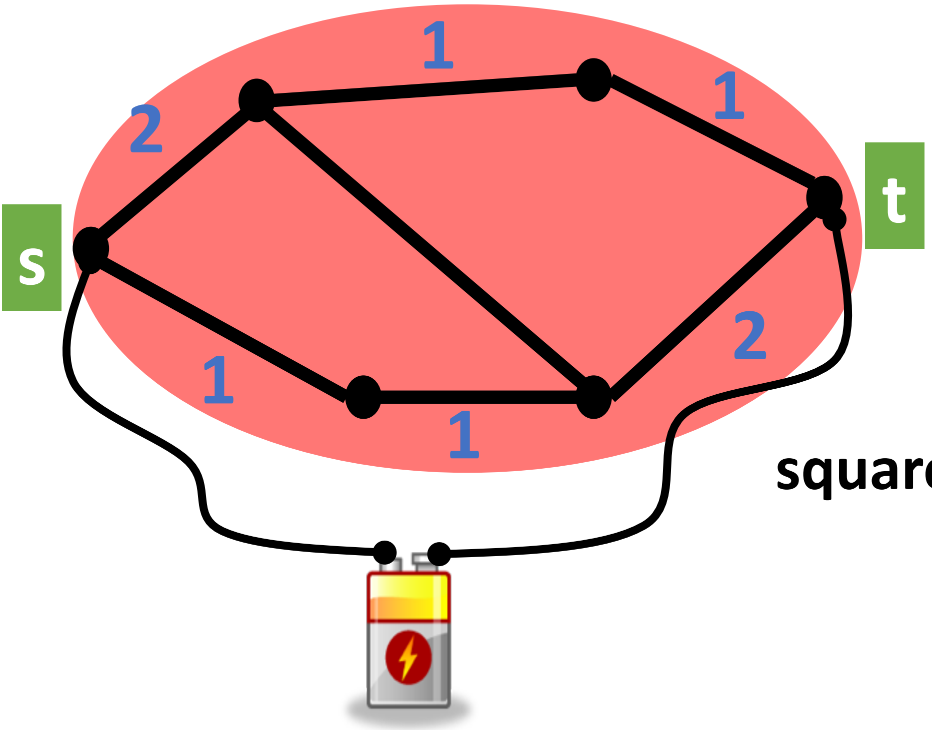
Initialize: $r = 1$

Solve least squares problem: $\min \sum r_i x_i^2$, $Ax = b$

Update r: $r_i \leftarrow r_i * \max\{(x_i/OPT)^2, 1\}$

This work

Natural IRLS method runs in $\tilde{O}(m^{1/3}/\varepsilon^{2/3}+1/\varepsilon^2)$ iterations



Guess OPT value: $\min |x|_\infty$, $Ax = b$ $\leq$ OPT (.5)

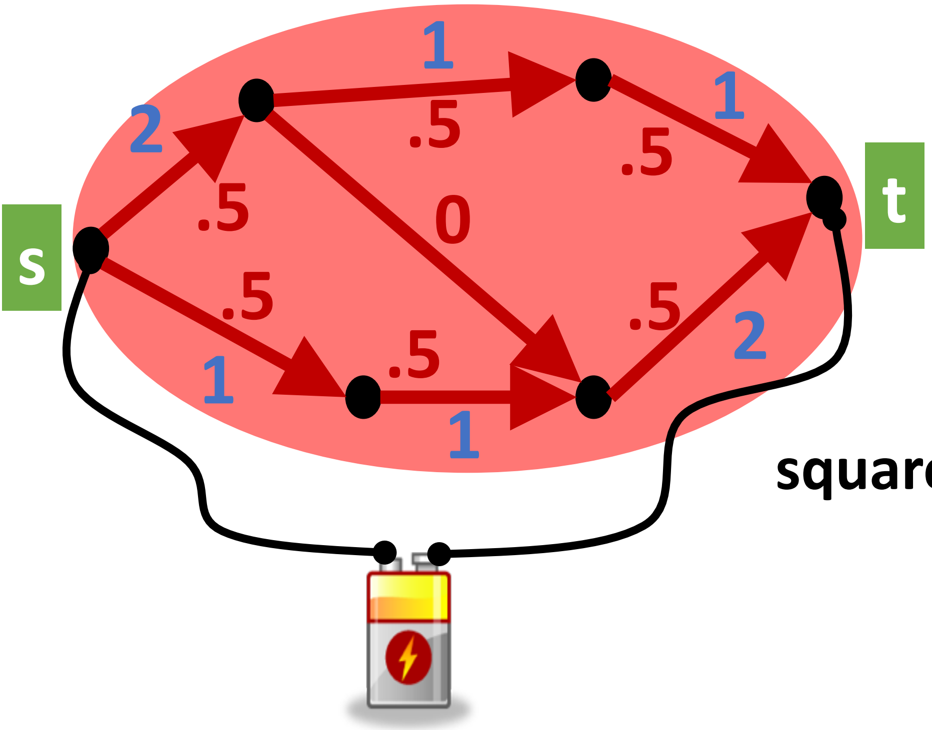
Initialize: $r = 1$

Solve least squares problem: $\min \sum r_i x_i^2$, $Ax = b$

Update r: $r_i \leftarrow r_i * \max\{(x_i/\text{OPT})^2, 1\}$

This work

Natural IRLS method runs in $\tilde{O}(m^{1/3}/\varepsilon^{2/3}+1/\varepsilon^2)$ iterations



Guess OPT value

$$\min \|x\|_\infty \quad Ax = b$$

$\leq$ OPT (.5)

Initialize

$$r = 1$$

Solve least squares problem

$$\min \sum r_i x_i^2 \quad Ax = b$$

Update r

$$r_i \leftarrow r_i * \max\{(x_i/\mathrm{OPT})^2, 1\}$$

Nonstandard Optimization Primitive

➜ Objective function is $\mathbf{max_{r\geq 0}\ min_{Ax=b}\ \sum r_i x_i^2/\sum r_i}$

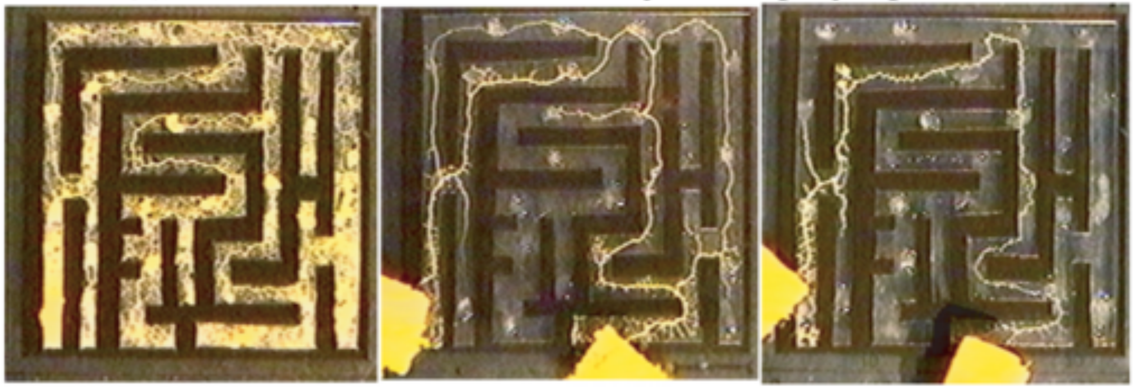
Nonstandard Optimization Primitive

➜ Objective function is $\mathbf{max}_{r \geq 0}\ \mathbf{min}_{Ax=b}\ \sum r_i x_i^2 / \sum r_i$

➜ Similar analysis to packing/covering LP **[Young '01]**

Nonstandard Optimization Primitive

- ➜ Objective function is $\mathbf{max}_{r\geq 0}\ \mathbf{min}_{Ax=b}\ \sum r_i x_i^2 / \sum r_i$
- ➜ Similar analysis to packing/covering LP **[Young '01]**
- ➜ ℓ_1 version is a type of "slime mold dynamics" **[Straszak-Vishnoi '16, '17]**



Nonstandard Optimization Primitive

- ➜ Objective function is $\mathbf{max}_{r\geq 0}\ \mathbf{min}_{Ax=b}\ \sum r_i x_i^2/\sum r_i$
- ➜ Similar analysis to packing/covering LP **[Young '01]**
- ➜ ℓ_1 version is a type of "slime mold dynamics" **[Straszak-Vishnoi '16, '17]**

Nonstandard Optimization Primitive

- ➜ Objective function is $\mathbf{max}_{r\geq 0}\ \mathbf{min}_{Ax=b}\ \sum r_i x_i^2 / \sum r_i$
- ➜ Similar analysis to packing/covering LP **[Young '01]**
- ➜ ℓ_1 version is a type of "slime mold dynamics" **[Straszak-Vishnoi '16, '17]**
- ➜ Any insights for new optimization methods?

Thank You!

More details at poster #208