Local Flow Partitioning for Faster Edge Connectivity

Monika Henzinger, Satish Rao, Di Wang

University of Vienna, UC Berkeley, UC Berkeley
Edge Connectivity

- **Edge-connectivity $\lambda$:** least number of edges whose removal disconnects the graph.
- **Minimum cut:** set of edges of minimum size whose removal disconnects the graph.
  - $\lambda = 2$

  $\lambda$ = size of minimum cut in **unweighted** graphs
# Prior Work

## Deterministic algorithm

|                |       | unweighted (multi-)graph |
|----------------|-------|--------------------------|
| Gabow’91       | $O(\lambda m \log n)$ |                          |
| Kawarabayashi & Thorup’15 | $O(m \log^{12} n)$ | simple graph              |
| Henzinger, Rao, W’17 | $O(m \log^{12} n \log \log^{12} n)$ | simple graph              |

## Randomized algorithm

|                |       | weighted graph |
|----------------|-------|----------------|
| Karger’00      | $O(m \log^{13} n)$ |                |

$n$ nodes, $m$ edges, min cut = $\lambda$

Simple graph: undirected, unweighted, no parallel edges

Multi-graph: can have parallel edges.
Theorem

G: simple, min degree $\delta \quad O(m)$

Multi-graph $\mathcal{G}$ with

$\min \deg \leq \lambda \leq \delta$

$m \downarrow \mathcal{G} = O(m/\delta)$ edges

Non-trivial min cut in $\mathcal{G}$

Min cut in $\mathcal{G}$

- **Trivial cut**: only 1 node on one side of the cut.
- The min degree $\delta$ bounds the edge connectivity $\lambda$

$\lambda \leq \delta$
Kawarabayashi-Thorup (KT)

- Theorem
  \[ G: \text{simple, min degree } \delta \quad \mathcal{O}(m) \quad \text{Multi-graph } \mathcal{G} \text{ with } \quad m \downarrow G = \mathcal{O}(m/\delta) \text{ edges} \]

  \[ \begin{align*}
  \text{Non-trivial min cut in } \mathcal{G} & \iff \text{Min cut in } \mathcal{G} \\
  \end{align*} \]

- Gabow’s algorithm on \( G \)
  \[ \mathcal{O}(\lambda m \downarrow G \log m) = \mathcal{O}(\lambda m / \delta) = \mathcal{O}(m) \]

- Assume \( \delta = \Omega(\log n) \quad \lambda \leq \delta \)
Low Conductance Cut

Conductance: $\phi(A) = \frac{|E(A,A)|}{\min\{vol(A), vol(A)\}}$

$vol(A) = \sum_{v \in A} \deg(v)$

Non-trivial cut of size $\leq \delta$ has low conductance!

2 nodes: $\geq 2\delta$ total degree
$\leq \delta$ edges across the cut
$\geq 2$ nodes $\Rightarrow \Omega(\delta)$ nodes
Low Conductance Cut

Conductance: \( \phi(A) = \frac{|E(A,A)|}{\min\{vol(A), vol(A')\}} \)

\( vol(A) = \sum_{v \in A} \deg(v) \)

Non-trivial cut of size \( \leq \delta \) has low conductance!

2 nodes: \( \geq 2\delta \) total degree
\( \leq \delta \) edges across the cut
\( \geq 2 \) nodes \( \Rightarrow \Omega(\delta) \) nodes

volume is \( \Omega(\delta^{1/2}) \)
\( \Rightarrow \) conductance \( O(1/\delta) \)
Local Graph Partitioning

Central tool in [KT’15], improved by us

Given $G$ with $m$ edges, find cut $(A,A)$

- Low conductance: $\phi(A) = O(1/\log m)$
- Local running time: $O(vol(A)\log \uparrow c m)$
  - Cannot afford $O(m)$ in recursive decomposition
PageRank/Diffusion [ACL’06]

**Input:** 1 unit of mass at a vertex v, rate of decay $\alpha$

Maintains 2 vectors in n-dimensional space:

- $p =$ “settled mass” and $r =$ “unsettled mass”
- **Initially:** $p = 0$, $r = 1$ at v and 0 everywhere else
- **Repeat:**
  - for every vertex $u$:
    - $p'(u) = p(u) + \alpha r(u)$ \text{mass settles}
    - $r'(u) = (1 - \alpha) r(u)/2$
    - For each neighbor $v$ of $u$:
      - $r'(v) = r(v) + (1 - \alpha)r(u)/(2\text{deg}(u))$ \text{mass pushed to neighbors}
  - $p = p'$, $r = r'$
PageRank/Diffusion [ACL’06]

- Input: starting distr., rate of decay $\alpha$
- Settle fraction $\alpha$ of residual mass per round
- Spread half of the remaining evenly to neighbors
- $\varepsilon$-approx. of limiting distribution in time $O(1/(\alpha \varepsilon))$
PageRank/Diffusion [ACL’06]

**Input:** starting distr., rate of decay $\alpha$

- Typical local partitioning result:

  $\exists$ conductance $O(\phi^{12}/\log m)$ cut

  Find conductance $\phi$ cut$^A$ in time $O(\text{vol}(A)/\phi^{12})$

- Quadratic loss in cut quality and running time
PageRank/Diffusion [ACL’06]

**Input:** starting distr., rate of decay $\alpha$

- Settle fraction $\alpha$ of residual mass per round
- Spread half of the remaining evenly to neighbors
- $\varepsilon$-approx. of limiting distribution in time $O(1/(\alpha\varepsilon))$
- Typical local partitioning result:
  $$\exists \text{ conductance } O(\phi^2 / \log m) \text{ cut}$$

Find conductance $\phi$ cut $A$ in time $O(\text{vol}(A)/\phi^2)$

- Quadratic loss in cut quality and running time
Flow-based Method

- Polylog loss in cut quality
- Difficult to make the running time local
Flow-based Method

- Polylog loss in cut quality
- Difficult to make the running time local

**Two-level structure** [HRW’17]

- *Unit-Flow*
  - Try to find **low conductance cut**
  - Running time "global" ~ size of instance

- *Excess Scaling*
  - Get running time local
  - Control instance size for Unit-Flow via value of unit.
Unit-Flow\((G, \Delta, \phi)\)

Called repeatedly on “partial” flow problems

**Input:**
- Graph \(G\)
- Source supply \(\Delta\): \(\forall v \Delta(v) \leq 2\ \text{deg}(v)\) units
- Parameters: target conductance \(\phi\)

**Flow Problem:**
- Each \(v\) has sink capacity \(\text{deg}(v)\) units.
- Edge capacities \(=1/\phi\) units.

![Diagram showing flow from source to sink](image)
Unit-Flow($G, \Delta, \phi$)

Variant of preflow push-relabel

**Preflow** $f: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$

- Antisymmetry: $f(u,v) = -f(v,u)$
- Non-deficient flows: $\forall v$, $\sum_{u \leftarrow v} f(v,u) \leq \Delta(v)$
- Respects edge capacities
  
  $f(v) = \sum_{u \leftarrow v} f(u,v) + \Delta(v)$
Unit-Flow($G, \Delta, \phi$)

Variant of preflow push-relabel

**Preflow** $f: V \times V \to \mathbb{R}$
Antisymmetry: $f(u,v) = -f(v,u)$
Non-deficient flows: $\forall v, \sum u \uparrow \bar{u} f(v,u) \leq \Delta(v)$
Respects edge capacities
$f(v) = \sum u \uparrow \bar{u} f(u,v) + \Delta(v)$

**Push-relabel algorithm:**
Each vertex has a height, starting at 0.
Repeatedly pick any $v$ with excess (i.e. $f(v) > \deg(v)$)

*Push:* send excess to lower neighbor along edges with residual capacity.
*Relabel:* if not possible, raise height of $v$ by 1.
Unit-Flow($G, \Delta, \phi$)

Key adaptations

- Upper-bound height by $h = \log m / \phi$
- Flow solution not guaranteed: Might not push all flow to sources
Unit-Flow\((G, \Delta, \phi)\)

Key adaptations

- Upper-bound height by \(h = \log m / \phi\),
- Flow solution not guaranteed
- But then \(\exists\) conductance \(O(\phi)\) “level cut”

Region growing argument
\((1 + \phi)^h m\)
Unit-Flow\( (G, \Delta, \phi) \)

Key adaptations

- Upper-bound height by \( h = \log m \)
- Flow solution not guaranteed
- \( \exists \) conductance \( O(\phi) \) “level cut”

\[ (1+\phi)^h \gg m \]

- Upper-bound excess on vertex
  - Maintain \( f(v) \leq 2\deg(v) \), assumed at start
  \[ \Rightarrow \text{Total excess } \leq \text{vol}(A) \text{ at the end} \]
Unit-Flow($G, \Delta, \phi$)

\[ f(\nu) = \# \text{ units of supply on } \nu \text{ at the end} \]

- **Either** routes all source supply to sinks

\[ \forall \nu: f(\nu) \leq \deg(\nu) \]

- **Or** finds conductance $O(\phi)$ cut $(A,A)$,

and total excess bounded by $vol(A)$

\[
\text{total excess} = \sum_{\nu} \max(0, f(\nu) - \deg(\nu)) \\
\leq \text{vol}(A)
\]

- **Explored subgraph volume** $\approx \sum \Delta(\nu) = \text{total units of flow}$
Unit-Flow($G, \Delta, \phi$)

\[ f(\nu) = \# \text{ units of supply on } \nu \text{ at the end} \]

- **Either** routes all source supply to sinks
  \[ \forall \nu : f(\nu) \leq \text{deg}(\nu) \]
- **Or** finds conductance $O(\phi)$ cut $(A,A)$,
  and total excess $\leq \text{vol}(A)$

**Running time:** $O(|\Delta|/\log m / \phi)$, $|\Delta| = \sum \nu \uparrow \Delta(\nu)$, proportional to volume of explored subgraph
**Running time:** \( O(|\Delta|/\log m / \phi) \), \( |\Delta| = \sum v \Delta(v) \)

But when cut \((A, A)\) is returned we need time \( O(vol(A) / \phi) \)

**Idea:**

- Repeatedly run Unit-Flow for doubling values of \(|\Delta|\) until Unit-Flow returns a cut \((A, A)\) with excess \( \geq \Omega(|\Delta|/\log n) \)
  - \( vol(A) \geq \Omega(|\Delta|/\log n) \)

- Can bound running time of all preceding calls to Unit-Flow by \( O(vol(A) \log \Omega n/\phi) \)

- Done by Excess Scaling
Idea:

- Repeatedly run Unit-Flow for doubling values of $|\Delta|$ until Unit-Flow returns a cut $(A,A)$ with $vol(A) \geq \Omega(|\Delta|/\log n)$
- Can bound running time of all preceding calls to Unit-Flow by $O(vol(A) \log \log n / \phi)$
- If never a “large enough” cut is returned then $\sum_{j} vol(A \downarrow j)$ is “small” and the (weighted) sum of the flows returned by all the Unit-Flow routes “almost all” flow
Excess Scaling

**Input:**
- Graph G
- Source supply $\Delta$, $|\Delta| = \sum v \uparrow \Delta(v) = 2m$

**Flow problem:**
- Each $v$ sink of capacity $\text{deg}(v)$
- Sufficient edge capacity for all calls to Unit-Flow
Excess Scaling

Source supply $\Delta$, $|\Delta| = \sum \Delta(v) = 2m$

Each $v$ sink of capacity $\text{deg}(v)$

Divide into “growing” phases for Unit-Flow

- Start with large enough unit value $\mu = \max_v \Delta(v)/2 \text{deg}(v)$
  
  $\Delta \downarrow 0 = \Delta/\mu \rightarrow \Delta \downarrow 0 (v) \leq 2\text{deg}(v)$,

Problem size: $2\text{deg}(v)$
Excess Scaling

Source supply $\Delta$, $|\Delta| = \sum_{v} \Delta(v) = 2m$

Each $v$ sink of capacity $\text{deg}(v)$

Divide into “growing” phases for Unit-Flow

- Start with large enough unit value $\mu = \max_v \Delta(v)/2 \text{deg}(v)$
  
  $\Delta \downarrow 0 = \Delta/\mu \rightarrow \Delta \downarrow 0 (v) \leq 2\text{deg}(v)$,

- Either returns low conductance cut $\text{or}$

  $\forall v: f(v) \leq \text{deg}(v)$

Problem size: $2\text{deg}(v)$
Excess Scaling

Source supply $\Delta$, $|\Delta| = \sum_{v \uparrow} \Delta(v) = 2m$

Each $v$ sink of capacity $\text{deg}(v)$

Divide into “growing” phases for Unit-Flow

- Start with large enough unit value $\mu = \max_{v} \Delta(v)/2 \text{deg}(v)$
  
  $\Delta \downarrow 0 = \Delta/\mu \rightarrow \Delta \downarrow 0 (v) \leq 2\text{deg}(v)$,

- Either returns low conductance cut: STOP
  
  or $\forall v: f(v) \leq \text{deg}(v)$: RESCALE and CALL Unit-Flow again
Excess Scaling

Source supply \( \Delta, |\Delta| = \sum_{v \uparrow} \Delta(v) = 2m \)

Each \( v \) sink of capacity \( \text{deg}(v) \)

- Start with large enough unit value \( \mu \) such that
  \( \forall v: \Delta \downarrow 0(v) = \Delta(v)/\mu \leq 2\text{deg}(v) \)
- Iteratively call Unit-Flow until low conductance cut with “large” volume is returned:
  - If Unit-Flow does not find such a cut, then \( \forall v: f(v) \leq \text{deg}(v) : \text{Set } \Delta \downarrow j+1 \approx 2f \downarrow j \), i.e. \( |\Delta| \) roughly doubles
  - Volume of explored subgraph, roughly doubles

Explored subgraph volume:

\[ 2\text{deg}(v) \rightarrow 4\text{deg}(v) \rightarrow 8\text{deg}(v) \rightarrow 16\text{deg}(v) \ldots \]
Excess Scaling

Low conductance cut in *local time*

- Terminate when encounter “good cut” = Low conductance + large volume
  - $j$-th Unit-Flow: running time $O(|\Delta \downarrow j| \log m / \phi)$
  - Running time of all previous Unit-flow: $O(|\Delta \downarrow j| \log m / \phi)$
Excess Scaling

Low conductance cut in **local time**

- Terminate when encounter “good cut” = Low conductance + large volume
  - $j$-th Unit-Flow: running time $O(|\Delta j|/\log m / \phi)$
  - Running time of all previous Unit-flow: $O(|\Delta j|/\log m / \phi)$
- Cut $A\downarrow j$ returned by last Unit-Flow
  - Low conductance: $O(\phi)$
  - Large volume: $\text{vol}(A) = \Omega(|\Delta j|/\log m)$
- Conductance $\phi$ cut $A$ in time $O(\text{vol}(A) \log^2 m / \phi)$
Excess Scaling

Low conductance cut in local time

- Terminate when encounter “good bottleneck”
  - $j$-th Unit-Flow: running time $O(|Δ↓j|/\log m/\phi)$
  - Running time of all previous Unit-flow: $O(|Δ↓j|/\log m/\phi)$
- Cut $A↓j$ returned by last Unit-Flow
  - Low conductance: $O(\phi)$
  - Large volume: $\text{vol}(A)=\Omega(|Δ↓j|/\log m)$
- Conductance $\phi$ cut $A$ in time $O(\text{vol}(A)\log^2 m/\phi)$
- Otherwise flow spread over $G$, almost all supply routed to sinks
| Excess Scaling + Unit-Flow | vs. | PageRank |
|---------------------------|-----|----------|
| Spread “stuff” to find bottleneck | Flow routing | Probability diffusion |
| Fail when no good enough “bottleneck”, so “stuff” spreads over entire graph |
### Excess Scaling + Unit-Flow vs. PageRank

| Spread “stuff” to find bottleneck | Fail when no good enough “bottleneck”, so “stuff” spreads over entire graph |
|-----------------------------------|--------------------------------------------------------------------------|
| Flow routing                      | Probability diffusion                                                     |
| Quality of cut vs. How easy to spread “stuff” | |

\[
U = O\left(\frac{1}{\phi}\right) \quad \alpha = O(\phi \uparrow 2)
\]
| Excess Scaling + Unit-Flow | vs. | PageRank |
|----------------------------|-----|----------|
| Spread “stuff” to find bottleneck | Flow routing | Probability diffusion |
| Fail when no good enough “bottleneck”, so “stuff” spreads over entire graph |  |
| Quality of cut vs. How easy to spread “stuff” |
| $U = O(1/\phi)$ | $\alpha = O(\phi^{1/2})$ |
| Quality of cut vs. Running time |
| $O(\text{vol}(A)/\phi)$ | $O(\text{vol}(A)/\phi^{1/2})$ |
Wrap-up

Flow-based local low conductance method

- polylog loss versus quadratic loss of PageRank

Framework developed in [KT’15]

Appropriate interface

Deterministic $O(m \log \log m \log m)$ algorithm for min cut in simple unweighted graphs
Open questions

Min cut in more general graphs:

- Determ. $o(mn)$ alg. for multi- or weighted graphs
- Directed graphs

Experimental evaluation

Further applications of flow-based local method:

- Local clustering (ICML‘17)