A Simple and Strongly-Local Flow-Based Method for Cut Improvement Supplemental material

Nate Veldt LVELDT@PURDUE.EDU
Mathematics Department, Purdue University, West Lafayette, IN 47906

David F. Gleich DGLEICH@PURDUE.EDU
Computer Science, Purdue University, West Lafayette, IN

Michael W. Mahoney MMAHONEY@ICSI.BERKELEY.EDU
International Computer Science Institute and Dept. of Statistics, University of California at Berkeley, Berkeley, CA 94720

1. Extra results and proofs

In this section, we include a few of the results we use that also appeared in other material – or had very similar proofs in other material – but restated in our notation for the reader's convenience.

Lemma 1 *If the minimum s-t cut of $G'_R(\alpha, \delta)$ for $\delta \geq 0$ is less than $\alpha \operatorname{vol}(R)$, then $\phi(S) < \alpha$, where S is the node set corresponding to the cut.*

Proof Recall that the min-cut objective can be stated as

$$\min_{S \subset V} \ \alpha \operatorname{vol}(R) + \partial S - \alpha \operatorname{vol}(R \cap S) + (\alpha f(R) + \alpha\delta) \operatorname{vol}(\bar{R} \cap S).$$

If the objective is less than $\alpha \operatorname{vol}(R)$, then

$$\begin{aligned} &\partial S - \alpha \operatorname{vol}(R \cap S) + (\alpha f(R) + \alpha\delta) \operatorname{vol}(\bar{R} \cap S) < 0 \\ \implies &\frac{\partial S}{\operatorname{vol}(R \cap S) - \varepsilon \operatorname{vol}(\bar{R} \cap S)} < \alpha, \end{aligned}$$

where $\varepsilon = f(R) + \delta$. All we need to show then is that

$$\operatorname{vol}(R \cap S) - \varepsilon \operatorname{vol}(\bar{R} \cap S) \leq \min\{\operatorname{vol}(S), \operatorname{vol}(\bar{S})\}$$

and it will follow that

$$\phi(S) = \frac{\partial S}{\min\{\operatorname{vol}(S), \operatorname{vol}(\bar{S})\}} < \alpha.$$

We first note that

$$\operatorname{vol}(R \cap S) - \varepsilon \operatorname{vol}(\bar{R} \cap S) \leq \operatorname{vol}(R \cap S) \leq \operatorname{vol}(S).$$

Also,

$$\begin{aligned} &\operatorname{vol}(R \cap S) - \varepsilon \operatorname{vol}(\bar{R} \cap S) \\ &\leq \operatorname{vol}(R \cap S) - f(R) \operatorname{vol}(\bar{R} \cap S) \\ &= \operatorname{vol}(R) - \operatorname{vol}(R \cap \bar{S}) - f(R) \operatorname{vol}(\bar{R}) + f(R) \operatorname{vol}(\bar{R} \cap \bar{S}) \\ &\leq \operatorname{vol}(R) - f(R) \operatorname{vol}(\bar{R}) + f(R) \operatorname{vol}(\bar{R} \cap \bar{S}) \\ &= f(R) \operatorname{vol}(\bar{R} \cap \bar{S}) \\ &\leq \operatorname{vol}(\bar{S}) \end{aligned}$$

so the result holds. ■

Both assertions in the following theorem are novel results regarding our algorithm SimpleLocal. They can be shown using the same proof techniques used in Lemma 2.2 of Andersen & Lang (2008), with slight alterations to include the locality parameter δ.

Theorem 4 Given an initial reference set $R \subset V$ with $\operatorname{vol}(R) \leq \operatorname{vol}(\bar{R})$, SimpleLocal returns a cut set S^* such that

1. if $C \subseteq R$, then $\phi(S^*) \leq \phi(C)$.
2. For all sets of nodes C such that
$$\frac{\operatorname{vol}(R \cap C)}{\operatorname{vol}(C)} \geq \frac{\operatorname{vol}(R)}{\operatorname{vol}(V)} + \gamma \frac{\operatorname{vol}(\bar{R})}{\operatorname{vol}(V)}$$
for some $\gamma > \delta$, we have $\phi(S^*) \leq \frac{1}{(\gamma - \delta)} \phi(C)$.

Proof We use the same proof outline as Andersen & Lang (2008), and reproduce many of the same steps for the convenience of the reader.

The first assertion holds because if $C \subseteq R$, $\bar{\phi}_R(C) = \phi(C)$, so

$$\phi(S^*) \leq \bar{\phi}_R(S^*) \leq \bar{\phi}_R(C) = \phi(C),$$

where $\bar{\phi}_R$ is used to denote quotient score introduced in equation (7) of the paper. We refer to this as the *modified* quotient score relative to R:

$$\bar{\phi}_R(C) = \frac{\partial C}{\mathrm{vol}(R \cap C) - \varepsilon\,\mathrm{vol}(\bar{R} \cap C)}.$$

To prove the second assertion we start by showing that $\bar{\phi}_R(C) \le \frac{1}{(\gamma-\delta)}\phi(C)$, which is true if and only if

$$\mathrm{vol}(C \cap R) - \varepsilon\,\mathrm{vol}(C \cap \bar{R}) \ge (\gamma - \delta)\,\mathrm{vol}(C).$$

To see this holds we apply the assumption made in the second assertion and simplify:

$$\begin{aligned}
\frac{\mathrm{vol}(R \cap C) - \varepsilon\,\mathrm{vol}(C \cap \bar{R})}{\mathrm{vol}(C)} &= \frac{\mathrm{vol}(C \cap R)}{\mathrm{vol}(C)} - \varepsilon\frac{\mathrm{vol}(C \cap \bar{R})}{\mathrm{vol}(C)} \\
&\ge \frac{\mathrm{vol}(R)}{\mathrm{vol}(V)} + \gamma\frac{\mathrm{vol}(\bar{R})}{\mathrm{vol}(V)} - (f(R) + \delta)\left(1 - \frac{\mathrm{vol}(R)}{\mathrm{vol}(V)} - \gamma\frac{\mathrm{vol}(\bar{R})}{\mathrm{vol}(V)}\right) \\
&= \gamma\frac{\mathrm{vol}(\bar{R})}{\mathrm{vol}(V)}(1 + f(R)) + \frac{\mathrm{vol}(R)}{\mathrm{vol}(V)}\left(1 - \frac{\mathrm{vol}(V)}{\mathrm{vol}(\bar{R})} + \frac{\mathrm{vol}(R)}{\mathrm{vol}(\bar{R})}\right) \\
&\quad - \delta\left(1 - \frac{\mathrm{vol}(R)}{\mathrm{vol}(V)} - \gamma\frac{\mathrm{vol}(\bar{R})}{\mathrm{vol}(V)}\right) \\
&= \gamma \cdot 1 + 0 - \delta\left(1 - \frac{\mathrm{vol}(R)}{\mathrm{vol}(V)} - \gamma\frac{\mathrm{vol}(\bar{R})}{\mathrm{vol}(V)}\right) \\
&\ge \gamma - \delta.
\end{aligned}$$

Since S^* is the set that minimizes $\bar{\phi}_R(S)$, we have

$$\phi(S^*) \le \bar{\phi}_R(S^*) \le \bar{\phi}_R(C) \le \frac{1}{(\gamma - \delta)}\phi(C).$$ ■

2. Empirical Runtime of SimpleLocal

In terms of the runtime, the spectral method is substantially faster in practice than our sequence of max-flow problems. (See Table 1 in the main text.) This arises due to a few factors. First, we are using a carefully engineered code for the spectral algorithm designed for speed. Second, we are using a general-purpose linear programming solver for the maximum-flow problems. Third, we are not exploiting any possible "warm-start" between independent flow solutions. We anticipate that a more careful implementation within our highly flexible three-stage framework would shrink the runtime gap considerably.

3. Experiment parameters for the MRI problem

We obtained a labeled MRI scan from the MICCAI-2012 challenge with $256 \times 287 \times 256$ voxels (around 18 million). (The MRI scans originated with the OASIS project, and labeled data was provided by Neuromorphometrics, Inc. `neuromorphometrics.com` under an academic subscription.) We assembled a nearest neighbor graph on this image using all 26 spatially adjacency voxels where each edge was weighted similar to Shi & Malik (2000). We used the function $e^{-(\sqrt{I_i}-\sqrt{I_j})^2/0.05^2}$ where $\sqrt{I_i}$ is the scan intensity at voxel i. Subsequently, we threshholded the graph at a minimum weight of 0.1 and scaled each edge weight to have minimum weight 1 so that the volume of a set was an upperbound on the number of edges contained. The final graph was connected except for 35 voxels and contained 467 million edges.

Seeding and SimpleLocal We picked 75 random voxels in the true image, then used SimpleLocal to refine the set R consisting of these 75 voxels and their immediate neighbors using a value of $\delta = 0.1$ to keep the computation local. The seed set is shown here in the supplement in Figure 1. The resulting set is show in 4(b).

Refinement The output from SimpleLocal can be further improved by growing the set by its neighborhood and varying δ. We call this "refinement" and used one step of refinement with $\delta = 0.5$. The result is in Figure 4(c).

Spectral We compare this against a highly-optimized strongly-local spectral method to minimize conductance using personalized PageRank vectors (Andersen et al., 2006), where the PageRank computation uses $\alpha = 0.99$. The spectral result is in the final subfigure Figure 4(d).

Parameter selection We picked parameters for the flow methods to ensure that the volume explored would be around 10 times the volume of the desired ventricle, and occassionally reduced the parameter δ if it seemed that the method was exploring too much or if the flow problems took too long. We picked the parameters for the spectral method until we found a set that meaningfully grew. Our particular technique attempts to avoid diffusing as much as possible and so we had to adjust the parameters to ensure that it moved beyond the seed set.

3.1. Near optimiality of Refined SimpleLocal

We can use our SimpleLocal and 3StageFlow primitives to attempt to identify the *best* and *largest* conductance set largely contained within the target ventricle. This is essentially the best result we could hope to achieve as the entire desired set has conductance larger than the set we identify. Thus, if we run a single iteration of 3StageFlow using the entire target set as R, $\alpha = 0.1291$ (the conductance of the target set), and $\delta = 15$, we will find a set that is almost exclusively contained within the target ventricle (Figure 2). This choice of δ is guided by the intuition that we want the set to be *almost exclusively* in the interior of the target, but small variations outside would be okay. The resulting solution set found has conductance 0.0621 and 2527 vertices. The difference between the refined set we generated

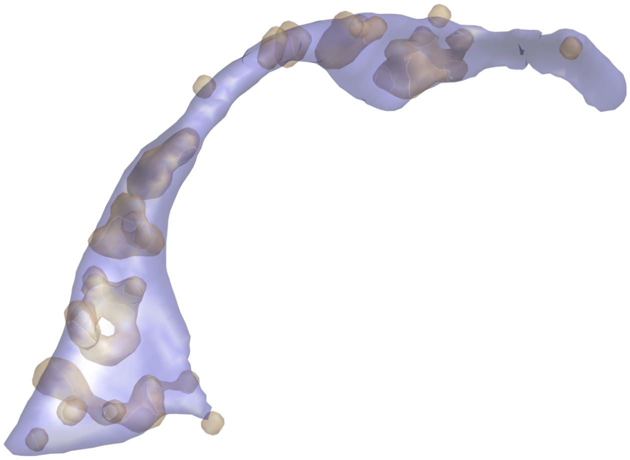


Figure 1. The seed set *R* for the MRI segmentation. As in the main body figures, the true ventricle is shown in blue and the set in orange.

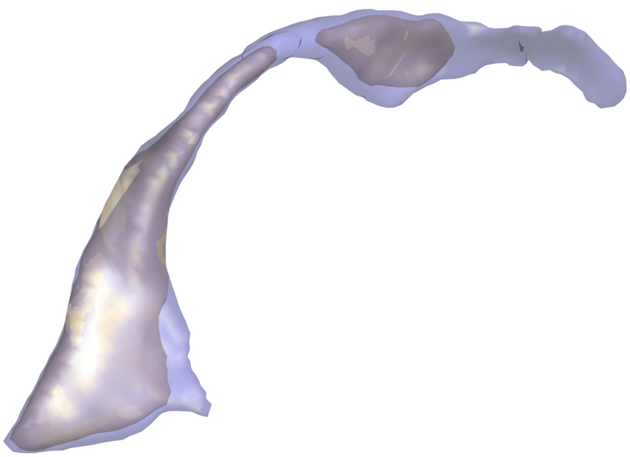
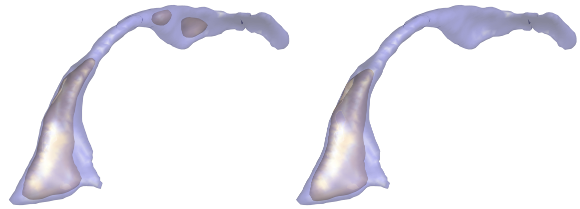


Figure 2. The largest set we identified with a small value of conductance inside the target ventricle. This is essentially the best set we could hope to identify from our flow techniques.



(a) Optimized spectral (b) Component of Refined

Figure 3. At left, we have another set from spectral that identifies a low-conductance set nearly strictly inside. At right, we show the best subset of the disconnected region identified by SimpleLocal and the refinement procedure. The spectral set has conductance 0.079 and the SimpleLocal component has conductance 0.0398. Note that the spectral set does not hug the boundary nearly as closely as the results from the SimpleLocal method in the main paper.

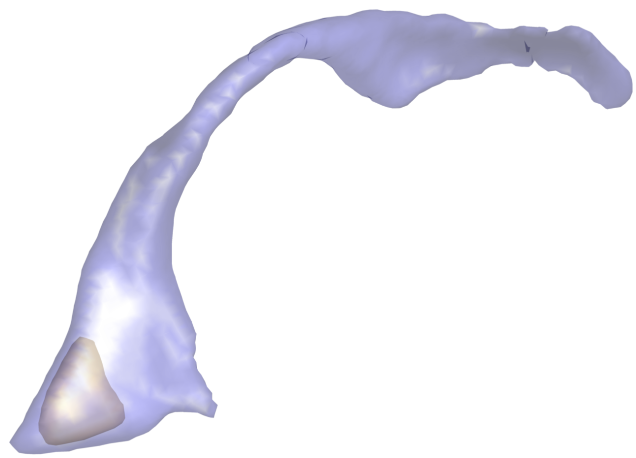


Figure 4. A tiny set of 295 vertices with conductance 0.048 buried deep within the ventricle. This set often attracts the flow-based method if the value of δ is set too high.

(Figure 4(c) in the main text) and this set is slight. Their intersection is 2317 voxels. So there is a slightly better set that SimpleLocal and the refinement procedure could have generated, but not by much.

3.2. Other good sets

We highlight a few other low-conductance sets we identified in the course of our experiments in Figure 3 and Figure 4. In the first figure, we show another set available from the spectral method that makes a boundary error in the other direction and ends up too far inside the set. A closely related set in Figure 3 is, perhaps, the optimal set contained within the the target ventricle. It has the lowest conductance score of any set we ever computed. One challenge with using the flow-based methods such as SimpleLocal is that they tend to quickly contract to very good, small sets. For instance, there is a set of 295 vertices with very good conductance (Figure 4). If the parameter δ is set too high, then this often causes the flow-based method to contract too much (e.g. we over-regularize) and identify a very precise small set. This feature could be useful in some applications where the conductance measure is a very good proxy for the desired output.

References

Andersen, Reid and Lang, Kevin. An algorithm for improving graph partitions. In *Proceedings of the 19th annual ACM-SIAM Symposium on Discrete Algorithms (SODA2008)*, pp. 651–660, January 2008.

Andersen, Reid, Chung, Fan, and Lang, Kevin. Local graph partitioning using PageRank vectors. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Sci-*

ence, 2006. URL http://www.math.ucsd.edu/~fan/wp/localpartition.pdf.

Shi, Jianbo and Malik, J. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, August 2000. doi: 10.1109/34.868688.