# Appendix

### Comparison between constraint (4) and other measures of connectivity

To form a better understanding of how conductance arises in this context, consider a (dual) flow formulation of connectedness: $S \in \Lambda_r$ if, for any $i \in S$, there exists a way to route one unit of flow from the root $r$ to $i$ with finite congestion. This can be strengthened by requiring that 1 unit of flow be routed from $r$ to $i$ with congestion $\frac{1}{k}$, yielding the notion of $k$-edge connectivity. Finally, we can strengthen the flow requirement further, by demanding that (1) $d_i$ units of flow be routed from $r$ to $i$, and (2) the flows from $r$ to all vertices $i \in S$ be routed *concurrently* with congestion $\frac{1}{\gamma}$. The maxflow-mincut theorem shows that such a flow routing exists if and only if the condition of (4) holds.

While the notions of connectivity and conductance converge to connectedness in the limit as $\gamma$ goes to 0, they display different behaviors for larger $\gamma$. In particular, conductance appears to be a more meaningful in the context of anomaly detection scenarios, where the anomalous set may be constructed by an unspecified diffusion process, such as the epidemic in the example of disease outbreak detection, which is unlikely to cross low-conductance cuts. Moreover, for noisy input graphs, conductance is a more robust notion than edge connectivity.

### Proof of Theorem 3.2

*Proof.* We prove the contrapositive. Given a vector $x$ such that

$$x^\top L_{G_S} x < \frac{\gamma^2}{2} x^\top L_{\mathrm{Star}_S^{(r)}} x,$$

we produce a cut $T \subseteq S, r \notin T$, with $\phi_{G_S}(T) < \gamma$. To do so, consider the embedding $y \in \mathbb{R}^n$ given by $y_i = |x_i - x_r|$. Then, it is easy to check that:

$$\begin{aligned} y^\top L_{G_S} y \le x^\top L_{G_S} x &< \frac{\gamma^2}{2} x^\top L_{\mathrm{Star}_S^{(r)}} x \\ &= \frac{\gamma^2}{2} \sum_{i \in S \setminus \{r\}} d_i y_i^2. \end{aligned}$$

By Lemma 3.1, a sweep cut of $y$ yields a cut $T$ such that $\phi_{G_S}(T) < \gamma$. Moreover, because $y_r = 0$, $r$ does not belong to $T$. ☐

### Proof of Theorem 3.3 and an alternative formulation using effective resistance

In this section we offer an alternative approach to obtain the inequality constraint (9) through electrical networks and the concept of effective resistance. Using this formulation, we then prove Theorem 3.3 that shows that the inequality constraint enforces connected through a simple rounding of $M$.

We shortly introduce the concept of effective resistance in the electrical network interpretation of graphs. In contrast to $s$-$t$ flow that interprets edge weights on a graph as flow capacities, electrical flow considers edge weights $w_{ij}$ as the conductance (inverse resistance) $\frac{1}{r_{ij}}$ between two nodes $i, j$.

Define the pseudoinverse of a Laplacian $L^+ = \sum_{i=2}^n \frac{1}{\lambda_i} v_i v_i^\top$ where $L = \sum_{i=2}^n \lambda_i v_i v_i^\top$ is its eigendecomposition, and we note that the minimum eigenvalue $\lambda_1$ corresponding to all 1 eigenvector is zero. Also note that $L_{G[M]}$ denotes the Laplacian of the subgraph with adjacency matrix $A \odot M$ ($\odot$ is the elementwise/Hadamard matrix product)[2] and let $L^+_{G[M]}$ denote its pseudoinverse.

Defining a vector of directional current flows into/out of nodes with $f$ (e.g. where positive elements are currents into the node and vice versa), the relationship between the vector of voltages $v$ and $f$ in an electrical circuit graph with resistances $r_{ij}$ are given by the relation $v = L^+ f$. This fact follows from Kirchoff's current and voltage laws (Vishnoi, 2012). The effective resistance $R_{ab}$ between any two nodes $a, b$ is then defined by the voltage difference $v_a - v_b = v^\top (e_a - e_b)$ when unit current is flowing between two nodes such that $f = e_a - e_b$ (or equivalently, inverse of the current flow with unit voltage difference between $a$ and $b$). We then have the identity $R_{ab} = (e_a - e_b)^\top v = (e_a - e_b)^\top L^+ (e_a - e_b)$.

Given a root/anchor node $r \in V$ that is assumed to be contained in the subgraph $S$, consider the case where a current flow of $d_i M_{ii}$ is present between $r$ and $i \in V$ on the induced graph with Laplacian $L_{G[M]}$. Letting $m_i = d_i M_{ii}(e_r - e_i)$, the voltage vector $v^i$ corresponding to these inputs is given by $v^i = L^+_{G[M]} m_i$. We then obtain the resistance between nodes $r$ and $i$ for input current $d_i M_{ii}$ with the identity $v^i_r - v^i_i = (e_r - e_i) L^+_{G[M]} m_i = \frac{1}{d_i M_{ii}} m_i^\top L^+_{G[M]} m_i$, which we intend to be finite for $i \in S$ (and thus $M_{ii} > 0$) if $G_S$ is connected. For a conductance threshold $\tau$, we would thus like to impose the constraint

$$m_i^\top L^+_{G[M]} m_i \le \frac{d_i M_{ii}}{\tau}, \tag{12}$$

for all $i \in V$, where the scaling with $M_{ii}$ serves to restrict the constraint to only the nodes $i \in S$. We also remark that the constraint (12) is independent of uniform multiplicative scaling of $M$, thus constraining $M$ to unit trace does not affect the choice of $\tau$.

We next unify the $n$ different resistance constraints to a single PSD constraint. Define the $2n \times 2n$ matrix $\mathcal{A}_M$ as

$$\mathcal{A}_M = \begin{pmatrix} L_{G[M]} & L_{\mathrm{Star}^{(r)}[M]} \\ L_{\mathrm{Star}^{(r)}[M]} & \frac{1}{\tau} L_{\mathrm{Star}^{(r)}[M]} \end{pmatrix},$$

[2] Corresponding to the original graph with edge weights scaled by $M_{ij}$.

for which we define our connectivity constraint to be $\mathcal{A}_M \succeq 0$. For $i \neq r$ consider the submatrix $\mathcal{A}_i$ formed by the top-left $n \times n$ submatrix (i.e. $L_{G[M]}$) and the $(n+i)$-th row and column, which results in

$$\mathcal{A}_i = \begin{pmatrix} L_{G[M]} & m_i^\top \\ m_i & \frac{d_i M_{ii}}{\tau} \end{pmatrix}.$$

Since $\mathcal{A}_M \succeq 0$ it implies that any such submatrix satisfies $\mathcal{A}_i \succeq 0$ and through the Schur complement condition that is equivalent to condition (12). We have thus shown that the condition $\mathcal{A}_M \succeq 0$ encapsulates the pairwise effective resistance constraints of the form (12). We note that we obtain a similar condition for the root node, where with currents $m_r \triangleq \sum_{j \neq r} m_j$ applied to the nodes and a convex combination of voltage differences, $m_r^\top L^+_{G[M]} m_r$ is being compared to $\frac{\sum_{j \neq r} M_{jj}}{\tau}$.

Finally, we can again utilize the Schur complement condition for positive semidefiniteness and simplify the PSD condition $\mathcal{A}_M \succeq 0$ on the $2n \times 2n$ matrix to a PSD condition on an $n \times n$ matrix, which directly leads to inequality (9) when we replace $\tau$ with $\frac{\gamma^2}{2}$.

Now we prove Theorem 3.3 using the above reformulation of (9). Assume there exists $i \in \hat{S}$ such that there exists no path between $r$ and $i$, i.e., the subgraph $G_{\hat{S}}$ is disconnected. Since $Q_\gamma(M) \succeq 0$, it follows that $\mathcal{A}_M \succeq 0$ and thus $\mathcal{A}_i \succeq 0$ for $\tau = \frac{\gamma^2}{2}$. Through the Schur complement lemma and (12) we have that $v_r^i - v_i^i \leq \frac{1}{\tau}$, i.e., that the voltage difference between nodes $r$ and $i$ is finite when a current flow of $d_i M_{ii}$ is applied that is non-zero (since $M_{ii} > 0$). This voltage difference is computed for the graph $G$ with edge weights given by $M_{ij}$. It is easy to see that if this voltage difference is finite, the voltage difference in the original graph $G$ is also finite since $M_{ij}$ are bounded. However a contradiction arises because there exists no path between $i$ and $r$ in $G$ and thus the effective resistance $R_{ri} = (e_r - e_i)^\top L^+ (e_r - e_i)$ is infinite.

**Proof of Theorem 3.4**

*Proof.* Consider an optimal solution $M$ of arbitrary rank. Construct a tentative rank-1 solution $m$ by taking $m_i = \sqrt{M_{ii}}$. Notice that, because $M \succeq 0$, we must have $M_{ij} \leq \sqrt{M_{ii} M_{jj}} = m_i m_j$. Then: $C \cdot mm^\top = \sum_{i,j} C_{ij} m_i m_j \geq C \cdot M$, by the non-negativity of $C$. Moreover, $Q_\gamma(mm^\top) \succeq Q_\gamma(M)$ as the weights of $G[mm^\top]$ are larger than the weights of $G[M]$. The theorem follows as $mm^T$ is a feasible rank-1 solution of value at least as large as the optimal. □

**Proof of Lemma 4.1**

*Proof.* Notice that: $\delta \cdot \nabla_Y f(Y^{(t)}) = \left(L_{G[M]} - \frac{\gamma^2}{2} \cdot L_{\mathrm{Star}^{(r)}[M]}\right)$. Because $I_n \cdot M = 1$, we have:

$$-\gamma^2 \cdot D \preceq -\frac{\gamma^2}{2} \cdot L_{\mathrm{Star}^{(r)}} \preceq \delta \cdot \nabla_Y f(Y^{(t)}) \preceq L_G \preceq 2D$$

It follows that:

$$\|D^{-1/2} \nabla_Y f(Y^{(t)}) D^{-1/2}\| \leq \frac{2}{\delta}$$

□