# Constrained fractional set programs and their application in local clustering and community detection

- ICML13

Author: Thomas Buhler

# Outline

1. **Background**
    a. Constrained balanced graph cuts for local clustering
    b. Constrained local community detection
2. Tight relaxations
    a. Unconstrained fractional set programs
    b. Constrained fractional set programs
3. Minimization of tight continuous relaxation
4. Tight relaxations of constrained maximum density problem
5. Experiments

# Constrained balanced graph cuts for local clustering

Problem: Find a local cluster around a given seed set.

Objective function (1):

$$\min_{C \subset V} \frac{\mathrm{cut}(C, \overline{C})}{\widehat{S}(C)}$$

$$\text{subject to}: \ \mathrm{vol}_g(C) \leq k, \quad \text{and} \quad J \subset C.$$

$\mathrm{vol}_g(C)$ is general volume of set C. $\mathrm{vol}_g(A) = \sum_{i \in A} g_i$

# Constrained local community detection

Problem: Search for a set C which has high association.

Objective function (2):

$$\max_{C \subset V} \frac{\text{assoc}(C)}{\text{vol}_g(C)}$$

$$\text{subject to} : k_1 \leq \text{vol}_h(C) \leq k_2, \quad \text{and} \quad J \subset C,$$

# Connection

In (2), If $\mathrm{vol}_g(C) = \mathrm{vol}_d(C)$

$$\frac{\mathrm{assoc}(C)}{\mathrm{vol}_g(C)} = \frac{\mathrm{assoc}(C)}{\mathrm{vol}_d(C)} = 1 - \frac{\mathrm{cut}(C,\overline{C})}{\mathrm{vol}_d(C)}$$

In (1). If $\widehat{S}(C) = \mathrm{vol}_d(C)$

$$\frac{\mathrm{cut}(C,\overline{C})}{\widehat{S}(C)} = \frac{\mathrm{cut}(C,\overline{C})}{\mathrm{vol}_d(C)}$$

# Contribution

All constrained non-negative fractional set programs have an equivalent tight continuous relaxation.

# Outline

1. Background
   a. Constrained balanced graph cuts for local clustering
   b. Constrained local community detection
2. **Tight relaxations**
   a. Unconstrained fractional set programs
   b. Constrained fractional set programs
3. Minimization of tight continuous relaxation
4. Tight relaxations of constrained maximum density problem
5. Experiments

# Unconstrained fractional set programs

Objective function:

$$\min_{C \subset V} \frac{\widehat{R}(C)}{\widehat{S}(C)} =: \widehat{Q}(C)$$

$$\text{subject to} : \widehat{M}_i(C) \leq k_i, \quad i = 1, \ldots, K$$

Assuming $\widehat{R}, \widehat{S}$ is non-negative and $\widehat{R}(\emptyset) = \widehat{S}(\emptyset) = 0$

# Unconstrained fractional set programs

Connection between the set-valued and the continuous space.

Thresholding:

Let $f \in \mathbb{R}^n$, we assume f is ordered in ascending order,

$f_1 \leq f_2 \leq \cdots \leq f_n$

$$C_i := \{j \in V | f_j \geq f_i\}, \qquad i = 1, \ldots, n.$$

# Unconstrained fractional set programs

Lovasz extension:

**Definition 1** *Let* $\widehat{R} : 2^V \to \mathbb{R}$ *be a set function with* $\widehat{R}(\emptyset) = 0$, *and* $f \in \mathbb{R}^n$ *in ascending order* $f_1 \leq f_2 \leq \cdots \leq f_n$. *The Lovasz extension* $R : \mathbb{R}^n \to \mathbb{R}$ *of* $\widehat{R}$ *is defined as* $R(f) = \sum_{i=1}^{n-1} \widehat{R}(C_{i+1}) \, (f_{i+1} - f_i) + \widehat{R}(V) f_1$.

$R(\mathbf{1}_C) = \widehat{R}(C)$ for all $C \subset V$

Use hat-symbol to denote set function.

# Unconstrained fractional set programs

Submodular set function:

**Definition 2** *A set function* $\widehat{R} : 2^V \to \mathbb{R}$ *is submodular if for all* $A, B \subset V$, $\widehat{R}(A \cup B) + \widehat{R}(A \cap B) \leq \widehat{R}(A) + \widehat{R}(B)$. *It is supermodular, if the converse inequality holds true, and modular if we have equality.*

# Unconstrained fractional set programs

Connection:

Lovasz extension of submodular set functions are convex function.

**Proposition 1** *Let* $R : \mathbb{R}^V \to \mathbb{R}$ *be the Lovasz extension of* $\widehat{R} : 2^V \to \mathbb{R}$*. Then,* $\widehat{R}$ *is submodular if and only if* $R$ *is convex. Furthermore, if* $\widehat{R}$ *is submodular, then* $\min_{A \subset V} \widehat{R}(A) = \min_{f \in [0,1]^n} R(f)$.

# Unconstrained fractional set programs

Properties:

**Proposition 2** *Let $R : \mathbb{R}^V \to \mathbb{R}$ be the Lovasz extension of $\widehat{R} : 2^V \to \mathbb{R}$. Then,*

- *$R$ is positively one-homogeneous*[4],
- *$R(f) \geq 0,\ \forall\, f \in \mathbb{R}^V$ and $R(\mathbf{1}) = 0$ if and only if $\widehat{R}(A) \geq 0,\ \forall\, A \subset V$ and $\widehat{R}(V) = 0$,*
- *Let $S : \mathbb{R}^V \to \mathbb{R}$ be the Lovasz extension of $\widehat{S} : 2^V \to \mathbb{R}$. Then, $\lambda_1 R + \lambda_2 S$ is the Lovasz extension of $\lambda_1 \widehat{R} + \lambda_2 \widehat{S}$, for all $\lambda_1, \lambda_2 \in \mathbb{R}$.*

# Unconstrained fractional set programs

Objective function:

$$\min_{C \subset V} \frac{\widehat{R}(C)}{\widehat{S}(C)} \quad \text{relax to} \quad \inf_{f \in \mathbb{R}^n_+} \frac{R(f)}{S(f)}$$

Given f, $C' = \mathop{\arg\min}_{C_i, i=1,\ldots,n} \frac{\widehat{R}(C_i)}{\widehat{S}(C_i)}$

# Unconstrained fractional set programs

**Theorem 1 (b)** *Let* $\widehat{R}, \widehat{S} : 2^V \to \mathbb{R}$ *be non-negative set functions and* $\widehat{R} := \widehat{R}_1 - \widehat{R}_2$ *and* $\widehat{S} := \widehat{S}_1 - \widehat{S}_2$ *be decompositions into differences of submodular set functions. Let the Lovasz extensions of* $\widehat{R}_1, \widehat{S}_2$ *be given by* $R_1, S_2$ *and let* $R_2', S_1'$ *be positively one-homogeneous convex functions with* $S_1'(\mathbf{1}_A) = \widehat{S}_1(A)$ *and* $R_2'(\mathbf{1}_A) = \widehat{R}_2(A)$ *such that* $S_1' - S_2$ *is non-negative. Define* $R := R_1 - R_2'$ *and* $S := S_1' - S_2$*. Then,*

$$\inf_{C \subset V} \frac{\widehat{R}(C)}{\widehat{S}(C)} = \inf_{f \in \mathbb{R}_+^V} \frac{R(f)}{S(f)} \, .$$

# Unconstrained fractional set programs

*Moreover, it holds for all* $f \in \mathbb{R}^n_+$, $\frac{R(f)}{S(f)} \geq \min_{i=1,\ldots,n} \frac{\widehat{R}(C_i)}{\widehat{S}(C_i)}$. *Thus a minimizer of the set ratio can be found by optimal thresholding. Let furthermore* $\widehat{R}(V) = \widehat{S}(V) = 0$, *then all the above statements hold if one replaces* $\mathbb{R}^n_+$ *with* $\mathbb{R}^n$.

Meaning: Every non-negative fractional set programs has a tight relaxation into a continuous fractional program.

# Unconstrained fractional set programs

**Lemma 1** *Let* $\widehat{R} : 2^V \to \mathbb{R}$ *be a submodular set function with* $\widehat{R}(\emptyset) = 0$. *Let* $R'$ *be a positively one-homogeneous convex function with* $R'(\mathbf{1}_A) = \widehat{R}(A)$ *for all* $A \subset V$. *Then, it holds* $\forall f \in \mathbb{R}^V_+$ *that*

$$R'(f) \leq \sum_{i=1}^{n-1} \widehat{R}(C_{i+1})\,(f_{i+1} - f_i) + f_1 \widehat{R}(V).$$

*Let furthermore* $\widehat{R}(V) = 0$, *then the above inequality holds for all* $f \in \mathbb{R}^V$.

Meaning: Lovasz extension of a submodular set function is an upper bound on any one-homogeneous convex function.

# Unconstrained fractional set programs

**Lemma 2** *Let* $\widehat{R}, \widehat{S} : 2^V \to \mathbb{R}$ *and* $R, S : \mathbb{R}^n \to \mathbb{R}$ *satisfy the assumptions of Theorem 1 (b). Then for all* $f \in \mathbb{R}^V_+$,

$$\frac{R(f)}{S(f)} \geq \min_{i=1,\ldots,n} \frac{\widehat{R}(C_i)}{\widehat{S}(C_i)} .$$

*Let furthermore* $\widehat{R}(V) = \widehat{S}(V) = 0$, *then the result holds for all* $f \in \mathbb{R}^V$.

# Unconstrained fractional set programs

**Proof of Theorem 1 (b):** Lemma 2 implies that

$$\inf_{f \in \mathbb{R}_+^V} \frac{R(f)}{S(f)} \geq \inf_{f \in \mathbb{R}_+^V} \min_{\substack{C_i \text{ def. by } f \\ i=1,\ldots,n}} \frac{\widehat{R}(C_i)}{\widehat{S}(C_i)} \geq \inf_{A \subset V} \frac{\widehat{R}(A)}{\widehat{S}(A)} .$$

On the other hand we have

$$\inf_{A \subset V} \frac{\widehat{R}(A)}{\widehat{S}(A)} = \inf_{A \subset V} \frac{R(\mathbf{1}_A)}{S(\mathbf{1}_A)} \geq \inf_{f \in \mathbb{R}_+^V} \frac{R(f)}{S(f)} ,$$

# Constrained fractional set programs

Transform constraints to penalty terms:

$$\widehat{M}_i(C) \leq k_i \quad \longrightarrow \quad \widehat{T}_i(C) = \begin{cases} \max\left\{0, \widehat{M}_i(C) - k_i\right\}, & C \neq \emptyset, \\ 0, & C = \emptyset. \end{cases}$$

# Constrained fractional set programs

Objective function:

$$\min_{C \subset V} \frac{\widehat{R}(C)}{\widehat{S}(C)} =: \widehat{Q}(C)$$

$$\text{subject to : } \widehat{M}_i(C) \leq k_i, \quad i = 1, \ldots, K$$

↓

$$\min_{C \subset V} \frac{\widehat{R}(C) + \gamma \sum_i^K \widehat{T}_i(C)}{\widehat{S}(C)} =: \widehat{Q}_\gamma(C)$$

# Constrained fractional set programs

Define $\theta$ is the minimum value of $\widehat{T}_i$`

$$\theta = \min_{i=1,\ldots,K} \left[ \min_{\widehat{M}_i(C) > k_i} \widehat{M}_i(C) - k_i \right]$$

# Constrained fractional set programs

**Theorem 2** *Let* $\widehat{R}, \widehat{S} : 2^V \to \mathbb{R}$ *be non-negative set functions and* $R$, $S$ *their Lovasz extensions. Let* $C_0 \subset V$ *be feasible and* $\widehat{S}(C_0) > 0$. *Denote by* $T$ *the Lovasz extension of* $\widehat{T}$. *Then, for* $\gamma > \frac{\widehat{R}(C_0)}{\theta \widehat{S}(C_0)} \max_{C \subset V} \widehat{S}(C)$,

$$\min_{\substack{\widehat{M}_i(C) \leq k_i, \\ i=1,\ldots,K}} \frac{\widehat{R}(C)}{\widehat{S}(C)} = \min_{f \in \mathbb{R}^n_+} \frac{R(f) + \gamma T(f)}{S(f)} := Q_\gamma(f)$$

*Moreover, for any* $f \in \mathbb{R}^n_+$ *with* $Q_\gamma(f) < \widehat{Q}_\gamma(C_0)$ *for the given* $\gamma$, *we have* $Q_\gamma(f) \geq \min_{i=1,\ldots,n} \widehat{Q}_\gamma(C_i)$, *and the minimizing set on the right hand side is feasible.*

# Constrained fractional set programs

Meaning: The set found by optimal thresholding of the solution of the continuous program is guaranteed to satisfy all constraints.

Proof:

Suppose $C^* \neq \emptyset$ is a minimizer of $\min_{f \in \mathbb{R}^n_+} \frac{R(f) + \gamma T(f)}{S(f)}$ and infeasible.

So, $\widehat{T}(C^*) \geq \theta$.

$$\widehat{Q}_\gamma(C^*) = \frac{\widehat{R}(C^*) + \gamma \widehat{T}(C^*)}{\widehat{S}(C^*)} \qquad (8)$$

$$\geq \frac{\gamma \widehat{T}(C^*)}{\widehat{S}(C^*)} \geq \frac{\gamma \widehat{T}(C^*)}{\max_{C \subset V} \widehat{S}(C)} \geq \frac{\gamma \theta}{\max_{C \subset V} \widehat{S}(C)} > \frac{\widehat{R}(C_0)}{\widehat{S}(C_0)} = \widehat{Q}_\gamma(C_0)$$

# Constrained fractional set programs

This contradicts the fact that $C^*$ is optimal.
So, it is feasible.

for the second statement, by lemma 2.
suppose $Q_\gamma(f) < \widehat{Q}_\gamma(C_0)$, then $Q_\gamma(f) \geq \min_{i=1,\dots,n} \widehat{Q}_\gamma(C_i)$
suppose $C^*$ is minimizer of right hand side, and infeasible,
so, $\widehat{Q}_\gamma(C^*) \geq \frac{\gamma\theta}{\max_{C\subset V} \widehat{S}(C)} > \widehat{Q}_\gamma(C_0)$
which lead to contradiction, so $C^*$ is feasible.

# Outline

1. Background
    a. Constrained balanced graph cuts for local clustering
    b. Constrained local community detection
2. Tight relaxations
    a. Unconstrained fractional set programs
    b. Constrained fractional set programs
3. **Minimization of tight continuous relaxation**
4. Tight relaxations of constrained maximum density problem
5. Experiments

# Minimization of tight continuous relaxation

Objective function:

$$\min_{f \in \mathbb{R}_+^n} \frac{R(f)}{S(f)} := Q(f)$$

**Proposition 3** *Every set function $\widehat{S}$ with $\widehat{S}(\emptyset) = 0$ can be written as $\widehat{S} = \widehat{S}_1 - \widehat{S}_2$, where $S_1$ and $S_2$ are submodular and $\widehat{S}_1(\emptyset) = \widehat{S}_2(\emptyset) = 0$. The Lovasz extension $S$ can be written as difference of convex functions.*

# Minimization of tight continuous relaxation

Algorithm:

**RatioDCA** Minimization of a non-negative ratio of one-homogeneous d.c functions over $\mathbb{R}^n_+$

1: **Initialization:** $f^0 \in \mathbb{R}^n_+$, $\lambda^0 = Q(f^0)$

2: **repeat**

3: $\quad f^{l+1} = \underset{u \in \mathbb{R}^n_+,\ \|u\|_2 \leq 1}{\arg\min} \left\{ R_1(u) - \langle u, r_2(f^l) \rangle + \lambda^l \big( S_2(u) - \langle u, s_1(f^l) \rangle \big) \right\}$

$\quad$ where $r_2(f^l) \in \partial R_2(f^l)$, $s_1(f^l) \in \partial S_1(f^l)$

4: $\quad \lambda^{l+1} = Q(f^{l+1})$

5: **until** $\frac{|\lambda^{l+1} - \lambda^l|}{\lambda^l} < \epsilon$

# Minimization of tight continuous relaxation

Convergence proof:

**Proposition 4** *The sequence $f^l$ produced by RatioDCA satisfies $Q(f^{l+1}) < Q(f^l)$ for all $l \geq 0$ or the sequence terminates.*

Proof:

$$\Phi_{f^l}(u) := R_1(u) - \langle u, r_2(f^l) \rangle + \lambda^l \big( S_2(u) - \langle u, s_1(f^l) \rangle \big)$$

The optimal of this function is non-positive.

# Minimization of tight continuous relaxation

$$\begin{aligned}\Phi_{f^l}(f^l) &= R_1(f^l) - \langle f^l, r_2(f^l)\rangle \\ &\quad + \lambda^l\big(S_2(f^l) - \langle f^l, s_1(f^l)\rangle\big) \\ &= R_1(f^l) - R_2(f^l) + \lambda^l\big(S_2(f^l) - S_1(f^l)\big) = 0\end{aligned}$$

$\langle f^l, r_2(f^l)\rangle = R_2(f^l)$ $\quad$ $\langle f^l, s_1(f^l)\rangle = S_1(f^l)$

# Minimization of tight continuous relaxation

$$
\begin{aligned}
0 &> \Phi_{f^l}(f^{l+1}) \\
&= R_1(f^{l+1}) - \langle f^{l+1}, r_2(f^l) \rangle \\
&\quad + \lambda^l \big( S_2(f^{l+1}) - \langle f^{l+1}, s_1(f^l) \rangle \big) \\
&\geq R_1(f^{l+1}) - R_2(f^{l+1}) + \lambda^l \big( S_2(f^{l+1}) - S_1(f^{l+1}) \big)
\end{aligned}
$$

$$S(f) \geq S(g) + \langle f - g, s(g) \rangle = \langle f, s(g) \rangle$$

so,

$$Q(f^{l+1}) = \frac{R_1(f^{l+1}) - R_2(f^{l+1})}{S_1(f^{l+1}) - S_2(f^{l+1})} < \lambda^l = Q(f^l)$$

# Outline

1. Background
   a. Constrained balanced graph cuts for local clustering
   b. Constrained local community detection
2. Tight relaxations
   a. Unconstrained fractional set programs
   b. Constrained fractional set programs
3. Minimization of tight continuous relaxation
4. **Tight relaxations of constrained maximum density problem**
5. Experiments

# Tight relaxations of constrained maximum density problem

Objective function:

$$\max_{C \subset V} \frac{\text{assoc}(C)}{\text{vol}_g(C)}$$

$$\text{subject to} : \text{vol}_h(C) \leq k, \quad \text{and} \quad J \subset C.$$

↓

$$\min_{\substack{C \subset V \\ \text{s.t.} J \subset C}} \frac{\text{vol}_g(C) + \gamma \widehat{T}_k(C)}{\text{assoc}(C)}$$

$$\widehat{T}_k(C) = \max\{0, \text{vol}_h(C) - k\} \qquad \gamma > \frac{\text{vol}_g(C_0)\,\text{vol}(V)}{\theta\,\text{assoc}(C_0)} \text{ for a feasible set } C_0 \subset V$$

## Tight relaxations of constrained maximum density problem

Objective function:

$$\min_{A \subset V \setminus J} \frac{\mathrm{vol}_g(A) + \mathrm{vol}_g(J) + \gamma \widehat{T}_{k'}(A)}{\mathrm{assoc}(A) + \mathrm{assoc}(J) + 2\mathrm{cut}(J, A)}$$

For technical reasons, we change:

$\mathrm{vol}_g(J)$ to $\mathrm{vol}_g(J)\widehat{P}(A)$ $\quad$ $\mathrm{assoc}(J)$ to $\mathrm{assoc}(J)\widehat{P}(A)$

$\widehat{P}(A) = 1$ for $A \neq \emptyset$ $\quad$ $\widehat{P}(\emptyset) = 0.$

## Tight relaxations of constrained maximum density problem

Objective function:

$$\min_{A \subset V \setminus J} \frac{\mathrm{vol}_g(A) + \mathrm{vol}_g(J)\widehat{P}(A) + \gamma \widehat{T}_{k'}(A)}{\mathrm{assoc}(A) + \mathrm{assoc}(J)\widehat{P}(A) + 2\mathrm{cut}(J, A)}$$

Because in this way, we will not consider A is empty.

So, the final solution is between: $A^* \cup J$ or $J$

# Tight relaxations of constrained maximum density problem

Lovasz extensions:

| Set function | Lovasz extension |
|---|---|
| $\mathrm{cut}(A, \overline{A})$ | $\frac{1}{2}\sum_{i,j}^{m} w_{ij}\lvert f_i - f_j\rvert$ |
| $\mathrm{vol}_g(A)$ | $\langle f, (g_i)_{i=1}^{m}\rangle$ |
| $\mathrm{assoc}(A)$ | $\left\langle f, (d_i^{(V\setminus J)})_{i=1}^{m}\right\rangle - \frac{1}{2}\sum_{i,j}^{m} w_{ij}\lvert f_i - f_j\rvert$ |
| $\widehat{P}(A)$ | $f_{\max}$ |
| $\widehat{T}_{k'}(A)$ | $\langle f, (h_i)_{i=1}^{m}\rangle - T_{k'}^{(2)}(f)$ |

## Tight relaxations of constrained maximum density problem

Objective function:

$$\min_{f \in \mathbb{R}_+^m} \frac{R_1(f) - R_2(f)}{S_1(f) - S_2(f)}$$

$$R_1(f) = \langle (g_i)_{i=1}^m + \gamma (h_i)_{i=1}^m, f \rangle + \mathrm{vol}_g(J) f_{\max},$$

$$S_1(f) = \langle (d_i)_{i=1}^m + (d_i^{(J)})_{i=1}^m, f \rangle + \mathrm{assoc}(J)\, f_{\max},$$

$$R_2(f) = \gamma T_{k'}^{(2)}(f)$$

$$S_2(f) = \frac{1}{2} \sum_{i,j}^m w_{ij} |f_i - f_j|$$

## Tight relaxations of constrained maximum density problem

Apply the RatioDCA algorithm, the crucial step is solving the inner problem in line 3.

$$\min_{\substack{f \in \mathbb{R}_+^m \\ \|f\|_2 \le 1}} \left\{ c_1 f_{\max} + \langle f, c_2 \rangle + \lambda^l \frac{1}{2} \sum_{i,j}^{m} w_{ij} |f_i - f_j| \right\}$$

# Tight relaxations of constrained maximum density problem

**Lemma 3** *The inner problem (15) is equivalent to*

$$-\min_{\substack{\|\alpha\|_\infty \le 1 \\ \alpha_{ij} = -\alpha_{ji}}} \min_{v \in S_m} \frac{1}{2} \left\| P_{\mathbb{R}^m_+} \left( -c_1 v - c_2 - \frac{\lambda^l}{2} A\alpha \right) \right\|_2^2$$

*where* $(A\alpha)_i := \sum_j w_{ij}(\alpha_{ij} - \alpha_{ji})$, $P_{\mathbb{R}^m_+}$ *denotes the projection on the positive orthant and* $S_m$ *is the simplex* $S_m = \{v \in \mathbb{R}^m \,|\, v_i \ge 0, \sum_{i=1}^m v_i = 1\}$.

# Tight relaxations of constrained maximum density problem

Algorithm:

**FISTA** for the inner problem

**Input:** Lipschitz constant $L$ of $\nabla\Psi$,

**Initialization:** $t_1 = 1$, $\alpha^1 \in \mathbb{R}^{|E|}$,

**repeat**

$$v = \arg\min_{u \in S_m} \left\| P_{\mathbb{R}^m_+}\left(-c_1 u - c_2 - \frac{\lambda^l}{2} A\alpha\right)\right\|_2^2$$

$$z = P_{\mathbb{R}^m_+}\left(-c_1 v - c_2 - \frac{\lambda^l}{2} A\alpha\right)$$

$$\beta_{rs}^{k+1} = P_{B_\infty(1)}\left(\alpha_{rs}^k + \frac{1}{L}\lambda^l w_{rs}(z_r - z_s)\right)$$

$$t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2},$$

$$\alpha_{rs}^{k+1} = \beta_{rs}^{k+1} + \frac{t_k - 1}{t_{k+1}}\left(\beta_{rs}^{k+1} - \beta_{rs}^k\right).$$

**until** duality gap $< \epsilon$

# Outline

1. Background
    a. Constrained balanced graph cuts for local clustering
    b. Constrained local community detection
2. Tight relaxations
    a. Unconstrained fractional set programs
    b. Constrained fractional set programs
3. Minimization of tight continuous relaxation
4. Tight relaxations of constrained maximum density problem
5. **Experiments**

# Experiments

LS: local spectral method by Mahoney et al.2012.
LRW: Lazy random walk by Andersen 2006.

CFSP: Method proposed in this paper.

Data: Large social networks of the Stanford Large Network Dataset Collection.

# Experiments

*Table 1.* Results for the constrained local normalized cut. Our solutions (CFSP) always satisfy all constraints and have smaller cuts than the two competing methods LS and LRW.

| | Method | $\leq$ 20% | $\leq$ 40% | $\leq$ 60% | $\leq$ 80% | $\leq$ 100% | Runtime |
|---|---|---|---|---|---|---|---|
| CA-GrQc | LRW | 0.1311 (0.0686) | 0.1005 (0.0542) | 0.0984 (0.0543) | 0.0920 (0.0439) | 0.0773 (0.0341) | 2 |
| (4158,13422) | LRW+CFSP | 0.1048 (0.0486) | 0.0695 (0.0318) | 0.0614 (0.0268) | 0.0614 (0.0268) | 0.0457 (0.0217) | 2 + 3 |
| | LS | 0.2014 (0.0958) | 0.1182 (0.0958) | 0.0685 (0.1089) | 0.0314 (0.0423) | 0.0217 (0.0259) | 6 |
| | LS+CFSP | 0.1366 (0.0914) | 0.0709 (0.0592) | 0.0340 (0.0494) | 0.0200 (0.0270) | 0.0147 (0.0120) | 6 + 3 |
| | CFSP | **0.0315 (0.0292)** | **0.0157 (0.0131)** | **0.0138 (0.0115)** | **0.0083 (0.0055)** | **0.0069 (0.0044)** | 31 |
| CA-HepTh | LRW | 0.2607 (0.0914) | 0.2157 (0.0533) | 0.2015 (0.0498) | 0.1954 (0.0491) | 0.1888 (0.0483) | 9 |
| (8638,24806) | LRW+CFSP | 0.2074 (0.1003) | 0.1076 (0.0561) | 0.0976 (0.0452) | 0.0882 (0.0305) | 0.0869 (0.0324) | 9 + 8 |
| | LS | 0.4125 (0.1079) | 0.3439 (0.0631) | 0.3089 (0.0839) | 0.2926 (0.0913) | 0.2778 (0.0923) | 13 |
| | LS+CFSP | 0.3258 (0.1236) | 0.1894 (0.1126) | 0.1274 (0.0986) | 0.0651 (0.0315) | 0.0618 (0.0324) | 13 + 9 |
| | CFSP | **0.0518 (0.0226)** | **0.0327 (0.0104)** | **0.0318 (0.0094)** | **0.0263 (0.0082)** | **0.0104 (0.0038)** | 58 |
| Cit-HepTh | LRW | 0.5052 (0.2208) | 0.4697 (0.2010) | 0.4373 (0.1962) | 0.4067 (0.1998) | 0.3807 (0.2224) | 15 |
| (27400,352021) | LRW+CFSP | **0.3888 (0.2261)** | **0.3249 (0.2072)** | 0.2960 (0.1778) | 0.2528 (0.1689) | 0.2476 (0.1928) | 15 + 368 |
| | LS | 0.5430 (0.2617) | 0.5099 (0.2524) | 0.4737 (0.2586) | 0.4290 (0.2773) | 0.3997 (0.2834) | 175 |
| | LS+CFSP | 0.4496 (0.2848) | 0.3585 (0.2185) | 0.3122 (0.2138) | 0.2074 (0.0814) | 0.1772 (0.0782) | 175 + 190 |
| | CFSP | 0.4693 (0.2676) | 0.3732 (0.2166) | **0.2683 (0.1494)** | **0.1748 (0.0683)** | **0.0752 (0.0233)** | 3704 |
| Cit-HepPh | LRW | 0.1784 (0.0541) | 0.1466 (0.0503) | 0.1234 (0.0256) | 0.1079 (0.0120) | 0.1048 (0.0062) | 19 |
| (34401,420784) | LRW+CFSP | 0.1365 (0.0305) | 0.1132 (0.0201) | 0.1070 (0.0181) | 0.0966 (0.0135) | 0.0948 (0.0052) | 19 + 219 |
| | LS | 0.1720 (0.0055) | 0.1292 (0.0224) | 0.1155 (0.0147) | 0.1107 (0.0062) | 0.1078 (0.0007) | 103 |
| | LS+CFSP | 0.1335 (0.0064) | **0.1064 (0.0114)** | **0.0965 (0.0091)** | 0.0944 (0.0061) | 0.0916 (0.0011) | 103 + 102 |
| | CFSP | **0.1181 (0.0143)** | 0.1127 (0.0101) | 0.1109 (0.0089) | **0.0928 (0.0039)** | **0.0913 (0.0015)** | 2666 |
| amazon0302 | LRW | 0.1768 (0.0833) | 0.1465 (0.0749) | 0.1336 (0.0601) | 0.1221 (0.0504) | 0.1120 (0.0429) | 336 |
| (262111,899792) | LRW+CFSP | 0.1072 (0.0666) | 0.0724 (0.0455) | 0.0577 (0.0419) | 0.0423 (0.0373) | 0.0344 (0.0294) | 336 + 608 |
| | LS | 0.2662 (0.1204) | 0.2496 (0.1155) | 0.2247 (0.1021) | 0.2066 (0.0892) | 0.1946 (0.0840) | 5765 |
| | LS+CFSP | 0.1775 (0.0807) | 0.1248 (0.0643) | 0.0923 (0.0675) | 0.0878 (0.0694) | 0.0641 (0.0435) | 5765 + 458 |
| | CFSP | **0.0194 (0.0063)** | **0.0095 (0.0043)** | **0.0072 (0.0031)** | **0.0056 (0.0024)** | **0.0050 (0.0022)** | 3007 |
| amazon0505 | LRW | 0.2472 (0.1112) | 0.2369 (0.1124) | 0.2249 (0.1132) | 0.2200 (0.1152) | 0.2163 (0.1183) | 210 |
| (410236,2439437) | LRW+CFSP | 0.1058 (0.0833) | 0.0636 (0.0319) | 0.0636 (0.0319) | 0.0636 (0.0319) | 0.0610 (0.0337) | 210 + 2061 |
| | LS | 0.4124 (0.1751) | 0.3704 (0.1864) | 0.3653 (0.1878) | 0.3576 (0.1919) | 0.3529 (0.1956) | 20558 |
| | LS+CFSP | 0.1300 (0.0935) | 0.0903 (0.0545) | 0.0782 (0.0587) | 0.0782 (0.0587) | 0.0782 (0.0587) | 20558 + 2900 |
| | CFSP | **0.0227 (0.0076)** | **0.0116 (0.0089)** | **0.0058 (0.0020)** | **0.0048 (0.0011)** | **0.0047 (0.0008)** | 13171 |

# Experiments

*Table 2.* Constrained local normalized Cheeger cuts of the solutions obtained by our method (note that we optimized the normalized cut) as well as the solutions of Lazy Random Walk (LRW) where we threshold in each step according to the normalized Cheeger cut objective

| | Method | $\leq$ 20% | $\leq$ 40% | $\leq$ 60% | $\leq$ 80% | $\leq$ 100% | Runtime (sec) |
|---|---|---|---|---|---|---|---|
| CA-GrQc | LRW | 0.1298 (0.0677) | 0.0992 (0.0536) | 0.0967 (0.0537) | 0.0894 (0.0418) | 0.0753 (0.0340) | 1 |
| | CFSP | **0.0312 (0.0289)** | **0.0153 (0.0128)** | **0.0133 (0.0110)** | **0.0079 (0.0051)** | **0.0064 (0.0040)** | 31 |
| CA-HepTh | LRW | 0.2601 (0.0911) | 0.2150 (0.0530) | 0.2005 (0.0495) | 0.1941 (0.0488) | 0.1873 (0.0481) | 1 |
| | CFSP | **0.0517 (0.0225)** | **0.0326 (0.0104)** | **0.0317 (0.0093)** | **0.0261 (0.0082)** | **0.0103 (0.0037)** | 58 |
| Cit-HepTh | LRW | 0.4967 (0.2300) | 0.4565 (0.2150) | 0.4179 (0.2174) | 0.3890 (0.2174) | 0.3705 (0.2307) | 10 |
| | CFSP | **0.4673 (0.2690)** | **0.3712 (0.2176)** | **0.2661 (0.1496)** | **0.1681 (0.0706)** | **0.0705 (0.0150)** | 3704 |
| Cit-HepPh | LRW | 0.1574 (0.0497) | 0.1104 (0.0364) | **0.0769 (0.0151)** | 0.0573 (0.0064) | **0.0566 (0.0062)** | 14 |
| | CFSP | **0.1168 (0.0156)** | **0.1067 (0.0138)** | 0.0986 (0.0202) | **0.0500 (0.0098)** | 0.0584 (0.0049) | 2666 |
| amazon0302 | LRW | 0.1768 (0.0833) | 0.1464 (0.0749) | 0.1335 (0.0600) | 0.1220 (0.0503) | 0.1118 (0.0428) | 241 |
| | CFSP | **0.0193 (0.0063)** | **0.0095 (0.0043)** | **0.0072 (0.0031)** | **0.0056 (0.0024)** | **0.0050 (0.0022)** | 3007 |
| amazon0505 | LRW | 0.2472 (0.1111) | 0.2369 (0.1124) | 0.2248 (0.1132) | 0.2200 (0.1152) | 0.2162 (0.1183) | 289 |
| | CFSP | **0.0227 (0.0076)** | **0.0116 (0.0089)** | **0.0058 (0.0020)** | **0.0048 (0.0011)** | **0.0047 (0.0008)** | 13171 |

Thanks