Supplementary material for Global graph kernels using geometric embeddings

Fredrik Johansson FREJOHK@CHALMERS.SE
Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

Vinay Jethava JETHAVA@CHALMERS.SE
Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

Devdatt Dubhashi DUBHASHI@CHALMERS.SE
Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

Chiranjib Bhattacharyya CHIRU@CSA.IISC.ERNET.IN
Indian Institute of Science, Bangalore 560012 Karnataka, India

A. Overview

This document is supplementary material for *Global graph kernels using geometric embeddings*. It is organized as follows. In Section B, we prove Lemmas 1, 2 and 4. In Section C, we prove the Theorems 1 and 2 about sample complexity of the Lovász ϑ kernel and the SVM-ϑ kernel. In Section D, we prove Lemma 3 about the margin of the Lovász ϑ kernel. In Section E, we give and prove a result about the margin of the SVM-ϑ kernel.

B. Proofs of smaller lemmas

We restate and prove the smaller lemmas from the paper.

Lemma 1 (Restated). *Given a graph $G = (V, E)$, with orthogonal representation U, as in (2), for any subset $B \subset V$, with H the subgraph of G induced by B, the following holds,*

$$\vartheta(H) \leq \vartheta_B(G) \leq \vartheta_V(G) = \vartheta(G) \ .$$

Proof. First, we note that

$$\vartheta(H) \leq \vartheta_B(G)$$

holds, by the definition of $\vartheta(G)$ in (2), as $\vartheta(H)$ is the smallest of all orthogonal representations of the subgraph induced by V. Second,

$$\vartheta_B(H) \leq \vartheta_V(G)$$

holds because of (2). This is clear, since $\vartheta_V(G)$ and $\vartheta_B(G)$ concern the same orthogonal representation, and thus adding nodes to B to form V implies adding indices to the maximization in (2). Clearly, the value can not decrease. The last equality holds by definition of ϑ. □

Lemma 2 (Restated). *The Lovász ϑ kernel, as defined in* (3)*, is a positive semi-definite kernel.*

Lemma 4 (Restated). *The* SVM-ϑ *kernel, as defined in* (9)*, is a positive semi-definite kernel.*

Proof of Lemmas 2 and 4. The kernels in (3) and (9) are instances of Haussler's R-convolution kernel (Haussler, 1999). The R-convolution kernel for points $x \in \chi$, each associated with a finite subset χ'_x of a common space χ', and a kernel $k : \chi' \times \chi' \to \mathbb{R}$ is defined as (Haussler, 1999)

$$K(x, y) = \sum_{(x',y') \in \chi'_x \times \chi'_y} k(x', y') \ .$$

Now, the Lovász ϑ kernel, (9) and SVM-ϑ kernel, (9) share a common form

$$K(G, G') = \sum_{B \subseteq V} \sum_{\substack{C \subseteq V' \\ |C|=|B|}} \frac{1}{Z_{|B|}} k(f_B(G), f_C(G'))$$

with $f_B(G) = \vartheta_B(G)$ for the Lovász ϑ kernel and $f_B(G) = \sum_{j \in B} \alpha_j(G)$ for the SVM-ϑ kernel.

In our setting, we set $\chi = \mathcal{G} = \{G^{(1)}, \ldots, G^{(M)}\}$ such that $G^{(m)} = (V^{(m)}, E^{(m)})$. Then, let $\chi' = \mathbb{R}$, and with each $x = G^{(m)}$, we associate a set

$$\chi'_x = \{f_B(G^{(m)}) : B \subseteq V^{(m)}\} \subset \chi'$$

We make a note that χ'_x may be a multiset, but a straight-forward extension of Lemma 1 in (Haussler, 1999) proves that this is also a p.s.d. kernel. □

C. Sample complexity of the ϑ kernels

We prove here Theorems 1 and 2 using a particular multiplicative Chernoff bound.

C.1. Sample complexity of the Lovász ϑ kernel

For convenience we restate the theorem as given in the main paper.

Theorem 1 (Restated). *For graphs of n nodes, each coordinate $\varphi(d)$ of the feature vector of the linear Lovász ϑ kernel can be estimated by $\hat{\varphi}(d)$ such that*

$$Pr\left[\hat{\varphi}(d) \geq (1+\epsilon)\varphi(d)\right] \leq O(1/n)$$
$$Pr\left[\hat{\varphi}(d) \leq (1-\epsilon)\varphi(d)\right] \leq O(1/n)$$

using $s_d = O(n \log n/\varepsilon^2)$ samples.

First, consider the following form of multiplicative Chernoff bound.

Lemma C.1. *Let $X = \frac{1}{s}(X_1 + \cdots + X_s)$, with $X_1, \ldots, X_s$ independent variables such that with probability 1, $0 \leq X_i \leq C$. Then,*

$$\begin{aligned} P[X \geq (1+\varepsilon)\mathbb{E}X] &\leq e^{-\frac{s\mathbb{E}X\varepsilon^2}{3C}} \\ P[X \leq (1-\varepsilon)\mathbb{E}X] &\leq e^{-\frac{s\mathbb{E}X\varepsilon^2}{2C}} . \end{aligned}$$

Proof of Theorem 1. We want to bound the error of a coordinate $\hat{\varphi}(d)$ in the sampled feature vector $\hat{\boldsymbol{\varphi}}$ of the Lovász ϑ kernel for a graph $G = (V, E)$. Recall that

$$\varphi(d) = \frac{1}{Z} \sum_{\substack{B \subseteq V \\ |B|=d}} \vartheta_B(G) .$$

We let

$$X = \hat{\varphi}(d) = \frac{1}{s} \sum_{r=1}^{s} \vartheta_{V_r}(G) ,$$

and thus

$$X_r = \vartheta_{V_r}(G) ,$$

for random subsets $V_r \subseteq V$, $r = 1, \ldots, s$, $|V_r| = d$. In order to use Lemma C.1, we bound $\mathbb{E}X$. We have that $1 \leq \vartheta(G) \leq n$ and together with (1), we get $1 \leq \vartheta_B(G) \leq n$ for $B \subseteq V$. Hence,

$$\mathbb{E}X \geq 1$$

and with $C = n$, we get

$$\begin{aligned} P[X \geq (1+\varepsilon)\mathbb{E}X] &\leq e^{-\frac{s\varepsilon^2}{3n}} \\ P[X \leq (1-\varepsilon)\mathbb{E}X] &\leq e^{-\frac{s\varepsilon^2}{2n}} . \end{aligned}$$

We make a note that $\mathbb{E}X = \varphi(d)$, by linearity of expectation. Choosing $s = Dn \log n/\varepsilon^2$ for some D, we obtain the result. □

C.2. Sample complexity of the SVM-ϑ kernel

We use the same Chernoff bound from the previous section to prove a result on the sample complexity of the SVM-ϑ kernel.

Theorem 2 (Restated) (Sample complexity of the SVM-ϑ kernel)**.** *For graphs of n nodes, each coordinate $\varphi(d)$ of the feature vector of the linear SVM-ϑ kernel can be estimated by $\hat{\varphi}(d)$ such that*

$$Pr\left[\hat{\varphi}(d) \geq (1+\epsilon)\varphi(d)\right] \leq O(1/n)$$
$$Pr\left[\hat{\varphi}(d) \leq (1-\epsilon)\varphi(d)\right] \leq O(1/n)$$

using $s_d = O(n^2 \log n/\varepsilon^2)$ samples.

Proof. To apply Lemma C.1, we choose

$$X_r := \sum_{j \in V_r} \alpha_j$$

and

$$X := \frac{1}{s} \sum_{r=1}^{s} X_r ,$$

for random subsets $V_r \subseteq V$, $r = 1, \ldots, s$, $|V_r| = d$. Now, by definition, because we sample subsets uniformly, and by linearity of expectation,

$$\mathbb{E}X = \varphi(d) = \binom{n}{d}^{-1} \sum_{\substack{B \subseteq V \\ |B|=d}} \sum_{j \in B} \alpha_j$$

and similar to the following section, we note We are now interested in bounding $\mathbb{E}X$ from below. We let α_0 denote the smallest maximum α,

$$\alpha_0 := \min_G \max_i \alpha_i(G)$$

and note that any graph G will have at least one node with this value associated with it. Thus, for any graph If we assume with $j_0 := \arg\max_j \alpha_j$, we have

$$\begin{aligned} \mathbb{E}X &= \binom{n}{d}^{-1} \sum_{\substack{B \subseteq V \\ |B|=d}} \sum_j \alpha_j \\ &\geq \binom{n}{d}^{-1} \sum_{\substack{B \subseteq V \\ |B|=d \\ j_0 \in B}} \alpha_0 \\ &= \binom{n}{d}^{-1} \binom{n-1}{d-1} \alpha_0 \\ &\geq \frac{d\alpha_0}{n} \end{aligned}$$

Now, we can prove that $\alpha_0 \geq 1/n$ if $\rho \geq 1$, with $\rho \geq -\lambda_n(A)$ where A is the adjacency matrix of G, as in (4). Assuming $\alpha_0 < 1/n$, clearly $\sum_i \alpha_i < 1$. We look now att the Karush-Kuhn-Tucker conditions for (8) and thus α_i,

$$\alpha_i + \frac{1}{\rho} \sum_{(i,j)\in E} A_{ij}\alpha_j = 1 + \mu_i$$
$$\text{such that } \ \mu_i\alpha_i = 0,\ \mu_i \geq 0\ .$$

We note that it can not be that $\forall i, \alpha_i = 0$. Thus, $\alpha_0 > 0$. We consider i such that $\alpha_i > 0$ (and then $\mu_i = 0$),

$$\begin{aligned} 1 &= \alpha_i + \frac{1}{\rho} \sum_{(i,j)\in E} A_{ij}\alpha_j \\ &\leq \alpha_i + \frac{1}{\rho} \sum_{j\neq i} \alpha_j \\ &< \frac{1}{n} + \frac{1}{\rho} \end{aligned}$$

We obtain a contradiction for $\rho \geq 1 - \frac{1}{n}$. Thus, for any such ρ, $\alpha_0 \geq \frac{1}{n}$

Further, we note that $\alpha_i \leq 1$ for all $i \in V$ and we can thus choose $C = d$. Thus using $\mathbb{E}X \geq d/n^2$ as derived above, we obtain the desired result by choosing $s = Dn^2 \log n/\varepsilon^2$ for some D. □

D. The margin of the Lovász ϑ kernel

We proceed to prove the following result for the Lovász ϑ kernel.

Lemma 3 (Restated). *There exist, with high probability, $Pr \geq 1 - O(1/n)$, a linear separator in linear Lovász ϑ kernel space, separating $G(n,p)$ and $G(n, 1-p, k)$ graphs, $k = 2t\sqrt{\frac{n(1-p)}{p}}$, where $p(1-p) = \Omega(n^{-1}\log^4 n)$, with margin*

$$\gamma \geq (t-c)\sqrt{\frac{n(1-p)}{p}} - o(\sqrt{n})\ ,$$

for some constant c, and large enough $t \geq 1$.

To help in proving Lemma 3, we restate known results about the value of ϑ for random and planted clique graphs.

Lemma D.1 (Coja-Oghlan (2005); Juhász (1982))**.** *For $G(n,p)$ graphs, for $0 < p \leq 0.99$, and $n > n_0$ for some $n_0 > 0$,*

$$\vartheta(G(n,p)) = \Theta\left(\sqrt{\frac{n(1-p)}{p}}\right)$$

with probability $\geq 1 - O(e^{-n})$.

Lemma D.2 (Jethava et al. (2014); Feige & Krauthgamer (2000))**.** *For $G = G(n, 1-p, k)$, where $p(1-p) = \Omega(n^{-1}\log^4 n)$ and $k = 2t\sqrt{n}\sqrt{\frac{n(1-p)}{p}}(1 + o(1))$ for any $t \geq 1$, with probability $\geq 1 - O(1/n)$,*

$$\vartheta(\bar{G}) = k$$

where $\bar{G}$ denotes the complement graph of G.

Proof (of Lemma 3). We will prove the assertion of Lemma 3 by bounding the margin using the feature vectors φ of the linear Lovász ϑ kernel.

For the linear kernel $k(x,y) = xy$, we can compute the feature vectors φ of the Lovász ϑ kernel explicitely, see (4) by,

$$\varphi(d) = \binom{n}{d}^{-1} \sum_{\substack{B\subseteq V \\ |B|=d}} \vartheta_B$$

for a graph $G = (V,E)$. We denote by $\boldsymbol{\varphi}_R(d)$, and $\boldsymbol{\varphi}_Q(d)$, the feature vectors of $G(n,p)$ and $\bar{G}(n, 1-p, k)$ graphs respectively. Showing that $\boldsymbol{\varphi}_R(d)$, and $\boldsymbol{\varphi}_Q(d)$ are separated by a certain margin gives us our desired result.

A simple lower bound on the margin can be obtained by noting that $\varphi_Q(d) \geq \varphi_R(d)$, leaving out errors. This holds because planting a clique k into $\bar{G}$ corresponds to removing edges from G, while $\bar{G}(n, 1-p)$ is equivalent to $G(n,p)$. Removing edges from a graph can only increase $\vartheta(G)$ and thus $\vartheta_B(G)$ and $\varphi(d)$ increase. Clearly, in this case, $\varphi_Q(d)$ and $\varphi_R(d)$ are linearly separable and we can place a separating line crossing the midpoint between the vectors. We have, with probability $\geq 1 - O(1/n)$,

$$\begin{aligned} 2\gamma &\geq \|\boldsymbol{\varphi}_Q - \boldsymbol{\varphi}_R\|_2 \\ &\geq |\varphi_Q(n) - \varphi_R(n)| \\ &\geq (2t-c)\sqrt{\frac{n(1-p)}{p}} - o(\sqrt{n}) \end{aligned}$$

for some constant c and error terms.

In the second inequality, we consider only the last coordinates of the feature vectors, and it holds because of the triangle inequality. This amounts to comparing only the values ϑ of the entire graphs. The third inequality holds by Lemmas D.1 and D.2. □

We note that we can get a better bound for γ by considering all coordinates of the feature vectors, or equivalently, subgraphs of $d \leq n$ nodes.

E. The margin of the SVM-ϑ kernel

We prove a result similar to that of the previous section, but now for the SVM-ϑ kernel. Following (Jethava et al., 2014) we will talk about *planted clique* graphs although most of the details of the proofs deal with the complement of such graphs. However, in terms of this application, they are equivalent.

Lemma E.1. *There exist, with high probability, $Pr \geq 1 - O(1/n)$, a linear separator in linear SVM-ϑ kernel space, separating $G(n,p)$ and $G(n, 1-p, k)$ graphs, $k = 2t\sqrt{\frac{n(1-p)}{p}}$, where $p(1-p) = \Omega(n^{-1}\log^4 n)$, with margin,*

$$\gamma \geq c \cdot t\sqrt{n} - O\left(\frac{\log n}{p}\right)$$

for some constant c, and large enough $t \geq 1$.

Proof. For the linear kernel $k(x,y) = xy$, we can compute the feature vectors φ of the SVM-ϑ kernel explicitely, similarly to (4) by,

$$\varphi^{(i)}(d) = \binom{n_i}{d}^{-1} \sum_{\substack{B \subseteq V^{(i)} \\ |B| = d}} \sum_{j \in B} \alpha_j^{(i)} \ .$$

Showing that the feature vectors $\boldsymbol{\varphi}_R$ and $\boldsymbol{\varphi}_Q$ of $G(n,p)$ and $\bar{G}(n, 1-p, k)$ are separated by a certain margin gives us our desired result.

We give some results obtained by (Jethava et al., 2014).

For a $G(n,p)$ graphs, we have for p in the regime $np \geq 1$, and with $\delta > 0$, $\rho = 2(1+\delta)\sqrt{np(1-p)}$, with $K = K_{LS}(G)$ as in (4), the optimizers of (8),

$$\beta := \alpha_i = 2(1+\delta)\sqrt{\frac{1-p}{np}} \pm \epsilon_R \quad \text{for all } i \in V \ ,$$

Similarly, we note that for $\bar{G}(n, 1-p, k)$ graphs with $S \subset V$, the planted clique, the following holds for $k = o(n)$, $K = K_{LS}(G)$ as in (4) and $\rho = (1+\delta)(2\sqrt{np(1-p)} + kp)$ with probability $\geq 1 - O(1/n)$,

$$\alpha_i = \begin{cases} \beta_1 \pm \epsilon_Q & \text{for } i \in S \\ \beta_2 \pm \epsilon_Q & \text{for } i \notin S \end{cases}$$

where for small δ,

$$\begin{aligned} \beta_1 &= \frac{(t+1)^2}{(t+1)^2 + \frac{1}{2}\sqrt{\frac{np}{1-p}}\frac{1}{1+\delta}} \ , \\ \beta_2 &= \frac{(t+1)}{(t+1) + \frac{1}{2}\sqrt{\frac{np}{1-p}}\frac{1}{1+\delta}} \ . \end{aligned}$$

Now, we have, for the d:th component of the feature vector $\varphi_R(d)$ of $G(n,p)$, with probability $\geq 1 - O(1/n)$,

$$\begin{aligned} \varphi_R(d) &= \frac{1}{Z} \sum_{\substack{B \subseteq V^{(i)} \\ |B| = d}} \sum_{j \in B} \beta + d \cdot \epsilon_R \\ &= d \cdot \beta + d \cdot \epsilon_R \end{aligned}$$

and for the feature vector $\varphi_Q(d)$ of $\bar{G}(n, 1-p, k)$, with sufficiently large n,

$$\begin{aligned} \varphi_Q(d) &= \frac{1}{Z} \sum_{\substack{B \subseteq V^{(i)} \\ |B| = d}} \sum_{\substack{j \in B \\ j \in S}} \beta_1 + \sum_{\substack{j \in B \\ j \notin S}} \beta_2 + d \cdot \epsilon_Q \\ &= d\left(\frac{k}{n}\beta_1 + \frac{(n-k)}{n}\beta_2\right) + d \cdot \epsilon_Q \end{aligned}$$

This gives us (since the margin between two points is at least the margin in one dimension), (factor 2 since margin from midpoint), with high probability,

$$\begin{aligned} 2\gamma &\geq |\varphi_Q(n) - \varphi_R(n)| - n \cdot (\epsilon_R + \epsilon_Q) \\ &\geq k \cdot \beta_1 + (n-k)\beta_2 - n \cdot \beta - n \cdot (\epsilon_R + \epsilon_Q) \\ &= \frac{k(t+1)^2}{(t+1)^2 + \frac{1}{u}} + \frac{(n-k)(t+1)}{(t+1) + \frac{1}{u}} - O\left(\frac{\log n}{p}\right) \\ &= t\sqrt{n}O(\delta) - O\left(\frac{\log n}{p}\right) \end{aligned}$$

with $u = 2(1+\delta)\sqrt{\frac{1-p}{np}}$ and $n \cdot (\epsilon_R + \epsilon_Q) = O(\log n / p)$ due to (Jethava et al., 2014). The first row holds with high probability since adding edges can only decrease $\vartheta(G)$. The last line holds due to (Jethava et al., 2014, Eq. 36). ☐

References

Coja-Oghlan, Amin. The lovász number of random graphs. *Combinatorics, Probability & Computing*, 14(4):439–465, 2005.

Feige, Uriel and Krauthgamer, Robert. Finding and certifying a large hidden clique in a semirandom graph. *Random Structures & Algorithms*, 16(2):195–208, 2000.

Haussler, David. Convolution kernels on discrete structures. Technical report, University of California at Santa Cruz, 1999.

Jethava, Vinay, Martinsson, Anders, Bhattacharyya, Chiranjib, and Dubhashi, Devdatt. Lovasz theta function, svms and finding dense subgraphs. *Journal of Machine Learning Research*, 14:3495–3536, 2014.

Juhász, Ferenc. The asymptotic behaviour of lovász' theta-function for random graphs. *Combinatorica*, 2(2):153–155, 1982.