

OPTIMAL SAMPLING SETS IN COGRAPHS

Dominique Guillot[†], Alejandro Parada-Mayorga[†], Sebastian Cioaba[†], and Gonzalo R. Arce[‡]

[†]Department of Mathematical Sciences, University of Delaware, Newark, DE, USA.

[‡]Department of Electrical and Computer Engineering, University of Delaware, Newark, DE, USA.

ABSTRACT

In this paper, we calculate the optimal sampling sets for bandlimited signals on cographs. We take into account the tree structure of the cograph to derive closed form results for the uniqueness sets of signals with a given bandwidth. These results do not require expensive spectral decompositions and represent a promising tool for the analysis of signals on graphs that can be approximated by cographs.

Index Terms— Cographs, graph sampling, sampling sets.

1. INTRODUCTION

The interest on the analysis of data defined on graph-like structures has increased dramatically in recent years [1], and different approaches have been proposed in order to deal with the ever growing size of the data [2]. Of particular importance is the *graph signal sampling* problem, where the objective is to identify a subset of nodes of a network that provides the minimum information required for the unique representation and low error reconstruction of a quantity of interest defined on the nodes [1, 2]. While many interesting approaches have been proposed to solve this problem for static signals [3, 4, 5, 6], and some others for time varying signals [7], these techniques are general and do not always exploit the structure of particular classes of graphs. Additionally, the approaches that offer the best results require spectral decompositions which can be, in some cases, computationally expensive [1].

Among these relevant classes of graphs are *cographs*, which can be described as graphs with no path of 4 vertices as an induced subgraph [8, 9]. Cographs have become of interest in applications that range from orthology detection [10, 11] up to community detection in large size graphs [12]. Additionally, cographs have become of relevance as they can often be used to approximate more complex graphs [13].

In this work, we exactly determine all the minimal uniqueness sets for static bandlimited signals on cographs. To this end, we exploit the particular structure of cographs and the way these graphs are built. We take advantage of the tree representation inherited by every cograph to compute the uniqueness sets in a fast and exact way, starting with simple low complexity uniqueness sets. The analysis presented

here leads to concrete and precise results that only require simple and sequential computations, and no spectral decompositions. We remark that these results have promising applications when the graphs of interest can be approximated by cographs. We show with numerical simulations that for different degrees of approximation between a graph and a cograph, the sampling sets calculated with the cograph offers low error reconstructions in comparison with random sampling and the state of the art techniques.

This paper is organized as follows. Section 2 introduces the basic notation and terminology of the paper, including the central concept of uniqueness set and basic definitions about cographs. Section 3 contains our new results about the uniqueness sets of cographs, as well as an algorithm for their calculation. We demonstrate the utility and promising applications of our results in Section 4 through numerical experiments. We conclude with a summary of our findings in Section 5.

2. BACKGROUND AND NOTATION

For any $n \in \mathbb{N}$, we let $[n] := \{1, \dots, n\}$. We identify the vector $\mathbf{f} = (f_1, \dots, f_n)^T \in \mathbb{R}^n$ with the function $\mathbf{f} : [n] \rightarrow \mathbb{R}$, where $\mathbf{f}(i) = f_i$. For any arbitrary $n \times n$ matrix \mathbf{A} , let us represent by $\mathbf{A}_{S,T}$ the submatrix of \mathbf{A} with rows in S and columns in T , where $S, T \subseteq [n]$. Similarly, for $\mathbf{v} \in \mathbb{R}^n$, \mathbf{v}_S denotes the restriction of the vector \mathbf{v} to its entries corresponding to indices in S . We denote the all ones and all zeros matrices of size $m \times n$ by $\mathbf{1}_{m \times n}$ and $\mathbf{0}_{m \times n}$ respectively, and we adopt the convention that $\mathbf{1}_{n \times 1} \equiv \mathbf{1}_n$ and $\mathbf{0}_{n \times 1} \equiv \mathbf{0}_n$.

Definition 2.1. Let U be a subspace of \mathbb{R}^n . We will say that a subset $S \subseteq [n]$ is a *uniqueness set for U* if for all $\mathbf{g}, \mathbf{h} \in U$, the condition $\mathbf{g}(i) = \mathbf{h}(i)$ for all $i \in S$ implies $\mathbf{g} = \mathbf{h}$. A uniqueness set will be said to be *minimal* if it does not contain a uniqueness set as a proper subset.

Hence, $S \subseteq [n]$ is a uniqueness set for U , if the values on S of any signal $\mathbf{f} \in U$ determine the signal on U . The following result provides a useful characterization of uniqueness sets.

Proposition 2.2. Let U be a k -dimensional subspace of \mathbb{R}^n and let $\{w_1, \dots, w_k\} \subseteq \mathbb{R}^n$ be any basis of U . Let \mathbf{W} be the $n \times k$ matrix with columns w_1, \dots, w_k . Then, the subset

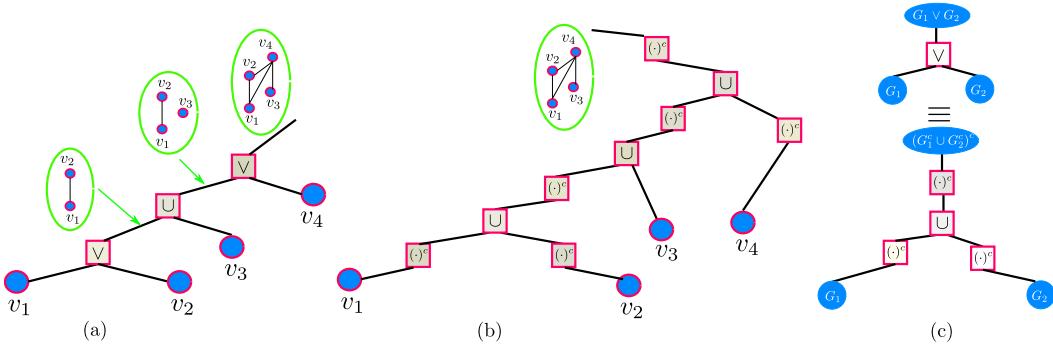


Fig. 1. (a) The cotree representation of the cograph $G = ((\{v_1\} \cup \{v_2\}) \cup \{v_3\}) \cup \{v_4\}$. (b) The cotree representation of the cograph $G = ((\{v_1\} \cup \{v_2\}) \cup \{v_3\}) \cup \{v_4\}$ using complements and unions. (c) A tree representation showing the equivalence $G_1 \vee G_2 = (G_1^c \cup G_2^c)^c$.

$S \subseteq [n]$ is a minimal uniqueness set for U if and only if $|S| = k$ and $\det \mathbf{W}_{S, \{1, \dots, k\}} \neq 0$.

Observe that Proposition 2.2 works for any subspace of \mathbb{R}^n . In particular, it applies to the important case of subspaces built from the eigenspaces of operators defined on a graph [2]. To elaborate, let $G = (V(G), E(G))$ be a simple graph with vertex set $V(G)$ and edge set $E(G)$. We denote by \mathbf{A} the adjacency matrix of G and by $\mathbf{L}_G := \mathbf{D} - \mathbf{A}$ the graph Laplacian of G , where $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$ and d_i is the degree of the i -th vertex of G . Let $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ denote the eigenvalues of \mathbf{L}_G , and let $\mathbf{1}_n = w_1, \dots, w_n \in \mathbb{R}^n$ be an associated orthogonal basis of eigenvectors. We identify signals (functions) $\mathbf{f} : V(G) \rightarrow \mathbb{R}$ with vectors in $\mathbb{R}^{|V(G)|}$. For $\omega \geq 0$ we define the Paley–Wiener space $\text{PW}_\omega(G)$ as $\text{PW}_\omega(G) := \text{span}\{w_i : \lambda_i \leq \omega\}$. Notice that $\text{PW}_\omega(G)$ is a vector space of dimension $\dim \text{PW}_\omega(G) = k$, with k being the largest integer in $[n]$ such that $\lambda_k \leq \omega$. We also define the modified Paley–Wiener space $\text{PW}_\omega^0(G)$ as $\text{PW}_\omega^0(G) := \text{span}\{w_i : i \geq 2 \text{ and } \lambda_i \leq \omega\}$, and the complementary Paley–Wiener space $\text{CPW}_\omega(G)$ as $\text{CPW}_\omega(G) := \text{span}\{w_i : \lambda_i \geq \omega\}$. In this work we adopt the convention of having $\mathbf{1}_n$ as the eigenvector associated to the eigenvalue $\lambda_1 = 0$, and the whole set of eigenvectors is assumed orthogonal. Proposition 2.2 implies the following characterization of the uniqueness sets for $\text{PW}_\omega(G)$.

Corollary 2.3. Let G be a simple graph on n vertices with Laplacian \mathbf{L}_G . Let $W = (w_1, \dots, w_n) \in \mathbb{R}^{n \times n}$ be any matrix whose columns form a basis of eigenvectors associated to the eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ of \mathbf{L}_G . Also, let $\omega \geq 0$ and let k be the largest integer in $[n]$ such that $\lambda_k \leq \omega$. Then a subset $S \subseteq V(G)$ is a minimal uniqueness set for $\text{PW}_\omega(G)$ if and only if $|S| = k = \dim \text{PW}_\omega(G)$ and the matrix $\mathbf{W}_{S, \{1, \dots, k\}}$ is non-singular.

Corollary 2.3 implies that if $\mathbf{f} \in \text{PW}_\omega(G)$ then there is a k dimensional vector α_k such that $\mathbf{f} = \mathbf{W}\alpha_k$. The components of α_k are known as the nonzero Graph Fourier coefficients of \mathbf{f} , and it is said that the bandwidth of \mathbf{f} is ω .

on the spectral axes (or is k on the discrete axes). Then, it is clear that \mathbf{f} has only k degrees of freedom determined by α_k , and the knowledge of \mathbf{f} on a set S of k well chosen nodes is enough to determine \mathbf{f} completely. The selection of such a subset S requires that the system of equations $\mathbf{f}_S = \mathbf{W}_{S, \{1, \dots, k\}}\alpha_k$ should have a unique solution for α_k , which happens when $\mathbf{W}_{S, \{1, \dots, k\}}$ is non-singular, i.e. when S is a uniqueness set.

2.1. Cographs

Let $G = (V(G), E(G))$, $H = (V(H), E(H))$ be two graphs with $V(G) \cap V(H) = \emptyset$. The union of the graphs G and H , denoted by $G \cup H$, is the graph with vertex set $V(G \cup H) = V(G) \cup V(H)$ and edge set $E(G \cup H) = E(G) \cup E(H)$. The join of G and H , denoted $G \vee H$, is the graph with vertex set $V(G \vee H) = V(G) \cup V(H)$ and where $ab \in E(G \vee H)$ if and only if either

- $ab \in E(G)$ or $ab \in E(H)$; or
- $a \in V(G)$, $b \in V(H)$ or $a \in V(H)$ and $b \in V(G)$.

With these definitions, we introduce the notion of a cograph.

Definition 2.4. A cograph (or a complement-reducible graph) is a graph defined recursively as follows:

1. Isolated vertices are cographs;
2. If G and H are cographs on disjoint vertex sets, then so is their join $G \vee H$;
3. If G and H are cographs on disjoint vertex sets, then so is their union $G \cup H$.

An equivalent definition can be given by only working with unions and complements. Recall that the complement of a graph G is the graph $G^c := (V(G), E(G)^c)$. Using the fact that $G \vee H = (G^c \cup H^c)^c$ one can show that a graph is a cograph if and only if it can be obtained from isolated vertices by only performing unions and complements [9]. In the remainder of the paper we only use this representation.

Figure 1 displays the tree representation (or cotree) of a cograph that is built according to Definition 2.4, and its equivalent cotree constructed using unions and complements. As

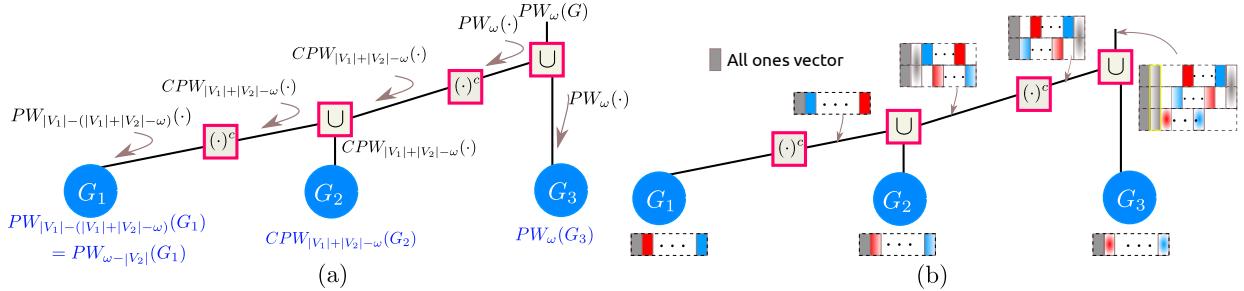


Fig. 2. Cotree representation of the cograph $G = (G_1^c \cup G_2)^c \cup G_3$ that is built from more elementary cographs G_1, G_2 and G_3 . (a) Indication on the cotree of the Paley-Weiner subspaces of G_1, G_2, G_3 involved in the calculation of the uniqueness set of $PW_\omega(G)$, and how the bandwidth changes when *moving* on the cotree. (b) Indication of how the set of eigenvectors associated to G are obtained from G_1, G_2, G_3 and the transformations involved when *moving* on the cotree.

we will indicate later, it is possible to take advantage of this representation to describe how the uniqueness sets of the cograph are related to the uniqueness sets of the building graphs used to generate the cograph.

3. UNIQUENESS SETS OF COGRAPHS

Taking into account that the representation of any cograph can be expressed in terms of unions and complements (see Figure 1), we establish in this section how the uniqueness sets are affected by these operations.

Lemma 3.1. *Let $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$ be two simple graphs, let $G := G_1 \cup G_2$, and let $\omega \geq 0$. Then*

1. $\dim PW_\omega(G) = \dim PW_\omega(G_1) + \dim PW_\omega(G_2)$.
2. S is a minimal uniqueness set for $PW_\omega(G)$ if and only if $S = S_1 \cup S_2$ where S_1 is a minimal uniqueness set for $PW_\omega(G_1)$ and S_2 is a minimal uniqueness set for $PW_\omega(G_2)$.

Proof. The result follows easily from the fact that $\mathbf{L}_G = \mathbf{L}_{G_1} \oplus \mathbf{L}_{G_2}$ and from Corollary 2.3. \square

Similarly, we can easily describe the spectrum of the complement of a graph on N vertices using the fact that $\mathbf{L}_{G^c} = N\mathbf{I}_N - \mathbf{1}_{N \times N} - \mathbf{L}_G$.

Lemma 3.2 (see e.g. [14, Section 1.3.2]). *Let G be a graph with Laplacian eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$ and associated eigenvectors $\mathbf{1}_N = u_1, u_2, \dots, u_N$. Then the complement graph G^c has Laplacian eigenvalues $\{0, N - \lambda_N, \dots, N - \lambda_2\}$, and eigenvectors $\mathbf{1}_N = v_1, v_2 = u_N, \dots, v_N = u_2$.*

Using the above lemma, we can now describe the Paley-Wiener space of the complement of a graph.

Theorem 3.3. *Let $G = (V(G), E(G))$ be a graph with $|V(G)| = N$. Then for any $\omega > 0$, we have $PW_{N-\omega}(G^c) = \mathbf{1}_N \oplus CPW_\omega(G)$ and $CPW_{N-\omega}(G^c) = PW_\omega^0(G)$.*

Proof. Let $0 = \mu_1, \mu_2, \dots, \mu_N$ denote the Laplacian eigenvalues of G^c . We know that $PW_{N-\omega}(G^c)$ is generated by all eigenvectors associated to the eigenvalues $\mu_i \leq N - \omega$, and by Lemma 3.2 the μ_i 's are given by $0, N - \lambda_N, \dots, N - \lambda_1$ with $i > 1$. This implies that $N - \lambda_i \leq N - \omega \Rightarrow \lambda_i \geq \omega$ then $PW_{N-\omega}(G^c) = \mathbf{1}_N \cup CPW_\omega(G)$. Similarly, we have that $CPW_{N-\omega}(G^c)$ is generated by the eigenvectors associated to the eigenvalues $\mu_i \geq N - \omega$ which implies $\lambda_i \leq \omega$. \square

Theorem 3.3 provides a succinct way to compute the uniqueness sets of a cograph via its cotree representation. For example, consider the illustration in Figure 2. Suppose the cograph G is constructed from simple cographs G_1, G_2, G_3 for which the uniqueness sets can be easily computed (e.g. they could be isolated vertices). From the cotree representation, it is possible to identify in a systematic way what are the Paley-Wiener spaces involved in the calculation of the uniqueness sets of G . Fig 2 (a) shows how the bandwidth and the type of Paley-Wiener space change as we go down the tree. Fig 2 (b) then illustrates how the basis for $PW_\omega(G)$ is built from the basis of G_1, G_2 and G_3 by doing simple transformations according to the operations performed in the tree.

3.1. Algorithm

Algorithm 1 summarizes the ideas derived above. In the cotree representation of the cograph we assume each node corresponds to the union of two graphs. These nodes are represented by $c(i)$ where $c(1)$ is the node at the top of the tree and $c(c_T)$ is the node at the bottom of the tree. Complements are indicated on the edges of the cotree (see Fig 2). The operator $PW_\omega(\cdot)$ indicates the calculation of the uniqueness sets for the Paley-Wiener space of bandwidth ω and $CPW_\omega(\cdot)$ indicates the uniqueness sets for the complementary Paley-Wiener space of bandwidth ω . The *descending branches* of each node are denoted by b_ℓ, b_r and $|b|$ denotes the number of nodes of all the elementary cographs G_i descending from the branch b . The symbol $\mathcal{P}(b \rightarrow)$ denotes the application of the operator \mathcal{P} on the branch b after the complement operation.

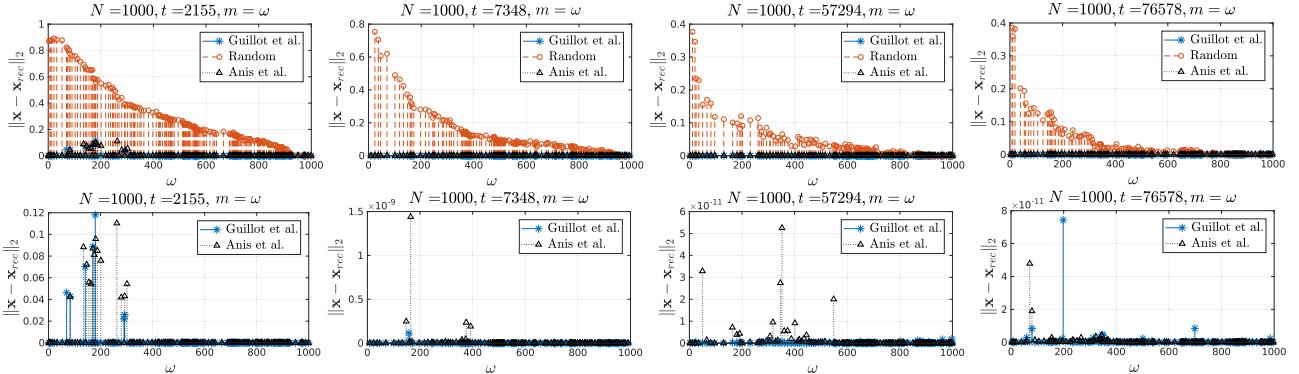


Fig. 3. First row: reconstruction error for random bandlimited signals defined on a graph G with cograph approximation \mathcal{G} . The number t indicates the number of edges that differ between the graph G and \mathcal{G} , whereas $m = \omega$ indicates that the number of samples is equal to the bandwidth of the signal. Second row: same results without the random sampling approach.

Algorithm 1 Uniqueness set for Cographs

Input: Cotree of cograph G, ω .

Output: A minimal uniqueness set \mathcal{U} of $PW_\omega(G)$.

```

Initialisation :  $\mathcal{U} = \emptyset$ 
 $\mathcal{P} = \mathcal{PW}_\omega(\cdot)$ .
2: for  $i = 1 : 1 : c_T$  do
   Get the branches  $b = b_\ell, b_r$  of  $c(i)$ . Then for each  $b$ :
4:   if  $b$  has no complement then
       $\mathcal{U}_b = \mathcal{P}(b)$ .
6:   else
       $\mathcal{P} = \mathcal{CPW}_{|b|-\omega}(\cdot)$ 
       $\mathcal{U}_b = \mathcal{P}(b \rightarrow)$ 
   end if
10:   $\mathcal{U} = \mathcal{U}_{b_\ell} \cup \mathcal{U}_{b_r} \cup \mathcal{U}$ 
   end for
12: return  $\mathcal{U}$ .

```

4. APPLICATIONS

We now provide numerical evidence to demonstrate how the uniqueness sets calculated from a cograph approximation of an arbitrary graph can lead to low reconstruction error, at a very low computational cost. As we show, for a wide range of the percentage of edges that differ between a graph and its cograph approximation, the sampling sets calculated from the cograph lead to competitive error. To assess the performance of our method, we compare its reconstruction error to the error obtained by selecting a sampling set (a) at random; and (b) using the general state-of-the-art method of Anis et al. [3].

For the experiment, a random cograph \mathcal{G} is created and a percentage of edges is altered creating a graph G . The uniqueness sets for $PW_\omega(\mathcal{G})$ are then determined and used as the sampling sets of $PW_\omega(G)$. The values of ω are selected such that they are always located in between two eigenvalues of \mathbf{L}_G . For each $PW_\omega(\mathcal{G})$, 100 signals are selected randomly and then sampled on the sampling sets obtained using Algo-

rithm 1. More specifically, the sampling of a signal \mathbf{f} on the graph G on a subset of nodes $S = \{s_1, \dots, s_m\} \subset V(G)$ is obtained by $\mathbf{f}_S = \mathbf{M}\mathbf{f}$ where \mathbf{M} is a binary matrix whose entries are given by $\mathbf{M} = [\delta_{s_1}, \dots, \delta_{s_m}]^\top$ and δ_v is the N -dimensional Kronecker column vector with center at v . From \mathbf{f}_S , it is possible to obtain a reconstructed version of \mathbf{f} as

$$\mathbf{f}_{rec} = \underset{\mathbf{z} \in \text{span}(\mathbf{U}_\omega)}{\text{argmin}} \|\mathbf{M}\mathbf{z} - \mathbf{f}_S\|_2^2 = \mathbf{U}_\omega (\mathbf{M}\mathbf{U}_\omega)^\dagger \mathbf{f}_S \quad (4.1)$$

where \mathbf{U}_ω is the matrix whose column vectors are the basis of $PW_\omega(G)$, and $(\mathbf{M}\mathbf{U}_\omega)^\dagger$ is the Moore-Penrose pseudo-inverse of $\mathbf{M}\mathbf{U}_\omega$ [15].

As shown in Figure 3, the sampling sets calculated from the cograph approximation lead to very competitive reconstruction errors that are typically smaller than the two competing methods that were considered. This suggests approximating a graph by a cograph provides a viable way to efficiently approximate sampling sets of very large graphs.

5. CONCLUSIONS

We presented a new method to exactly compute all the uniqueness sets of cographs. By exploiting the cotree structure of the cograph, it is possible to determine its uniqueness sets by only performing sequential and simple operations on the uniqueness sets of very simple graphs. Numerical simulations show how a cograph approximation can be exploited to approximate uniqueness sets at very low cost, leading to low error reconstructions for a wide range of approximation degrees. This represents a very promising alternative for the calculation of sampling sets in large size networks. We also note that our method allows one to compute all the uniqueness sets of a cograph in closed form. As a consequence, cographs form an interesting test set of graphs that can be used in theoretical studies to better understand the properties of uniqueness sets, and to provide new insights about them.

6. REFERENCES

- [1] P. Djuric and C. Richard, *Cooperative and Graph Signal Processing: Principles and Applications*, Elsevier Science, 2018.
- [2] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, “Graph Signal Processing: Overview, Challenges and Applications,” *arXiv:1712.00468*, Mar. 2018.
- [3] A. Anis, A. Gadde, and A. Ortega, “Efficient sampling set selection for bandlimited graph signals using graph spectral proxies,” *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3775–3789, July 2016.
- [4] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, “Discrete signal processing on graphs: Sampling theory,” *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6510–6523, Dec 2015.
- [5] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, “Random sampling of bandlimited signals on graphs,” *Applied and Computational Harmonic Analysis*, vol. 44, no. 2, pp. 446 – 475, 2018.
- [6] A. Parada-Mayorga, D.L. Lau, J. Giraldo, and G. Arce, “Blue-Noise Sampling on Graphs,” *arXiv:1811.12542*, Dec. 2018.
- [7] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, “Sampling of graph signals with successive local aggregations,” *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1832–1843, April 2016.
- [8] R. B. Bapat, *Graphs and matrices*, vol. 27, Springer, 2010.
- [9] D.G. Corneil, H. Lerchs, and L.S. Burlingham, “Complement reducible graphs,” *Discrete Applied Mathematics*, vol. 3, no. 3, pp. 163–174, 1981.
- [10] F. Kühnl and M. Lechner, “orthoDeprime: A tool for heuristic cograph editing on estimated orthology graphs,” 2014.
- [11] M. Hellmuth, M. Hernandez-Rosales, K.T. Huber, V. Moulton, P. Stadler, and N. Wieseke, “Orthology relations, symbolic ultrametrics, and cographs,” *Journal of mathematical biology*, vol. 66, pp. 399–420, 03 2013.
- [12] J. Songwei, G. Lin, G. Yong, J. Nastos, Y. Wang, X. Zhang, and H. Wang, “Defining and identifying co-graph communities in complex networks,” *New Journal of Physics*, vol. 17, 01 2015.
- [13] M. Hellmuth, A. Fritz, N. Wieseke, and P. Stadler, “Techniques for the cograph editing problem: Module merge is equivalent to editing P4’s,” 09 2015.
- [14] A. E. Brouwer and W. H. Haemers, *Spectra of Graphs*, Universitext. Springer New York, 2011.
- [15] N. Tremblay, P.O. Amblard, and S. Barthelmé, “Graph sampling with determinantal processes,” in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 1674–1678.