Determinantal Learning for Drone Networks

Xiangliu Tu, Chiranjib Saha, *Member, IEEE*, Harpreet S. Dhillon, *Fellow, IEEE*

***Abstract*—Several key optimization problems in wireless networks can be formulated as *subset selection* problems with the objective of selecting a subset of items from the ground set to optimize an objective function. Recently, a determinantal point process-based learning (DPPL) algorithm, which captures the correlation among items and learns the *quality-similarity* trade-off between them, has emerged as an appealing solution for efficiently solving such problems. A critical component of the DPPL framework is the *similarity matrix*, which models correlation across items and needs to be positive semidefinite (PSD). Popular methods for constructing the similarity matrix, such as cosine similarity and the covariance function, enforce both decomposability and symmetry in the matrix structure. However, these conditions might not always hold in practical scenarios. We overcome this limitation by developing a new way of constructing similarity matrices using the *Gershgorin Circle Theorem*. This is inspired by the application of DPPL to drone cellular networks with directional antennas in which the requirements of symmetry and decomposability are violated, thus necessitating a new method of defining similarity. We rigorously demonstrate the efficacy of the proposed solution by solving link scheduling problems in drone cellular networks.**

I. Introduction

Subset selection problems encompass a large range of resource management problems in wireless networks, such as power control, link scheduling, network utility maximization, and beamformer design. The objective is to identify the optimal subset from a ground set based on a specific objective function. The common strategy is to design heuristic algorithms to find a local optimum with acceptable complexity using a variety of approaches from optimization, such as geometric programming (GP), integer linear or non-linear programming [1]–[4]. However, most of these methodologies are NP-complete, which renders their implementation increasingly challenging as the network size grows. To address the scalability issue, we turn our attention to the determinantal point process (DPP) from stochastic geometry (SG) which has already found applications in some machine learning problems, such as recommender systems and document summarization [5], [6]. The main idea is to view the optimal subset as a realization of a DPP such that items with high quality and low similarity (with each other) are selected. This reduces the subset selection to sampling from a DPP whose parameters need to be trained for a given subset selection problem.

In the wireless community, the DPP has been applied to model and analyze cellular networks [7], [8]. Departing from this direction, we recently proposed to use finite DPPs [9] and the associated learning framework to solve link scheduling problems efficiently in ad hoc networks [10]. A key feature of the DPPL is its ability to capture diversity among items through the similarity model. However, a significant challenge in applying this framework to new settings is that the similarity matrices must be PSD. Common methods used to construct this matrix, such as cosine similarity or covariance function [11], guarantee PSD property at the expense of inducing additional decomposability and symmetry constraints on the matrix, which might not always be satisfied in practice. For instance, in drone networks where base stations (BSs) are equipped with directional antennas, these constraints are clearly violated when considering the similarity among drone-BS links. Inspired by this challenge, we provide a new method to construct a valid similarity model based on the Gershgorin Circle Theorem, which allows the DPPL to handle the specific challenges of directional antennas and capture more complicated correlation structures.

In order to provide a concrete context for our discussion, we will focus our analysis on the link scheduling problem, which is a fundamental and challenging issue in wireless networks. In the literature, this classical problem has been handled through a variety of heuristic algorithms. For instance, integer programming has been used to develop efficient algorithms for handling scheduling problems characterized by the non-convex non-linear objective functions [1], [2]. GP has also been used for a variety of scheduling problems, such as MaxWeighted link scheduling in multihop wireless networks, link activations in multiple-input multiple-output (MIMO) networks, and power/rate allocation [4], [12]–[14]. However, most of these heuristic approaches are NP-complete. Therefore, alternate approaches have long been desirable. As we will discuss shortly, DPPs provide one such approach by treating the optimal solution as a realization of a point process.

In the field of ML, DPPs play a crucial role in a range of problems, such as classification [15], document summarization [6], and recommondation [5]. In such scenarios, DPPs effectively capture the implicit balance between the quality and similarity among items. The authors of [9], [16] have proposed a DPPs-based machine learning framework, which highlights the ability of parameterized DPPs to learn quality and similarity models. This work primarily focuses on DPPs characterized by *L-ensemble*. This formalism requires the matrix L to be PSD, as the probability assigned by the DPPs is proportional to the determinant of the submatrix of L [17]. Common approaches for generating L, such as in [11], incorporate decomposability and symmetry within the similarity model, which may not always hold in practice.

As mentioned above, DPPs have also recently found applications in the optimization of wireless networks [7], [8]. The authors of [18] proposed a novel class of data-driven

X. Tu and H. S. Dhillon are with Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061, USA. Email: {xiangliutu, hdhillon}@vt.edu. C. Saha is with the Qualcomm Standards and Industry Organization, Qualcomm Technologies Inc., San Diego, CA 92121 USA. Email: csaha@qti.qualcomm.com. The support of the US NSF (Grant CNS-1923807) is gratefully acknowledged.

SG models using DPPs, trained to mimic the properties of some hard-core point processes used for wireless network modeling in a finite window. Our work in [10] was the first to introduce the DPPL framework as a solution for general subset selection problems. It demonstrated the scalability of DPPL to find the optimal subsets when applied to large-scale networks. Building on this, the authors of [19] applied this idea to the problem of single-group multicast beamforming. The DPPL was particularly applied to select the bottleneck subset from the multicast group by selecting channels with lower channel gains that are also as orthogonal as possible to each other. While these works provide promising initial steps in the development of the DPPL framework, unnecessary conditions placed on the construction of the similarity matrix (such as symmetry and decomposability) limit the applicability of this approach, which is the main inspiration behind this paper.

Contributions. The main contribution of this paper is to develop a new approach for constructing a similarity model for the DPPL framework by using the Gershgorin Circle Theorem. While the DPPL framework has exhibited its notable efficiency in solving general subset selection problems, current approaches for constructing the similarity matrix introduce unnecessary constraints, which makes it challenging to apply the DPPL framework to new settings. Our proposed generative method effectively overcomes these limitations. We demonstrate the effectiveness of the DPPL framework and the new asymmetric similarity model in solving the link scheduling problems within drone networks where the BS is equipped with directional antennas. Simulation results demonstrate that our approach obtains near-optimal performance by enabling DPPL to capture more complicated correlation structures.

II. Determinantal Point Process

A. Definition of DPPs

To introduce the DPPL framework, we first define DPPs based on finite point processes using the *L-ensemble* formalism [17]. For discrete sets, DPPs are the probability measures over all subsets of a finite ground set $\mathcal{Y} = \{1, \ldots, N\}$. For any set $Y \subseteq \mathcal{Y}$, a DPP defined using a PSD matrix L has the following distribution:

$$\mathcal{P}_L(Y) \equiv \mathcal{P}_L(\mathbf{Y} = Y) = \frac{\det(L_Y)}{\det(L+I)}, \tag{1}$$

where $L_Y = [L_{ij}]_{i,j \in Y}$. If the matrix L is symmetric, it can be represented as a Gram matrix in the form $L = D^T D$, where D denotes a corresponding matrix. For further decomposition, the columns of D can be written as the product of a scalar g and a normalized feature vector ϕ. Consequently, the elements of matrix L can be decomposed as $L_{ij} = g_i \phi_i^\top \phi_j g_j$, where $\|\phi\|^2 = 1$. For $Y = \{i, j\}$, the probability $\mathcal{P}(Y = \{i, j\})$ is proportional to $g_i^2 g_j^2 \left(1 - \phi_i^\top \phi_j \phi_j^\top \phi_i\right)$. This value increases with g (which represents the quality of items) and decreases with $\phi_i^\top \phi_j$ (which represents the similarity between items). For general PSD matrix L, the similarity between items i and j can be quantified as

$$S_{ij} = \frac{L_{ij}}{\sqrt{L_{ii} L_{jj}}}, \forall i, j \in \mathbf{Y}, \tag{2}$$

where the corresponding similarity matrix is denoted by $S_Y = [S_{ij}]_{i,j \in Y}$. The elements of matrix L can be represented as $L_{ij} = g_i S_{ij} g_j$. The probability assigned by DPP to the subset Y is proportional to $\det(L_Y) = \left(\prod_{i \in Y} g_i^2\right) \cdot \det(S_Y)$, which decreases with the similarity and increases with the quality of items in Y. In other words, DPP will favor the selection of subsets in which each element has a high quality (say high signal strength) and low similarity (say low mutual interference). A visual representation of this tradeoff is provided in Fig. 1.

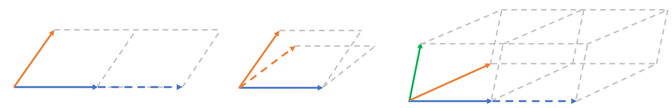

Fig. 1. In DPP, the probability of occurrence of a set Y depends on the volume of the parallelopiped with sides g_i and angles proportional to $S_{i,j}$. Changes in the parallelopiped as we (a) increase the quality of one item, (b) decrease the similarity of times, and (c) increase the quality of one item.

B. Conditional DPPs

We now focus on DPP-based modeling where we assume that the output Y is distributed as a conditional DPP given input X. We denote $\mathcal{Y}(X)$ as the collection of all possible subsets for given X. The conditional DPP assigns probability to every possible subset $Y \subseteq \mathcal{Y}(X)$ based on the matrix L that is parameterized in terms of a generic $\boldsymbol{\theta}$ as

$$\mathcal{P}_{\boldsymbol{\theta}}(\mathbf{Y} = Y \mid X) = \frac{\det(L_Y(X; \boldsymbol{\theta}))}{\det(L(X; \boldsymbol{\theta}) + I)}. \tag{3}$$

Now, assume that we have a sequence of data samples $\left\{\left(X^{(t)}, Y^{(t)}\right)\right\}_{t=1}^{T}$, which are drawn independently from a distribution over pairs $(X, Y) \in \mathcal{X} \times 2^{\mathcal{Y}(X)}$, where $\mathcal{X}$ is the input space. The objective now is to learn an appropriate $\boldsymbol{\theta}$ depending on the input data. The estimation/learning problem can be formulated as

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}; X), \tag{4}$$

where $\mathcal{L}(\boldsymbol{\theta}; X)$ is the log-likelihood function defined as

$$\mathcal{L}(\boldsymbol{\theta}; X) = \log \prod_{t=1}^{T} \mathcal{P}_{\boldsymbol{\theta}}\left(Y^{(t)} \mid X^{(t)}\right) \tag{5}$$

$$= \sum_{t=1}^{T} \Big\{ \log\det\left(L_{Y^{(t)}}\left(X^{(t)}; \boldsymbol{\theta}\right)\right) - \log\det\left(L\left(X^{(t)}; \boldsymbol{\theta}\right) + I\right) \Big\}. \tag{6}$$

If the gradient of the objection function $\mathcal{L}(\boldsymbol{\theta}; X)$ exists and is computable, standard algorithms such as gradient ascent or L-BFGS [20] can be leveraged to find $\hat{\boldsymbol{\theta}}$.

C. Inference using DPPs

We use the following methods for estimating $\hat{Y}$ given X.

1) Sampling from DPP: The first approach involves drawing a random sample from the DPP, i.e., $\mathbf{Y} \sim \mathcal{P}_{\boldsymbol{\theta}^*,\sigma^*}(\cdot|X)$ and set $\hat{Y} = \mathbf{Y}$. This approach begins by drawing a random sample from a specific type of DPP, known as the elementary DPP. A DPP on $\mathcal{Y}$ is called *elementary* if every eigenvalue of its marginal kernel K lies in $\{0, 1\}$. Therefore, an elementary DPP can be represented as $\mathcal{P}^V$ where $V = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is the set of k orthonormal vectors such that $K^V = \sum_{\mathbf{v}\in V} \mathbf{v}\mathbf{v}^\top$. Thus, if $\mathbf{Y} \sim \mathcal{P}^V$, then $|\mathbf{Y}| = |V|$ almost surely. With this, we know that $\mathcal{P}^V(Y) = \mathcal{P}^V(Y \subseteq \mathbf{Y}) = \det(K_Y^V)$. Next, we aim to draw a $k = |V|$ sample Y. Without loss of generality, assume $Y = \{1, 2, 3, \ldots, k\}$. Let $B = \left[\mathbf{v}_1^\top, \ldots, \mathbf{v}_k^\top\right]^\top$. Then, $K^V = BB^\top$ and $\det(K_Y^V) = (\text{Vol}(\{\mathbf{b}_i\}_{i\in Y}))^2$, where $\text{Vol}(\{\mathbf{b}_i\}_{i\in Y})$ is the volume of the parallelepiped spanned by the column vectors $\mathbf{b}_i$ of B. Now, $\det(K_Y^V) = (\text{Vol}(\{\mathbf{b}_i\}_{i\in Y}))^2 = \|\mathbf{b}_1\|^2 \times \left\|\mathbf{b}_2^{(1)}\right\|^2 \times \cdots \times \left\|\mathbf{b}_k^{(1,\ldots,k-1)}\right\|^2$ where $\mathbf{b}_i^{(1)} = \text{Proj}_{\perp b_1}\mathbf{b}_i$ denotes the projection of $\{\mathbf{b}_i\}$ onto the subspace orthogonal to $\mathbf{b}_1$. Thus, the j^{th} step ($j > 1$) of the sampling scheme assuming $y_1 = 1, \ldots, y_{y_{j-1}} = j - 1$ is to select $y_j = j$ with probability proportional to $\left\|\mathbf{b}_j^{(1,\ldots,j-1)}\right\|^2$ and project $\left\{\mathbf{b}_j^{(1,\ldots,j-1)}\right\}$ to the subspace orthogonal to $\mathbf{b}_j^{(1,\ldots,j-1)}$. Therefore, it can be guaranteed that $\mathcal{P}^V(Y) = \det\left(K_Y^V\right)$. Given an eigen decomposition of L, the DPP sampling algorithm discussed thus far is summarized in Alg. 1.

Algorithm 1 Sampling form a DPP

```
procedure SampleDPP(L)
    Eigen decomposition of L: L = \sum_{n=1}^{N} \lambda_n v_n v_n^T
    J = ∅
    for n = 1,...,N do
        J ← J ∪ {n} with probability \lambda_n / (1+\lambda_n)
        V ← {V_n}_{n∈J}
        Y ← ∅
        B = [b_1, ..., b_n] ← V^T
    for 1 to |V| do
        select i from \mathcal{Y} with probability ∝ ||b_i||^2
        Y ← Y ∪ {i}
        b_j ← Proj_{⊥b_i} b_j
    return Y
```

2) MAP inference: Another approach is to obtain the maximum a posteriori (MAP) set, i.e., $\hat{Y} = \arg\max_{Y\subseteq\mathcal{Y}(X)} \mathcal{P}_{\boldsymbol{\theta}^*,\sigma^*(Y|X)}$. However, finding $\hat{Y}$ is an NP-hard problem because of the exponential order search space $Y \subseteq \mathcal{Y}(X)$. That said, [21] offers a possible near-optimal MAP inference scheme for DPPs that will be used in the numerical simulations of this paper.

III. Link Scheduling

To illustrate the capability of the DPPL framework, we apply it to a cellular network in which drones serve as users (UEs). The BS sites are centrally located in a 19-cell hexagonal grid with inter-site distance (ISD) = 500 meters. Each hexagonal cell has three sectors and each sector is served by a different BS. The BSs of three sectors of one cell are co-located at the center. The azimuth direction of the beams of each sector is offset by $120°$. The drones are dropped uniformly and independently at random. Each drone is equipped with an omnidirectional antenna and has a height from the ground uniformly distributed between 1.5 meters to 300 meters. The BS antenna is downtilted at $100°$ and the height of the BS is 25 meters. After dropping the drones, we perform radio-distance-based association, where each drone camps to the sector with minimum pathloss. However, if we schecule one drone per sector on the same time-frequency resources, it may create significant self interference because of line of sight propagation conditions. Therefore, our objective is to implement a DPP-based inter-cell scheduler that selects a subset of simultaneously active BS-drone links at a given resource. We denote the set of BSs as $\Phi_{\rm b}$ and the set of drones that can be co-scheduled as $\Phi_{\rm u}$. Note that $|\Phi_{\rm b}| = |\Phi_{\rm u}|$. The cellular network connectivity scenario explained above can be described as the directed bipartite graph $\mathcal{G} := \{\Phi_{\rm b}, \Phi_{\rm u}, \mathcal{E}\}$, where $\mathcal{E} := \{(t, r)\}$ is the set of BS-drone pairs based on cell association, $t \in \Phi_{\rm b}$ and $r \in \Phi_{\rm u}$. In this case, our objective is to find the optimal subset of co-active links to maximize the sum-rate of the network:

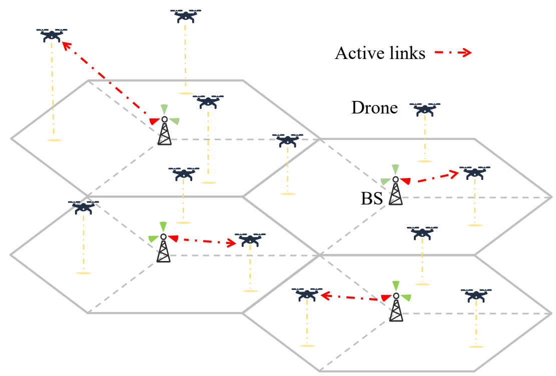

Fig. 2. An illustration of the co-scheduled drones (subset $\mathcal{E}^*$).

$$\text{maximize} \quad \sum_{e_i\in\mathcal{E}} \log_2(1 + \gamma_i), \tag{7a}$$

$$\text{subject to} \quad \gamma_i = \frac{P_i\zeta_{ii}}{\sum_{e_j\in\mathcal{E}}^{j\neq i} P_j\zeta_{ji} + N}, \tag{7b}$$

$$P_i \in \{p_l, p_h\}, \tag{7c}$$

where P_i is the downlink transmit power, N is the noise power, and ζ_{ji} denotes the channel gain between BS j and drones i. Our goal is to find the optimal subset of simultaneously active links denoted as $\mathcal{E}^* \subseteq \mathcal{E}$. In Fig. 2, we illustrate a realization of the active links $\mathcal{E}^*$ in the cellular-connected drone network with 4 three-sector cells.

A. Similarity and Quality Model

In the DPPL framework, the similarity model plays a critical role in modeling correlation among items with the required condition of being PSD. In this specific case of link scheduling, the similarity model is designed to capture the mutual interference between links. The application of directional antennas causes non-symmetric interference patterns

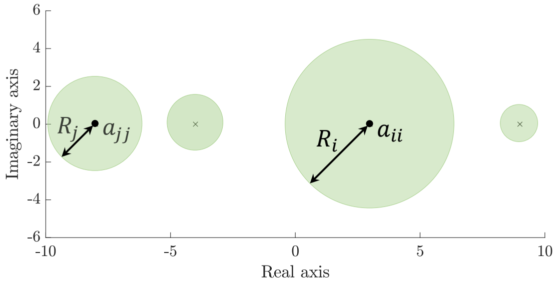

Fig. 3. An Illustration of the Gershgorin discs $D(a_{ii}, R_i(A))$ centered at a_{ii} with radius $R_i(A), i = 1, 2, ..., n$. The green circles are Gershgorin discs that describe the domain for the eigenvalues of matrix A. Specifically, eigenvalues of matrix A will lie within at least one of the Gershgorin discs.

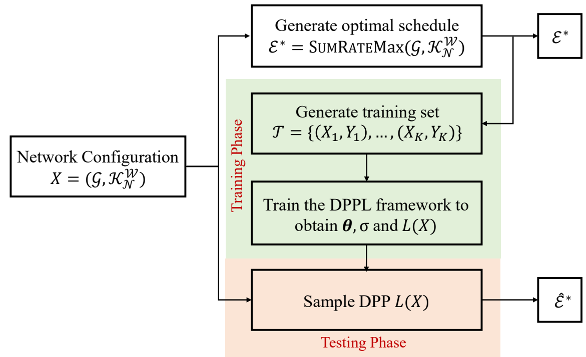

Fig. 4. The block diagram of the DPPL for the link scheduling problem.

among links, leading to an asymmetric similarity matrix. As a result, popular methods of constructing the similarity matrix, such as cosine similarity and the covariance function, are not applicable to this case. This inspired us to develop the following method for constructing more general similarity matrices.

1) Similarity Model: In this work, we use an interference-based parametric similarity model that will try to minimize mutual interference between simultaneously active links. Defining $I_{ij} = P_i \zeta_{ij}$ as the interference caused by BS i to drone j, we define the similarity model as:

$$S(X; \sigma) = \sigma S(X) , \tag{8}$$

where $\sigma \in \mathbb{R}^+$ is a scalar and $S(X)$ is a fixed matrix constructed by I_{ij} which is defined as follows:

$$S_{ij}(X) = \begin{cases} \max(\{R_i(S) \mid i \in \Phi_\text{b}\}) , & i = j \\ I_{ij}, & i \neq j \end{cases} , \tag{9}$$

where $i \in \Phi_\text{b}$ is the index of BS, $j \in \Phi_\text{u}$ is the index of drone, and $R_i(S) = \sum_{j \neq i} |S_{ij}|$. The inclusion of directional antennas implies that I_{ij} does not necessarily equal I_{ji}, indicating asymmetry in S. This asymmetry requires rigorous proof that the matrix L constructed with the above similarity S continues to fulfill the PSD condition. Given quality $g \in \mathbb{R}^+$ and $L_{ij} = g_i S_{ij} g_j$, it follows that similarity matrix S being PSD is the necessary and sufficient condition for L to be PSD. To prove that S is PSD, we first explore the relationship between the elements and the eigenvalues of matrices through the Gershgorin Circle Theorem.

Theorem 1. *Let $A = [a_{ij}] \in \boldsymbol{M}_n$, where $\boldsymbol{M}_n$ is the set of n-by-n matrices. Let $R_i(A) = \sum_{j \neq i} |a_{ij}|, i = 1, 2, ..., n$, denote the absolute row sums of A. The i-th Gershgorin disc is:*

$$D(a_{ii}, R_i(A)) = \{\lambda \in \mathbb{C} : |\lambda - a_{ii}| \leq R_i(A)\} . \tag{10}$$

All eigenvalues of matrix A are in the union of Gershgorin discs, represented as

$$G(A) = \bigcup_{i=1}^{n} \{\lambda \in \mathbb{C} : |\lambda - a_{ii}| \leq R_i(A)\} . \tag{11}$$

Proof: Please see [22, Theorem 6.1.1] for the proof. ■

The Gershgorin Circle Theorem states that every eigenvalue of matrix A lies within at least one of the Gershgorin discs $D(a_{ii}, R_i(A))$, as illustrated in Fig. 3.

For real matrices, Gershgorin discs will reduce to intervals $T_i = [a_{ii} - R_i, a_{ii} + R_i]$. A real matrix (such as S) is PSD if and only if all of its eigenvalues are non-negative. This can be guaranteed by ensuring that all intervals T_i lie on the positive side of the axis, which can be represented by $a_{ii} \geq 0$ and $|a_{ii}| \geq R_i(A)$ for all $i \in \Phi_\text{b}$. We state these conditions formally in the next Proposition.

Proposition 1. *Let $A = [a_{ij}] \in \boldsymbol{M}_n$ is a positive semidefinite matrix if:*

1) $|a_{ii}| \geq R_i(A), i = 1, 2, ..., n,$
2) $a_{ii} \geq 0.$

Now, one can easily construct a PSD matrix using constraints from Proposition 1. Note that under this construction, S neither needs to be symmetric nor decomposable. It is noteworthy that even though our similarity matrix was based on interference in this work, Proposition 1 is more general and can be applied to new settings in which the similarity matrix might depend on other aspects or features of the problem.

2) Quality Model: Coming to the quality, links having a high signal-to-interference-plus-noise ratio (SINR) should naturally be preferred, thus we parameterize the quality model based on SINR as follows:

$$g_i(X; \theta) = \theta \cdot \text{SINR}_i, \tag{12}$$

where SINR_i is the SINR of link i.

B. Training DPPL framework

Now, we implement the proposed DPPL framework to solve the optimal subset selection problem. A sequence of networks and the corresponding optimal subsets obtained by GP serve as the training data sets. Using GP, we obtain approximate solution to (7) by treating transmit power values as continuous variables and then discretizing the resulting optimal values using threashold p_{th}. In the training phase, set $X_k = \left(\mathcal{K}_{\mathcal{N}_t, \mathcal{N}_r}^{\mathcal{W}}, \mathcal{E}, \mathcal{E}^*\right)_k$ represents the k^{th} realization of the network and its optimal subset. The ground set of the DPP is represented as $\mathcal{Y}(X) = \mathcal{E}$. Denote the subset estimated by DPPL in the testing phase as $\hat{\mathcal{E}}^*$. The block diagram of the DPPL framework is presented in Fig. 4.

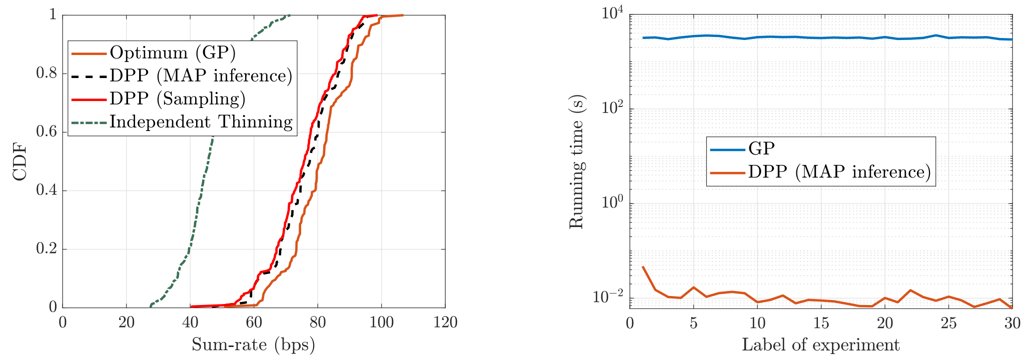

Fig. 5. The numerical simulation results of a cellular network with 19 cells and 57 BS-drone links: (a) comparison of DPP and independent thinning with the optimal sum-rate generated by the GP algorithm, and (b) comparison of the running time of GP and DPP under the same environments.

C. Results and Discussion

Now, we demonstrate the performance of DPPL through numerical simulations. We set $p_h = 46$ dBm, $p_l = 0$ dBm, active threshold $p_{th} = 15$ dBm and use $K = 200$ independent realizations of the network as the training set $\mathcal{T}$. We plot the empirical cumulative distribution functions (CDFs) of the sum-rate obtained from GP, DPPL, and independent thinning in Fig. 5(a). We clearly see that the sum-rate achieved by DPPL is close to the optimal sum-rate.

Further, in order to compare the computational efficiency of DPPL and GP, we select 30 realizations of the network arbitrarily and obtain the optimal schedules using both approaches. The results are presented in Fig. 5(b). It is evident that solving the scheduling problem with GP is roughly 10^5 slower than solving the same problem using DPPL. This is not surprising since GP *solves* the optimization problem, whereas DPPL simply obtains the optimal solution through *sampling*.

IV. Conclusion

Common methods used to construct similarity models limit the applicability of DPPL in scenarios that involve complex asymmetric correlations among items. To address this issue, we developed a new approach based on the Gershgorin Circle Theorem and created an interference-based similarity model for solving three-dimensional link scheduling problems in drone networks. We validated the effectiveness of the new approach through numerical simulations. The results show that this method overcomes limitations of the existing approaches that unnecessarily impose symmetry and decomposability on the similarity matrix, thereby extending the capability of DPPL to capture more general correlation structures.

References

[1] M. W. Cooper, "A survey of methods for pure nonlinear integer programming," *Management Science*, vol. 27, no. 3, pp. 353–361.

[2] S. Burer and A. N. Letchford, "Non-convex mixed-integer nonlinear programming: A survey," *Surveys in Operations Research and Management Science*, vol. 17, no. 2, pp. 97–106, 2012.

[3] M. Chiang, "Geometric programming for communication systems," *Foundations and Trends in Commun. and Info. Theory*, vol. 2, no. 1–2, 2005.

[4] M. Chiang, C. W. Tan, D. P. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Trans. on Wireless Commun.*, vol. 6, no. 7, pp. 2640–2651, Jul. 2007.

[5] M. Wilhelm, A. Ramanathan, A. Bonomo, S. Jain, E. H. Chi, and J. Gillenwater, "Practical diversified recommendations on YouTube with determinantal point processes," in *Proc., ACM Intl. Conf. on Info. and Knowledge Management*, Oct. 2018, pp. 2165–2173.

[6] S. Cho, L. Lebanoff, H. Foroosh, and F. Liu, "Improving the similarity measure of determinantal point processes for extractive multi-document summarization," *arXiv preprint arXiv:1906.00072*, 2019.

[7] Y. Li, F. Baccelli, H. S. Dhillon, and J. G. Andrews, "Statistical modeling and probabilistic analysis of cellular networks with determinantal point processes," *IEEE Trans. on commun.*, vol. 63, no. 9, pp. 3405–3422, Sep. 2015.

[8] N. Miyoshi and T. Shirai, "A cellular network model with Ginibre configured base stations," *Advances in Applied Probability*, vol. 46, no. 3, pp. 832–845, Feb. 2014.

[9] A. Kulesza and B. Taskar, "Determinantal point processes for machine learning," *Foundations and Trends in Machine Learning*, vol. 5, no. 2–3, pp. 123–286, Dec. 2012.

[10] C. Saha and H. S. Dhillon, "Machine learning meets stochastic geometry: Determinantal subset selection for wireless networks," in *Proc., IEEE Globecom*, Dec. 2019.

[11] F. Lavancier, J. Møller, and E. Rubak, "Determinantal point process models and statistical inference," *Journal of the Royal Statistical Society*, Sep. 2015.

[12] P. C. Weeraddana, M. Codreanu, M. Latva-aho, A. Ephremides, and C. Fischione, "Weighted Sum-Rate Maximization in Wireless Networks: A Review," *Foundations and Trends in Networking*, vol. 6, no. 1–2, Oct. 2012.

[13] T. ElBatt and A. Ephremides, "Joint scheduling and power control for wireless ad hoc networks," *IEEE Trans. on Wireless Commun.*, vol. 3, no. 1, pp. 74–85, Jan. 2004.

[14] J. Tang, G. Xue, C. Chandler, and W. Zhang, "Link scheduling with power control for throughput enhancement in multihop wireless networks," *IEEE Trans. on Veh. Technology*, vol. 55, no. 3, pp. 733–742, May 2006.

[15] P. Xie, R. Salakhutdinov, L. Mou, and E. P. Xing, "Deep determinantal point process for large-scale multi-label classification," in *Proc., IEEE Intl. Conf. on Computer Vision (ICCV)*, Oct. 2017, pp. 473–482.

[16] J. A. Kulesza, *Learning with determinantal point processes*. University of Pennsylvania, 2012.

[17] A. Borodin and E. M. Rains, "Eynard–Mehta theorem, Schur process, and their Pfaffian analogs," *Journal of Statistical Physics*, vol. 121, no. 3, pp. 291–317, Nov. 2005.

[18] B. Blaszczyszyn and H. P. Keeler, "Determinantal thinning of point processes with network learning applications," in *Proc., IEEE Wireless Commun. and Networking Conf. (WCNC)*. IEEE, Oct. 2019.

[19] L. Liu, Y. Wang, C. Hua, and J. Jian, "A learning approach for efficient multicast beamforming based on determinantal point process," *IEEE Trans. on Wireless Commun.*, vol. 21, no. 9, pp. 7427–7442, Sep. 2022.

[20] J. Nocedal, "Updating quasi-newton matrices with limited storage," *Mathematics of Computation*, vol. 35, no. 151, pp. 773–782, 1980.

[21] J. Gillenwater, A. Kulesza, and B. Taskar, "Near-optimal map inference for determinantal point processes," *Advances in Neural Information Processing Systems*, vol. 25, Dec. 2012.

[22] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 2012.