Inference and Generating Method for Extremely Sparse Networks

Anonymous Author(s)
Affiliation
Address
email

Abstract

Generative models for real-world networks face a fundamental challenge in reconciling the desirable property of exchangeability with the empirical observation of sparsity. The Caron-Fox framework, which leverages Kallenberg exchangeability and Completely Random Measures (CRMs), provides a principled approach to this problem. However, existing models within this class typically generate graphs where the number of edges scales super-linearly with the number of nodes ($E \propto N^{1+\epsilon}$). In this work, we present a novel CRM with rapid variation that, when integrated into the Caron-Fox model, generates graph sequences in an extremely sparse regime. We also derive a posterior inference method to fit our model to an observed graph. This workshop paper summarizes the key results of our recent publication, introducing the model, outlining its theoretical underpinnings, and presenting the inference procedure.

1 Introduction

The proliferation of large-scale, graph-structured data across diverse scientific fields, from microbiology to the social sciences, has revealed a fascinating truth: despite their varied origins, real-world networks consistently exhibit a set of common structural properties. A key characteristic is *sparsity*, where the number of existing edges is a mere fraction of the $\binom{N}{2}$ potential connections in a graph with N nodes. Another is the *scale-free phenomenon*, in which the network's degree distribution presents a heavy tail, signifying the presence of a few highly connected "hub" nodes alongside a vast majority of nodes with very few links. Additional defining features include the *small-world phenomenon*[1] and the formation of distinct *community structures*.

Developing generative models that can faithfully reproduce these characteristics, especially sparsity, presents a fundamental statistical dilemma. An intuitive and mathematically convenient property for such a model is *node-exchangeability*, which dictates that the graph's probability distribution should remain *invariant* regardless of how its nodes are labeled. However, this very property is fundamentally at odds with sparsity. The well-known Aldous-Hoover theorem [1, 2] formalizes this conflict, demonstrating that any node-exchangeable graph must fall into one of two extremes: it is either dense (almost all possible edges exist) or empty, failing to capture the sparse nature of real-world networks.

To navigate this challenge, the research community has pursued several ideas. One popular approach, exemplified by preferential attachment models [3, 4], resolves the issue by completely forgoing any form of exchangeability. A more recent and powerful line of inquiry, initiated by Caron and Fox [5], circumvents the limitation in a more nuanced way. Instead of abandoning exchangeability altogether, this framework adopts a weaker, more flexible notion known as *Kallenberg exchangeability* (see also

[1]The small-world phenomenon is the observation that, in many real-world networks, any two nodes can be connected through a short chain of intermediary links.

[6, 7, 8]). This principled relaxation successfully reconciles the model with sparsity and has spurred a rich body of work, leading to new models that incorporate other complex features like overlapping communities [9], clustering [10], dynamic evolution [11], and core-periphery organization [12].

Our work is situated within this latter framework, leveraging Bayesian nonparametric methods, specifically Completely Random Measures (CRMs). Existing models in this class have successfully generated sparse graph sequences where the number of edges, E, scales as $N^{1+\epsilon}$ for a given number of nodes N (where $0 < \epsilon < 1$). In our recent publication [13], we introduced a novel CRM that achieves a significantly sparser regime. Our model generates graphs where the edge count scales as $E = \Theta(N\ell(N))$, with ℓ being a slowly varying function. We term graph sequences exhibiting this near-linear scaling *extremely sparse*.

The statistical properties of our model make it possible to derive a posterior inference method that can be used to fit our model to an observed graph. The performance of this hierarchical HMC method is illustrated with experiments on both synthetic and large-scale real datasets.

The purpose of this workshop paper is to present the main results and implications of this new model as detailed in [13]. The remainder is structured as follows. In Section 2, we introduce necessary preliminaries on the Caron-Fox model. In Section 3, we present the main contribution of [13]—a novel CRM with rapid variation that generates extremely sparse sequences of networks within the Caron-Fox model, and explain how to sample from them. In Section 4, we outline the inference procedure. The results are illustrated with experiments throughout the paper, which can be reproduced using the code available at `https://anonymous.4open.science/r/rapidly_varying_crm-7COB/README.md`. Detailed proofs and additional details can be found in the main article [13].

2 The Caron-Fox model

The Caron-Fox model [5, 7, 8] provides a powerful framework for overcoming the limitations of the Aldous-Hoover theorem [1, 2]. Instead of representing graphs with exchangeable arrays, this model uses **exchangeable measures**. In this representation, nodes are embedded at locations $\theta_i \in \mathbb{R}$, and the set of edges is described by a random measure on the plane:

$$Z = \sum_{i,j} Z_{i,j}\delta_{(\theta_i,\theta_j)},$$

where $Z_{i,j} = Z_{j,i}$ is a binary variable indicating the presence or absence of an edge between nodes θ_i and θ_j. A sequence of finite graphs $(\mathcal{G}_t)_{t>0}$ is derived by restricting the point process Z to the squares $[0,t]^2$ and retaining only the points that have at least one edge. In the following, we denote N_t as the number of nodes in $\mathcal{G}_t$ and $N_t^{(e)}$ as the number of edges in $\mathcal{G}_t$.

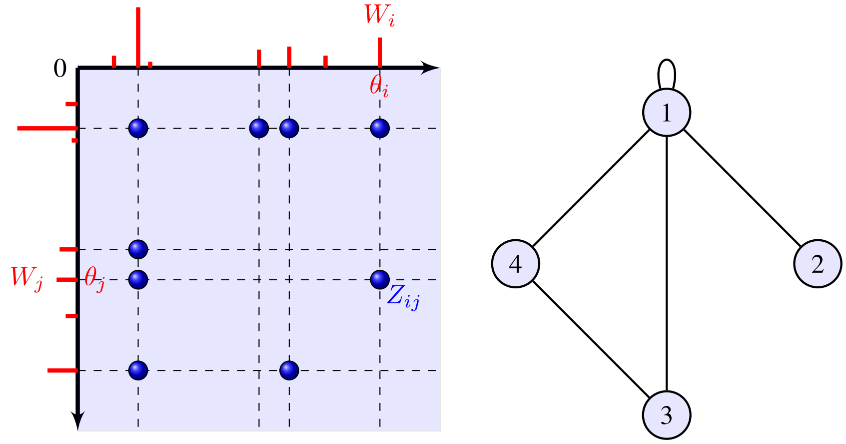
This model was introduced in [5], where it was shown that sparse exchangeable graphs can be obtained in this framework by treating Z as an exchangeable random measure. In [7], the authors introduced the term *graphex processes*, which we use here. Simultaneously, [7] and [8] analysed properties of graphex processes, explaining how graphex processes serve as a good notion of limit for sequences of sparse graphs, analogous to how graphons serve as a limit for dense graphs. Further properties, including convergence results for sparse graph sequences and sampling methods, were derived by [14, 15, 16] and [8]. In [10], the asymptotic properties of graphex processes were investigated.

Following the notation of [10] (which aligns with that of [7], excluding terms corresponding to stars and isolated edges), we can parameterize the graph with a symmetric measurable function $W : \mathbb{R}_+^2 \to [0,1]$ that dictates how the edges are formed. For each $i < j$:

$$Z_{i,j} \mid (\theta_k, \vartheta_k)_{k=1,2,\dots} \sim \text{Bernoulli}(W(\vartheta_i, \vartheta_j)),$$

where $(\theta_k, \vartheta_k)_{k=1,2,\dots}$ is a unit-rate Poisson process on $\mathbb{R}_+^2$.

In this framework, the **Caron-Fox model** is defined by:



(a) (b)

Figure 1: Point process representation of a random graph. Each node i is embedded in $\mathbb{R}_+$ at some location θ_i and is associated with a sociability parameter W_i. An edge between nodes θ_i and θ_j is represented by a point at locations (θ_i, θ_j) and (θ_j, θ_i) in $\mathbb{R}_+^2$ (Figure 1, 5). The corresponding graph is plotted in (b).

$$W(x, y) = \begin{cases} 1 - e^{-2\bar{\rho}^{-1}(x)\bar{\rho}^{-1}(y)} & \text{if } x \neq y, \\ 1 - e^{-(\bar{\rho}^{-1}(x))^2} & \text{if } x = y, \end{cases}$$

where ρ is a **Lévy measure** such that $\int_0^\infty \min(1, x)\nu(dx) < \infty$ and $\bar{\rho}(x) := \int_x^\infty \rho(w)\, dw$ is the corresponding tail Lévy intensity. The Lévy measure ρ controls the properties of the produced network.

To demonstrate that a sparse regime is achievable within this framework, [5] use a Generalized Gamma measure (GG):

$$\rho(dw) = \frac{1}{\Gamma(1-\sigma)} w^{-1-\sigma} e^{-\beta w}\, dw.$$

They prove that if $\sigma < 0$, an increasing sequence of graphs sampled from this model is dense, while if $\sigma \geq 0$ and $\beta > 0$, such a sequence is sparse. They also provide a posterior inference algorithm for fitting the GG-Caron-Fox model to real data, demonstrating that this model is highly effective. In [10], further exploration of the graphex model revealed four asymptotic regimes: a dense regime, a sparse almost-dense regime, a sparse regime exhibiting power-law behaviour, and an almost extremely sparse regime.

The parameters of the graphex model are easily interpretable. In the Caron-Fox framework, the quantity $W_i = \rho^{-1}(\vartheta_i)$ is often interpreted as a measure of sociability. This interpretation arises from viewing the model as a scenario where potential node i enters a room at time θ_i and attempts to link with the nodes already present. The greater the sociability of both individuals, the more likely they are to establish a connection.

3 Mixture of Generalized Gamma Process and extremely sparse graphs

In [13], we define a Lévy intensity that realizes the extremely sparse regime in the Caron-Fox class and propose a corresponding approximate sampling method. We achieve this regime using a mixture of Generalized Gamma (GG) processes with the following Lévy intensity:

$$\rho_{\mathrm{mGG}}(w; \beta, c, \eta) = \eta \int_0^1 \frac{s c^s}{\Gamma(1-s)} w^{-1-s} e^{-\beta w / c} \mathrm{d}s, \tag{1}$$

The model parameters have the following interpretations:

- β is the exponential tilting parameter; if $\beta > 0$, it tunes the exponential decay of large weights.
- c is a scaling parameter: if $G \sim \mathrm{CRM}(\rho_{\mathrm{mGG}}(\cdot;\beta,1,\eta),H)$, then $cG \sim \mathrm{CRM}(\rho_{\mathrm{mGG}}(\cdot;\beta,c,\eta),H)$.
- η is a rate parameter: if $G_1 \sim \mathrm{CRM}(\rho_{\mathrm{mGG}}(\cdot;\beta,c,\eta_1),H)$ and $G_2 \sim \mathrm{CRM}(\rho_{\mathrm{mGG}}(\cdot;\beta,c,\eta_2),H)$, then $G_1+G_2 \sim \mathrm{CRM}(\rho_{\mathrm{mGG}}(\cdot;\beta,c,\eta_1+\eta_2),H)$.

A key advantage of this Lévy measure is that its Laplace exponent admits an analytic expression:

$$\psi_{\mathrm{mGG}}(t;\beta,c,\eta) := \int_0^\infty \left(1-e^{-tw}\right)\rho_{\mathrm{mGG}}(w)dw = \eta\left(\psi(\beta+ct)-\psi(\beta)\right), \tag{2}$$

where $\psi(t) = \frac{t-1}{\log t}$ for $0 < t < 1$, with $\psi(0) = 0$ and $\psi(1) = 1$.

As shown in [13, Corollary 7], graphs generated using this Lévy intensity, ρ_{mGG}, are extremely sparse:

Theorem 1. *[13, Corollary 7] Almost surely as t goes to infinity*

$$N_t^{(e)} = \Theta(N_t \log(N_t)).$$

We also obtain an interesting result for the degree distribution of these graphs. For $j \in \mathbb{N}$ and $t \in \mathbb{R}$, let $N_{t,j}$ denote the number of nodes with degree j in $\mathcal{G}_t$. In the following theorem, C represents a positive constant.

Theorem 2. *[13, Proposition 8] Almost surely as t goes to infinity*

$$N_{t,j} \sim E(N_{t,j}) \sim \begin{cases} t^2 \frac{C}{\log(t)} & j=1 \\ \frac{t^2}{j(j-1)} \frac{C}{\log^2(t)} & j \geq 2 \end{cases}$$

Let $\tilde{N}_{t,2} = \sum_{k\geq 2} N_{t,k}$ denote the number of nodes with a degree of at least 2. Then, for all $j \geq 2$, we have

$$\frac{N_{t,j}}{\tilde{N}_{t,2}} \underset{t\to\infty}{\longrightarrow} \frac{1}{j(j-1)}.$$

This result implies that the degree distribution for nodes with a degree of at least 2 follows a power law with an exponent of 2.

A key property expected of a tractable statistical model is the ability to sample from it. However, since CRMs are infinite-dimensional objects that cannot be fully represented on a computer, only approximate sampling is possible. Several approximation schemes for CRM sampling exist. In [13], we propose a size-biased method, which is advantageous due to its straightforward implementation:

Proposition 1. *[13, Proposition 3] Let $\xi_1, \xi_2, \ldots$ be the ordered points of a unit-rate Poisson process on $(0,\infty)$; that is, ξ_1, $\xi_2-\xi_1$, $\xi_3-\xi_2$, ... are iid unit-rate exponential random variable. We have the following size-biased construction for $G \sim \mathrm{CRM}(\rho_{\mathrm{mGG}}(\cdot;\beta,c,\eta),H)$. For $j \geq 1$,*

$$\begin{aligned} T_j &= \psi^{-1}\left(\frac{\xi_j}{\eta}+\psi(\beta)\right)-\beta, \\ S_j \mid \{T_j=t\} &\sim p(s|t) \propto s(t+\beta)^s \mathbf{1}_{\{s\in(0,1)\}}, \\ W_j'|\{T_j=t,S_j=s\} &\sim \mathrm{Gamma}(1-s,t+\beta), \\ W_j &= cW_j', \end{aligned}$$

Each step in this sampling procedure can be performed using only standard functions. The asymptotic L_1 error of this size-biased approximation is explored in [13, Proposition 4]. With the result from Proposition 1, we can generate graphs. In Figure 2, we compare our method to the non-mixture Generalized Gamma model of Caron and Fox [5] and the Barabási–Albert model [3]. All three models are known to produce sparse graph sequences that exhibit power-law degree distributions. The expected asymptotic behavior of our model is given in Theorem 1 and Theorem 2. For the Generalized

Gamma model (here with $\alpha = 0.5$), the number of edges scales asymptotically as $N_t^{(e)} \sim N_t^{\frac{4}{3}}$, and its degree distribution follows a power law with an exponent of 1.5. For the Barabási–Albert model, the number of edges is linear in the number of nodes, and its asymptotic degree distribution follows a power law with an exponent of 3.

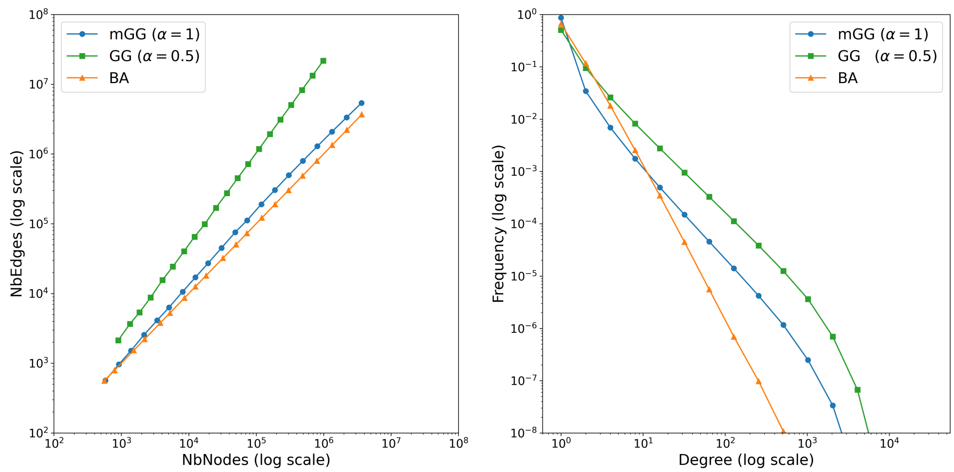


Figure 2: Examination of the mGG graph properties (●) with parameters $\alpha = 1$, $\tau = 0$, $c = 1$, and $\beta = 1$ for various values of η ranging from 50 to 6000, resulting in graphs of different sizes. Comparison with the Generalized Gamma CRM (■) with parameters $\tau = 1$ and $\sigma = 0.5$, and with the Barabási–Albert model (▲). For every configuration we simulate 20 graph samples and plot the mean of the quantity of interest.

4 Inference Algorithm

Posterior inference is a crucial property of a statistical model, as it allows the model to be fit to observational data. Here, we briefly describe the MCMC algorithm introduced in [13] to approximate the posterior density over the model parameters. Assume we have observed a graph $\mathcal{G}$ sampled from our mGG graph model. We aim to infer the sociability parameter W_i for each node and the model parameters $\phi = (\beta, c, \eta)$. We consider the following improper priors:

$$p(\beta) \propto \frac{1}{\beta}, \;\; p(c) \propto \frac{1}{c}, \;\; p(\eta) \propto \frac{1}{\eta}. \tag{3}$$

We introduce two sets of auxiliary variables. First, as in [5], we introduce latent count variables $\tilde{q}_{ij}$ with the conditional distribution, for $i < j$,

$$\tilde{q}_{ij} \mid Z, W \sim \begin{cases} \delta_0 & \text{if } Z_{ij} = 0, \\ \text{tPoisson}\left(2W_iW_j\right) & \text{if } Z_{ij} = 1, i \neq j, \\ \text{tPoisson}\left(W_i^2\right) & \text{if } Z_{ij} = 1, i = j, \end{cases} \tag{4}$$

where $\text{tPoisson}(\lambda)$ is the zero-truncated Poisson distribution, and with $\tilde{q}_{ji} = \tilde{q}_{ij}$. Second, to take advantage of the mixture representation of the Lévy intensity, we introduce local indices of variations $S_i \in (0, 1)$ for each node i, as in the size-biased algorithm. In the following, we also write w_* for the sum of the sociabilities of nodes with no connections in $\mathcal{G}$. The algorithm is presented in Algorithm 1.

Algorithm 1 Posterior inference [13, Algorithm 1]

Step 1: Update the weights w_i and the latent space variables s_i given the other variables, using a Hamiltonian Monte Carlo step.
Step 2: Update (ϕ, w_*) given the other variables, using a Gibbs sampler update.
Step 3: Update the latent counts $(\tilde{q}_{ij})$ given the other variables using (4).

To illustrate the performance of the inference algorithm, [13] proposes different experiments on both synthetic and real data. We present the results on one of the real datasets here. We consider the **Flickr** dataset, a crawl of the Flickr social network, a photo and video sharing platform. The dataset contains all links between users (`https://socialnetworks.mpi-sws.org/data-imc2007.html`, 17). This dataset consists of a network with 1,861,232 nodes and 15,555,041 edges; the mean degree is 33.4, while the maximum degree is 54,472.

To train our model, we extracted a subgraph containing approximately 5% of the nodes from the network using p-sampling [16]. For Flickr, we used $p = 0.14$, resulting in a random subgraph with 85,613 nodes and 299,358 edges, with a mean degree of 13.9 and a maximum degree of 2890.

To assess model fit, we apply p-sampling to split the dataset into a training set (approximately 5% of the nodes) and a test set (the remaining nodes). We then compare the empirical degree distribution of the test set with the predictive degree distribution generated by our model, using hyperparameters inferred from the training data. Specifically, we simulate 200 graphs from the posterior predictive distribution. Parameters are drawn from the post-burn-in MCMC samples, and we scale $\hat{\eta}$ by $(1-p)/p$ to match the expected size of the target graph.

Figure 3 shows the comparison between the empirical and predictive degree distributions. Overall, the model provides a good fit for this dataset, successfully capturing the heavy-tailed behavior characteristic of real-world networks.

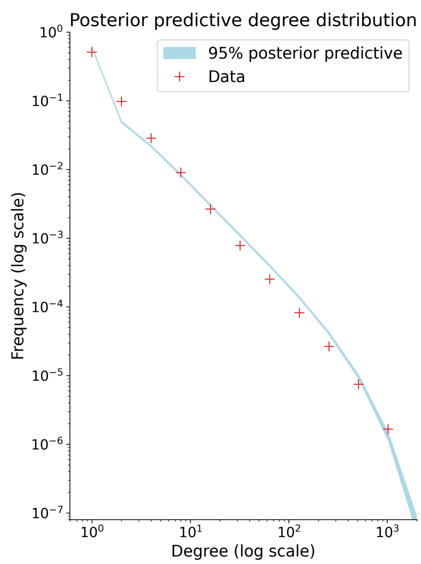


Figure 3: Posterior predictive degree distribution for the Flickr dataset. The predictive distribution is shown in blue, and the empirical distribution of the test set is shown in red.

5 Conclusion

We have seen that [13] has introduced a novel generative model for the difficult problem of generating extremely sparse sequences of graphs with good statistical properties. The model presents theoretical guarantees, a simple size-biased sampling procedure, and an efficient inference algorithm. The new CRM introduced to produce this model has a wider range of applications and could also be used in clustering or partition models. For a more detailed discussion, we refer the reader to the final section of [13].

References

[1] David J. Aldous. Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4):581–598, December 1981. ISSN 0047-259X. doi: 10.1016/0047-259X(81)90099-3.

[2] Douglas N. Hoover. Relations on Probability Spaces and Arrays of Random Variables, 1979.

[3] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[4] Béla Bollobás. *Random Graphs*. Cambridge University Press, August 2001. ISBN 978-0-521-79722-1.

[5] François Caron and Emily B. Fox. Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 79:1295–1366, 2017.

[6] Olav Kallenberg. *Probabilistic Symmetries and Invariance Principles*. Springer Science & Business Media, December 2005. ISBN 978-0-387-28861-1.

[7] Victor Veitch and Daniel M. Roy. The Class of Random Graphs Arising from Exchangeable Random Measures. *arXiv:1512.03099*, December 2015.

[8] Christian Borgs, Jennifer T Chayes, Henry Cohn, and Nina Holden. Sparse exchangeable graphs and their limits via graphon processes. *Journal of Machine Learning Research*, 18(210):1–71, 2018.

[9] Adrien Todeschini, Xenia Miscouridou, and François Caron. Exchangeable random measures for sparse and modular graphs with overlapping communities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(2):487–520, 2020. ISSN 1467-9868. doi: 10.1111/rssb.12363.

[10] François Caron, Francesca Panero, and Judith Rousseau. On sparsity, power-law, and clustering properties of graphex processes. *Advances in Applied Probability*, 55(4):1211–1253, December 2023. ISSN 0001-8678, 1475-6064. doi: 10.1017/apr.2022.75.

[11] Cian Naik, François Caron, Judith Rousseau, Yee Whye Teh, and Konstantina Palla. Bayesian nonparametrics for sparse dynamic networks. *Machine Learning and Knowledge Discovery in Databases*, pages 191–206, 2023.

[12] Cian Naik, François Caron, and Judith Rousseau. Sparse networks with core-periphery structure. *Electronic Journal of Statistics*, 15(1):1814–1868, January 2021. ISSN 1935-7524, 1935-7524. doi: 10.1214/21-EJS1819.

[13] Anonymous. Rapidly Varying Completely Random Measures for Modeling Extremely Sparse Networks. *arXiv:2505.13206*, May 2025.

[14] Svante Janson. On convergence for graphexes. *arXiv:1702.06389*, March 2022.

[15] Svante Janson. Graphons, cut norm and distance, couplings and rearrangements. *arXiv:1009.2376*, June 2011.

[16] Victor Veitch and Daniel M. Roy. Sampling and estimation for (sparse) exchangeable graphs. *The Annals of Statistics*, 47(6):3274 – 3299, December 2019. ISSN 0090-5364. doi: 10.1214/18-AOS1778.

[17] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, pages 29–42, San Diego California USA, October 2007. ACM. ISBN 978-1-59593-908-1. doi: 10.1145/1298306.1298311.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction accurately reflect the paper's actual contributions.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitations are discussed in our Conclusion section.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: Proofs of all results are included in full detail in the main article [13], which is cited at multiple time in the paper, pointing out the relevant section in the main article.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: As stated in the introduction, all the code used to perform our experiments is available in an anonymous repository.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: As stated in the introduction, all the code used to perform our experiments is available in an anonymous repository.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: All details necessary to understand the results are provided in the main paper [13].

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: For all experiments, results are averaged over several repetitions. Variations from the mean are negligible.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: All this information are available in the main paper [13].

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conducted in the paper is mainly theoretical and uses only publicly available datasets, which do not contain any sensitive information.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: The work performed is mainly theoretical, and we do not foresee any societal impact.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The work performed is mainly theoretical and doesn't pose such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: We do not use any dataset.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: We do not release such assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.