Predicting similarity of random graphs for QAOA MaxCut

Eesh Gupta[1]
Department of Physics and Astronomy, Rutgers University, New Brunswick, New Jersey

(Dated: 18 November 2022)

Hybrid quantum-classical algorithms like Quantum Approximate Optimization Algorithm (QAOA) have great potential for demonstrating quantum advantage in practical applications like solving MaxCut problem. Combining a classical optimizer that finds optimal parameters and a quantum device that computes the energy, this algorithm runs until the classical optimizer converges to the minimum energy solution. To reduce the time burden in optimizing an instance, A. Galda et al. [arXiv:2106.07531[quant-ph] (2021)] proposed improving the quality of initial parameters by choosing them to be optimal parameters of similar instances. It was proposed that similar graphs (a donor and an acceptor) have similar global optimal parameters. However, computing such similarity can be expensive in time, depending on the size of the graphs. Therefore, we created a cheaper metric that predicts similarity between two graphs using the true similarity between their constituent subgraphs. Our metric was consistent with the true similarity for only 15% of the random graph pairs in our dataset. Subsequent analysis revealed that even if two subgraphs do not share global optimal parameters, they could share a set of local optimal parameters with corresponding local optimal energies being within 98% of their global optimal energies. In fact, a set of parameters were found to optimize all the subgraphs and hence all their parent graphs as well. Hence, our metric, reliant on the true similarity of subgraphs, underestimated the similarity for most graph pairs, resulting in its poor performance. Moreover, this analysis motivated a novel scheme to predict optimal parameters of a graph using the shared local optimal parameters of its constituent subgraphs. Our findings present a pathway to use local properties of instances to accelerate hybrid quantum-classical algorithms.

I. INTRODUCTION

Noisy Intermediate Scale Quantum devices may soon solve problems of practical importance, doing so faster than classical computers and hence giving us the quantum advantage. One of these avenues is combinatorial optimization problems like MaxCut which has applications in electronic circuitry, statistical physics and computer science[1]. To address this problem, Farhi et. al.[2] devised QAOA, a hybrid quantum classical algorithm which variationally arrives at the minimum solution. It involves a classical optimizer which guessing parameters and quantum computer computing the energy as a function of those parameters.

Convergence to optimal parameters is, however, a time consuming task for QAOA, making it prone to quantum errors. To reduce the number of optimization iterations, Galda et. al.[3] suggested reusing optimal parameters of similar instances. Their technique is motivated by Brandao et. al.[4]'s findings which revealed that the value of QAOA objective function for 3-regular graphs of fixed size concentrated for fixed parameters. As a result, those instances share the same optimal parameters. While Brandao et. al. reasoned such *transferability* of optimal parameters was due to the popularity of the tree subgraph, Galda et. al. attributed it to the similar energy landscapes of 3-regular subgraphs. The latter claim suggests that transferability between two graphs is reliant on the transferability of their constituent subgraphs.

Nevertheless, Galda et al's approach of reusing optimal parameters of similar instances still faces a critical challenge: determining the similarity of two instances, a donor and an acceptor. In other words, we need to quantify how well do the donor's optimal parameters transfer onto the acceptor. A natural suggestion would be to compute acceptor's approximation ratio as a function of donor's optimal parameters. However, since we do not know acceptor's maximum energy, we cannot compute the approximation ratio. However, given Galda et al's claim about reliance of tranferability between the donor and the acceptor on their subgraphs, we can still *predict* the approximation ratio or similarity between them.

In this paper, we create a similarity metric that predicts the tranferability of a donor graph's optimal parameters onto the acceptor graph. In analyzing the metric's performance on random graphs, we also develop a novel approach to predict the optimal parameters for a graph without using a donor. Both contributions rely on local properties of graphs, namely their subgraphs. These techniques can further speed up convergence to optimal parameters and hence improve performance of QAOA on random graphs.

II. BACKGROUND

A. QAOA

Combinatorial optimization problems like MaxCut are usually defined on size N bit strings and m clauses. Each of these m clauses is a constraint on those bit strings. If a bit string $z = z_1 z_2 \dots z_N$ satisfies a clause C_α, then $C_\alpha(z) = 1$. Otherwise, $C_\alpha(z) = -1$. Together, all clauses constitute the cost function $C(z) = \sum_{\alpha=1}^{m} C_\alpha(z)$. To prepare such problems on the quantum computer, QAOA maps them onto a 2^N dimensional Hilbert space with computational basis states $|z\rangle$ and diagonalized operator $\hat{C}$ such that

$$\hat{C}|z\rangle = C(z)|z\rangle . \tag{1}$$

Each call to quantum computer involves two steps: preparing the trial state $|\gamma,\beta\rangle_p$ and measuring $\langle C\rangle_p = \langle\gamma,\beta|_p C|\gamma,\beta\rangle_p$. The trial state is expressed as

$$|\gamma,\beta\rangle_p = U_B(\beta_p)U_C(\gamma_p)\dots U_B(\beta_1)U_C(\gamma_1)|s\rangle \tag{2}$$

where $\hat{B} = \sum_{i=0}^{N} \hat{X}_i$ is the mixer hamiltonian, $U_G(\theta) = e^{-i\theta\hat{G}}$ is the exponentiation operator of $\hat{G}$, $\gamma_i \in [0, 2\pi], \beta_i \in [0, \pi]$ are angles, and $|s\rangle = |+\rangle^{\otimes N}$ is the uniform superposition state. After preparing the trial state, we can measure $\langle C \rangle_p$ through the hamiltonian averaging produdure[5] as

$$\langle C \rangle_p = \sum_{\alpha=1}^{m} \langle C\alpha \rangle_p \tag{3}$$

The computed energy as a function of parameters $\gamma = (\gamma_1, \ldots, \gamma_p), \beta = (\beta_1, \ldots, \beta_p)$ is given to a classical optimization algorithm which then guesses a new of parameters. These exchanges between quantum computer and classical optimizer continue until the latter converges to an optima.

B. MaxCut

Given an unweighted and undirected graph $G = (V, E)$ with $|V| = N$ vertices and $|E| = m$ edges, MaxCut aims to partition the vertices in two complementary sets A and B such that the number of edges between the two sets is maximized. Then, each bit z_i in the bitstring $z = z_1 z_2 \ldots z_N$ encodes whether the i'th vertex belongs to A ($z_i = 0$) or B ($z_i = 1$). The associated cost function to be maximized is

$$C = \sum_{jk \in E} C_{jk} \tag{4}$$

where each clause C_{jk} encodes whether the edge jk is contained entirely within one set or shared between the two sets. These clauses or local hamiltonians take the following form:

$$C_{jk} = \frac{1 - \hat{Z}_j \hat{Z}_k}{2} \tag{5}$$

C. Subgraphs

Computation of the expectation value of each local hamiltonian or edge C_{jk} can be further simpified through commutation properties of pauli matrices (Eq 16 in reference[2]). This simplification leaves $\langle C_{jk} \rangle$ containing only qubits corresponding to vertices at most p steps away from the central edge jk. We will refer to the resulting smaller graph as a *subgraph* f_g. As a result, the cost function simplifies to

$$\langle C \rangle_p = \sum_{g \in \{G\}} n(g) \langle f_g \rangle \tag{6}$$

where $\{G\}$ is the set of all the distinct subgraphs of graph G and $n_G(g)$ is the number of edges in graph G with the same subgraph f_g. For most graphs, computing the cost function using subgraphs (equation 6) is faster than doing without as shown in equation 3. In the following sections, we will exploit such subgraph decomposition for predicting similarity as well as predicting optimal parameters for random graphs.

III. METHODS: DATASET GENERATION

We first generated degree sequences for random graphs of fixed size of 20 nodes with different number of even degree nodes $(0, 2, 4, \ldots 20)$. For each of these 11 degree sequences, we generated 5 non-isomorphic random graphs, giving us a total of 55 random graphs for our dataset. Since none of the nodes in the graphs had a degree exceeding 6, we could easily generate the set of all possible 56 random subgraphs of depth $p = 1$.

Next, we found *global* optima for graphs by running QAOA algorithm on the quantum simulator QTensor[6]. Randomly initializing the parameters, the optimizer RMSprop from the pytorch package performed 200 iterations, with the largest of the energies in the set being the local optima. Due to presence of multiple local optima on random graphs' solution space, we followed the multi-start approach[7] by repeating the optimization procedure 20 times. Each multistart produced local optimal parameters $(\gamma_{Gi}, \beta_{Gi})$. From these 20 different optimal parameters, we picked the parameters $(\gamma_{G*}, \beta_{G*})$ for which G had the maximum cut. Subgraphs were also optimized using the same procedure.

IV. RESULTS

A. Predicting Similarity

Before comparing our similarity metric against true similarity, we will first describe how these two quantities were computed. True similarity between a donor $D = (V_D, E_D)$ and an acceptor $A = (V_A, E_A)$ is the approximation ratio obtained when we apply donor's optimal parameters $(\gamma_{D*}, \beta_{D*})$ onto the accepto,

$$\text{True Similarity}(D, A) = \frac{\text{A}(\gamma_{D*}, \beta_{D*})}{\text{A}(\gamma_{A*}, \beta_{A*})} \tag{7}$$

where $A((\gamma_{\text{G}*}, \beta_{\text{G}*}))$ is the acceptor A's energy as a function of graph G's best optimal parameters . On the other hand, the similarity metric combines the true similarity of all the pairs consisting of the donor's subgraphs and the acceptor's subgraph as

$$\text{SM}(D, A) = \sum_{d \in \{D\}} \sum_{a \in \{A\}} n_D(d) n_A(a) \frac{\text{True Similarity}(d, a)}{|E_D| \cdot |E_A|} \tag{8}$$

where SM is Similarity Metric and $|E_D| \cdot |E_A|$ is the total number of subgraph pairs.

First, we compared our similarity metric against the true similarity for all the possible graph pairs out of the 55 graphs in our dataset. The resulting distribution in figure 2 shows two clusters: one near the 45 degree line and the other that is concentrated on the line $y = 1$. While the former reflects agreement between the two metrics, the latter shows inconsistency. Given the sizes of these clusters, our similarity under approximates true similarity for 85% of the graph pairs in the dataset.

To address the inconsistency in figure 2, we considered the distribution of optimal parameters of donors and their subgraphs. As a case study, figure 1 considers a subgraph pair of two graphs with high true similarity but low similarlity metric. While the donor subgraph's optimal parameters don't transfer well, the donor's best optimal parameters do. In fact, the latter optimizes most subgraphs as shown in figure 3. As a result, figure 4 shows that these same parameters optimize most random graphs as well. While a majority of donors had their best parameters localized near these universal optimal parameters $((\beta/2\pi, \gamma/2\pi) = (0.118, 0.916)$ and $(0.383, 0.082))$ as shown in figure 5, the same is not true for subgraphs as shown in figure 6. Their optimal parameters were spread over non-universal regions on the parameter landscape. Since our similarity metric is dependent on the distrbution of these subgraphs' best optima and hence significantly affected by the non-universal optima, it underestimated similarity for most of the graph pairs.

This analysis forced us to rethink the definition of True Similarity. Our earlier findings reveal that most donor-acceptor random graph pairs are transferable if the donor's optimal parameters are universal. However, in some experimental settings, due to inadequate number of iterations or multistarts, QAOA may report non-universal optimal parameters as the donor's best optimal parameters. When tranferred onto an acceptor, we no longer have the guarantee of a high approximation ratio. To reflect these situations, we redefined true similarity as

$$\text{True Similarity}'(D,A) = \frac{1}{20}\sum_{i=1}^{20}\frac{\text{A}(\gamma_{D_i},\beta_{D_i})}{\text{A}(\gamma_{A*},\beta_{A*})} \tag{9}$$

In averaging over donor's optimal parameters from each multi-start, we avoid overestimating or underestimating transferability in practical settings.

As a result of this change, the performance of our similarity metric improved. The clusters we observed in figure 2 merged to show a linear relationship between our similarity metric and the new True Similarity' in figure 7. Yet, we still found our similarity metric undestimating similarity between graph pairs by 0.05 units on average. To probe this inconsistency, we hypothesized that the parity of graph pairs affects our prediction of similarity. Graphs with $0, 2$ or 4 even degree nodes were deemed *odd*. On the other hand, graphs with $16, 18$ or 20 even degree nodes were deemed *even*. The remaining graphs were left unclassified. Lastly, the self classification refers to graph pairs where the donor and acceptor are the same graph. Our results in figure 8 show that while our metric was consistent with true similarity of the different parity graph pairs, it was inconsistent for the same parity and the self graph pairs. Yet, the reason behind such inconsistency was still unclear from these findings.

Consequently, we tested whether mutual similarity of graphs affects our prediction for similarity for the same similarity and the self graph pairs. Mutual similarity of graph refers to whether QAOA parameters are tranferable among its subgraphs. Numerically, it is

$$\text{MS}(G) = \sum_{d\in\{G\}}\sum_{a\neq d\in\{G\}} n_G(d)n_G(a)\frac{\text{True Similarity'}(d,a)}{T_G} \tag{10}$$

where MS is Mutual Similarity and

$$T_G = \sum_{d\in\{G\}}\sum_{a\neq d\in\{G\}} n_G(d)\cdot n_G(a)$$

is the total number of subgraph pairs within the graph. The investigation of mutual similarity is motivated by Galda et. al.'s conjecture[3] that high mutual transferability among the donor and the acceptor is an essential condition for good transferability between them. Validating this claim, our results in figures 9, 10 show that graph pairs with lower mutual similarity for the donor also reported higher inconsistency between their similarity metric and true similarity. These results suggest that donor graphs with low mutual similarity may have some subgraphs with *bad* local optima. The corresponding *bad* optimal parameters may not optimize most subgraphs in the donor and hence the parent donor graph as well, resulting in its low mutual similarity. Since the donor and acceptor have similar energy landscapes, owing to their classification as a same parity or a self graph pair, the acceptor and hence its consituent subgraphs may also not share the *bad* optimal parameters. As a result, these *bad* optima will lower the subgraph similarity. However, true similarity between the donor and acceptor graphs is not reliant on optimal parameters of subgraphs, and hence unaffected by the donor's *bad* local optima. Therefore, subgraph similarity metric will underestimate true similarity of a graph pair if the donor has low mutual similarity.

B. Predicting Optimal Parameters

Alternatively, we can predict the optimal parameters of an acceptor directly without using a donor. As we were debugging the previous approach, we found that energy landscapes of most sugraphs had 4 local optima, and that these subgraphs shared some of their local optimal parameters with other subgraphs as well as parent graphs. Thus, our prediction for optimal parameter for a parent graph can be the optimal parameters shared by most of its constituent subgraphs.

For each subgraph, we combined the results of 50 multistarts each with 200 steps. This generated 4 clusters of local optima for most subgraphs. Then, we identified the centroids of these 4 clusters. Collecting 4 centroids for every subgraph, we then reapplied the clustering technique to find the center of the most densely populated region of subgraph centroids. This center of centroids was our prediction for the optimal parameters for the parent graph. Note whereas the unweighted approach described above identified 4 centroids for each subgraph, the weighted approach multiplied the number of centroids for each subgraph by its weight or frequency in the parent graph. Our results in figure 11 show the distribution of the resultant approximation ratios. Given that the mean unweighted and weighted approximation ratio exceeded 0.998, our technique was successful.

V. CONCLUSIONS

We demonstrated that subgraphs can be used to predict both similarity between two random graphs as well as their optimal parameters. Accounting for local optima of subgraphs was critical in both cases. We also found evidence for "universally good" parameters for random graphs, which extends the earlier studies on fixed angle conjecture for regular graphs[8],[4]. Taken together, these techniques can make QAOA non-variational or at worst, reduce the number of iterations for convergence.

However, our work was limited in scope: we only considered 55 twenty-node random graphs with maximum node degree of 6. A possible future direction could be verifying whether such predictive powers hold for more diverse set of random graphs, varying in size and node degree. Another future direction is exploring these techniques for higher depth QAOA circuits.

VI. ACKNOWLEDGEMENTS

I would like to thank my supervisor Yuri Alexeev as well as colleagues Alexey Galda, Ilya Safro, Danylo Lykov, Xiaoyuan Liu and Rusland Shaydulin. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

VII. REFERENCES

[1] F. Barahona, M. Grötschel, M. Jünger, and G. Reinelt, "An application of combinatorial optimization to statistical physics and circuit layout design," Operations Research **36**, 493–513 (1988).

[2] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," (2014), arXiv:1411.4028 [quant-ph].

[3] A. Galda, X. Liu, D. Lykov, Y. Alexeev, and I. Safro, "Transferability of optimal qaoa parameters between random graphs," (2021), arXiv:2106.07531 [quant-ph].

[4] F. G. S. L. Brandao, M. Broughton, E. Farhi, S. Gutmann, and H. Neven, "For fixed control parameters the quantum approximate optimization algorithm's objective function value concentrates for typical instances," (2018), arXiv:1812.04170 [quant-ph].

[5] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, "The theory of variational hybrid quantum-classical algorithms," New Journal of Physics **18**, 023023 (2016).

[6] D. Lykov, R. Schutski, A. Galda, V. Vinokur, and Y. Alexeev, "Tensor network quantum simulator with step-dependent parallelization," (2020), arXiv:2012.02430 [quant-ph].

[7] R. Shaydulin, I. Safro, and J. Larson, "Multistart methods for quantum approximate optimization," 2019 IEEE High Performance Extreme Computing Conference (HPEC) (2019), 10.1109/hpec.2019.8916288.

[8] J. Wurtz and D. Lykov, "The fixed angle conjecture for qaoa on regular maxcut graphs," (2021), arXiv:2107.00677 [quant-ph].

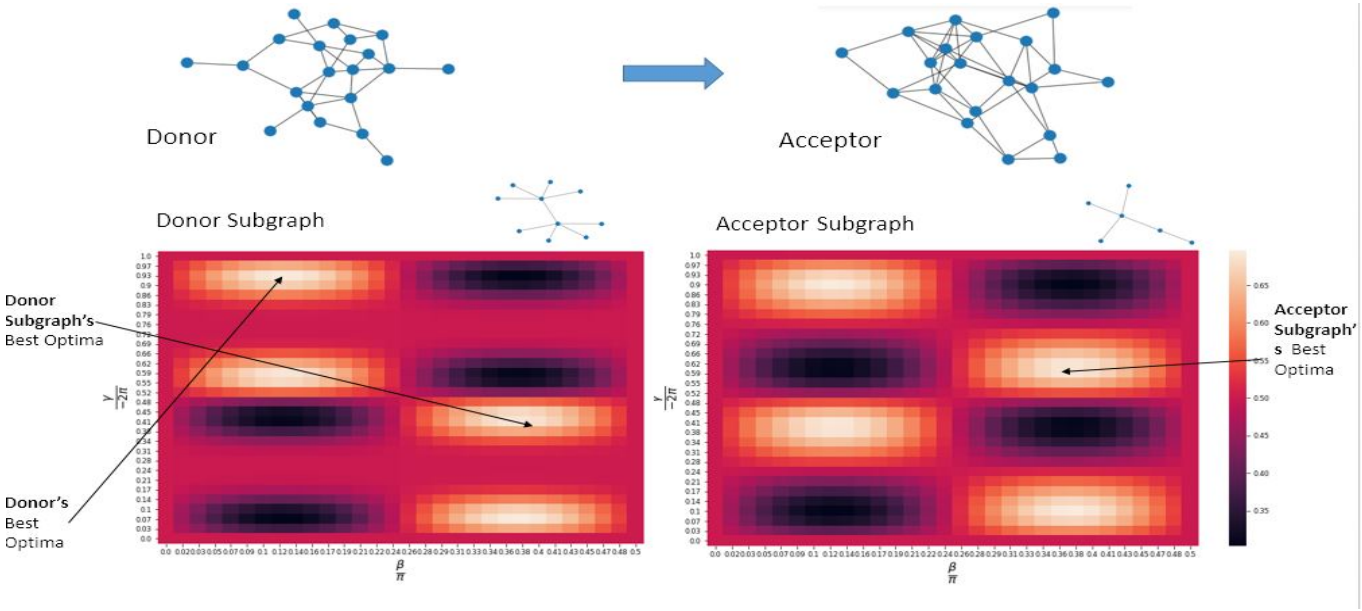

FIG. 1. Case Study of a subgraph pair of a graph pair with inconsistent similarity metric (0.65) and true similarity (0.99). If we tranfer the donor subgraph's best optimal parameters, we obtain an approximation ratio of 0.45. On the other hand, if we tranfer the donor's best optimal parameters, the ratio is 0.99.

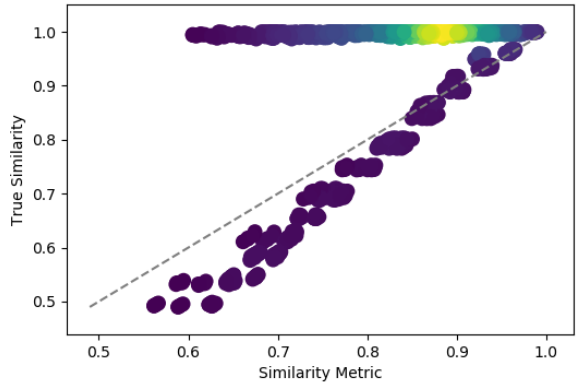

FIG. 2. Comparing subgraph similarity metric against true similarity for 55^2 graph pairs. The color indicates density of points in the surrounding region where yellow is high density and purple is low density.

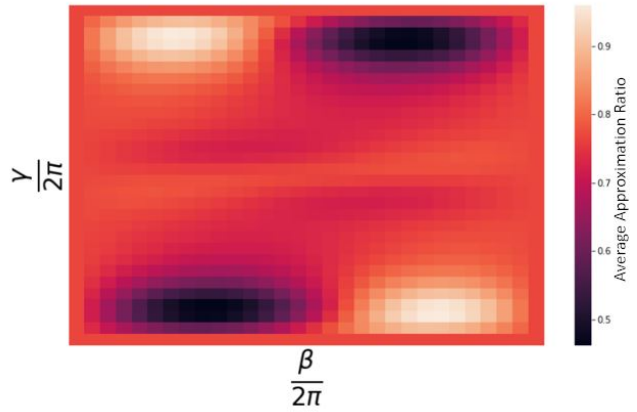

FIG. 3. Averaging approximation ratios of all 56 random subgraphs across the parameter landscape.

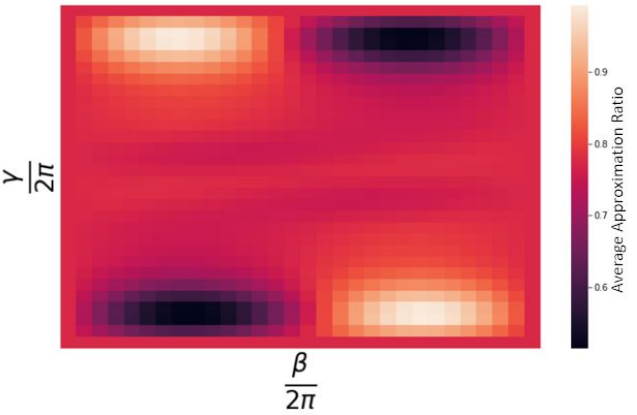

FIG. 4. Averaging approximation ratios of all 55 random graphs across the parameter landscape.

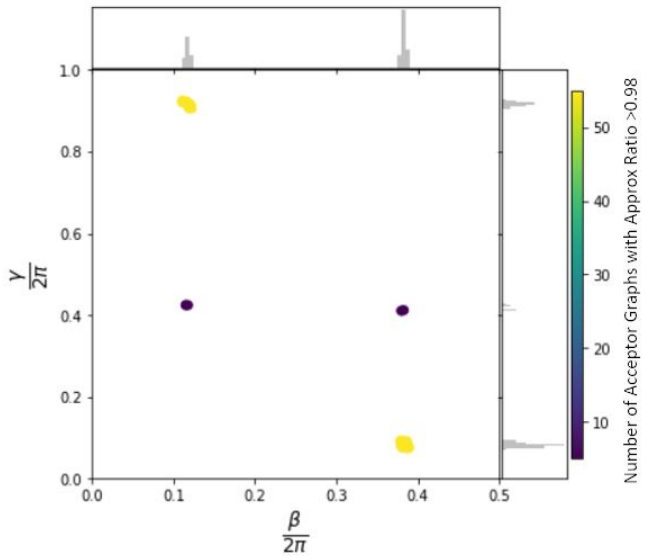

FIG. 5. Distribution of best optimal paramters of all the 55 random graphs in our dataset. The color of each point (β_i, γ_i) denotes the number of random graphs in the dataset that report high approximation ratios as a function of (β_i, γ_i). The histograms further highlight the marginal distribution of these optimal parameters.

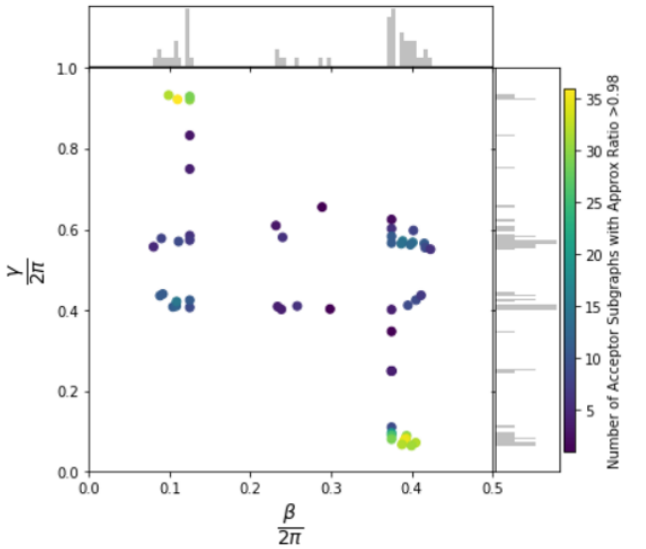

FIG. 6. Distribution of best optimal paramters of all the 56 random subgraphs in our dataset. The color of each point (β_i, γ_i) denotes the number of subgraphs in the dataset that report high approximation ratios as a function of (β_i, γ_i). The histograms further highlight the marginal distribution of these optimal parameters.

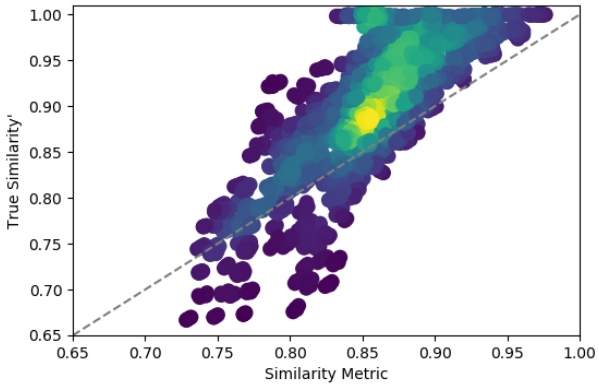

FIG. 7. Comparing subgraph similarity metric against the redefined true similarity for 55^2 graph pairs. The color indicates density of points in the surrounding region where yellow is high density and purple is low density.

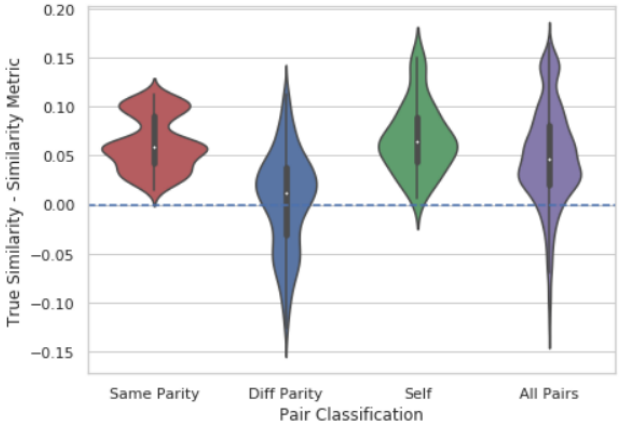

FIG. 8. Effect of parity of graphs in each graph pair on the difference between its redefined true similarity and similarity metric.

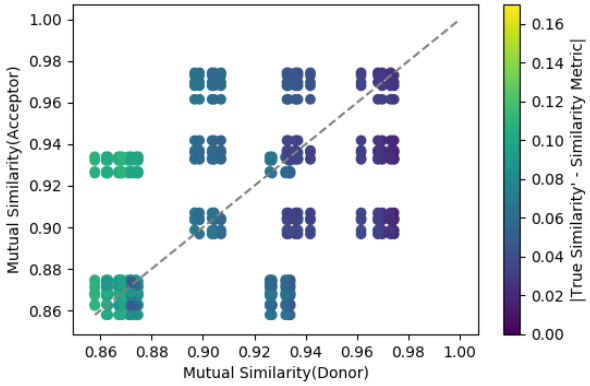

FIG. 9. Effect of Mutual Similarity on the fifference between redefined true similarity and similarity metric of the same parity graph pairs

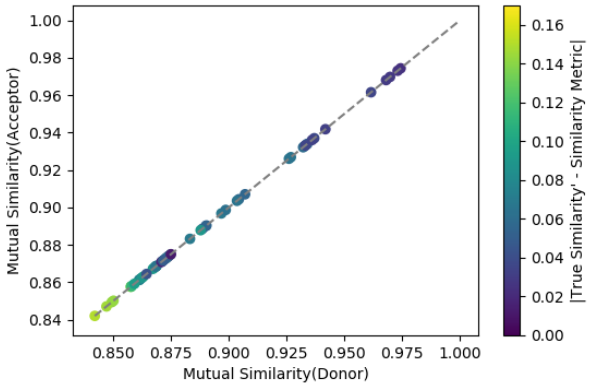

FIG. 10. Effect of Mutual Similarity on the fifference between redefined true similarity and similarity metric of "self" classified graph pairs

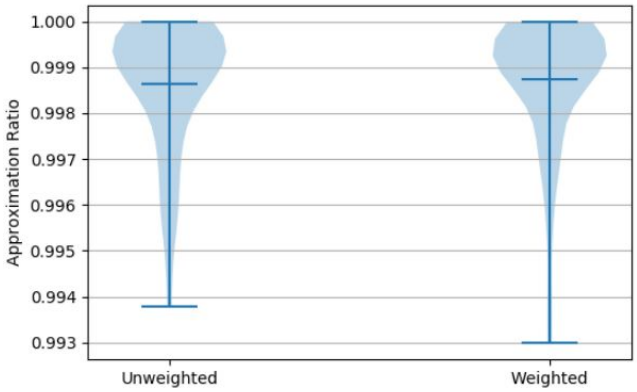

FIG. 11. Approximation ratios of random graphs using the unweighted and weighted overlapping optima technique.