Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Gaussian Boson Sampling for Efficient Dense Subgraph Identification in Omic-driven Gene Networks During Anthropogenic Methane Emission

Kin Tung Michael Ho
Imperial College London
Ching Jui Lai
National Cheng Kung University
Giin Yu Amy Tan
City University of Hong Kong
Po-Heng Lee
po-heng.lee@imperial.ac.uk

Imperial College London

Article

Keywords:

Posted Date: May 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9485151/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Gaussian Boson Sampling for Efficient Dense Subgraph Identification in Omic-driven Gene Networks During Anthropogenic Methane Emission

Kin Tung Michael Ho[1]*, Ching-Jui Lai[2], Giin Yu Amy Tan[3], Po-Heng Lee[1]**

[1] Department of Civil and Environmental Engineering, Imperial College London, London, United Kingdom

[2] Department of Mathematics, National Cheng Kung University, Tainan, Taiwan

[3] Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong SAR, China

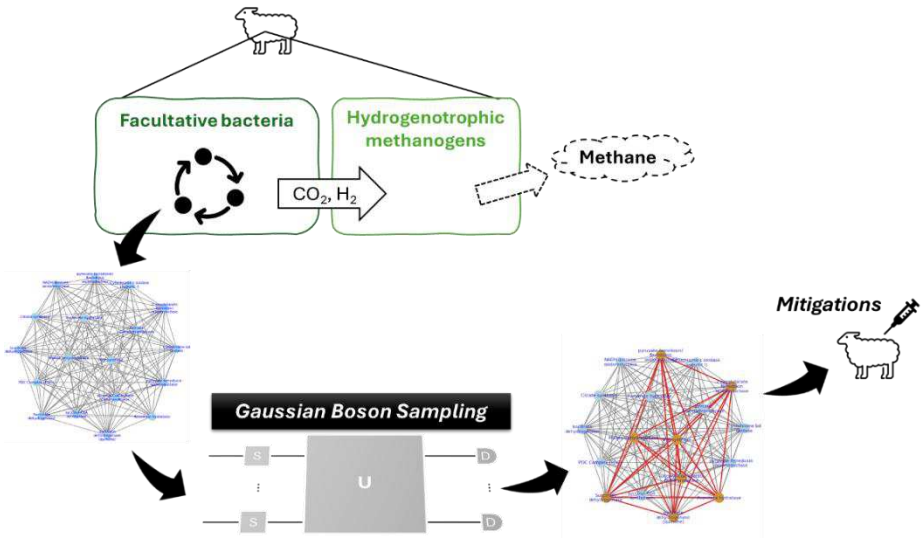
* First author: kin-tung-michael.ho20@imperial.ac.uk

** Corresponding author: po-heng.lee@imperial.ac.uk; phone: +44 020 7594 5993

Abstract

This work presents a Gaussian Boson Sampling (GBS) method to identify densely connected subgraphs from omics-derived gene networks. The method considers directional influence and utilises the intrinsic sampling bias of GBS towards densely connected nodes, providing an efficient strategy for dense subgraph identification compared with classical methods. We applied GBS to gene pathway networks of rumen from sheep gut under low and high methane yield conditions. The resulting key gene clusters reveal directional relationships consistent with known biochemical metabolic reactions in the microbe, supporting a mechanistic interpretation that electron carrier formation leads to enhanced hydrogenotrophic methanogenesis. Benchmarking against the classical Greedy Peel algorithm shows that GBS consistently recovers identical densest subgraphs but requires 5 to 10 times fewer subgraph evaluations. These results confirm that GBS can efficiently identify biologically meaningful functional modules in metabolic networks without the heavy computational overhead of exhaustive classical searches. Our findings support further investigation into the use of GBS for subgraph search in large multi-omic networks where dense modules exhibit functional significance.

Graphical abstract



Introduction

Biological systems can be modelled as networks in which nodes represent genes or proteins and edges denote functional interactions (Koutrouli et al., 2020). Such representations have been applied to metabolic pathway analysis, protein–protein interaction mapping, and gene co-expression profiling, enabling systematic characterisation of cellular organisation (Abdel-Hafiz et al., 2022; Everett et al., 2006; Feng et al., 2011; Hu et al., 2005; R. Li et al., 2022). Within these networks, densely connected regions are of particular interest because they often correspond to coherent gene or enzyme clusters that collectively characterise metabolic performance. Their identification can therefore be formalised as a dense subgraph identification task, which becomes increasingly challenging as network size grows.

Network construction commonly uses meta-omic data from high-throughput sequencing to infer relationships among functional elements. Metagenomics quantifies gene potential, whereas metatranscriptomics measures gene expression activity. These measurements can be integrated to compute associative statistics, including relative ratios and correlation coefficients, which are encoded as edge weights. The resulting weighted graphs capture the relative strengths between genes across samples or experimental conditions.

Extending this framework, multi-omics studies integrating two or more omics layers further enable analysis of coordinated molecular changes in biological systems. Hayes et al. (2024) emphasised that multi-omics provides a more comprehensive framework for identifying key metabolic drivers via vertical integration (different omic layers of the same sample), horizontal integration (the same omic layer across different samples), and diagonal integration (different omic layers across different samples). However, these schemes produce larger, higher-dimensional networks, thereby expanding the combinatorial search space for dense subgraph and increasing both computational burden and susceptibility to noise.

Classical dense subgraph methods are typically categorised as exact or approximate (Gudapati et al., 2021; Lanciano et al., 2024). A naïve exact approach exhaustively evaluates every possible subgraph; for a graph with n nodes, this requires examining 2^n candidate subgraphs, rendering brute-force search computationally infeasible even for moderately sized networks. Approximate heuristic methods improve scalability but reduce optimality guarantees. For example, greedy peeling algorithms iteratively remove nodes based on density and implicitly evaluates edge contributions across up to $\frac{n(n-1)}{2}$ possible pairwise interactions. While polynomial-time and practically efficient, such heuristics can be sensitive to graph structure and may discard nodes belonging to small, highly interconnected modules. This trade-off between exhaustive optimality and heuristic efficiency motivates alternative strategies for better exploring dense regions without explicit enumeration of the entire search space.

Gaussian Boson Sampling (GBS) is a photonic quantum computing method that generates samples from a probability distribution determined by a matrix representing the network graph (Arrazola & Bromley, 2018; Deng et al., 2023). Each photon detection event corresponds to a subset of modes that can be interpreted as a candidate subgraph. The sampling probability is proportional to the Hafnian function, which is empirically associated with the subgraph density. This induced bias enables GBS to preferentially select dense subgraphs, enhancing the stochastic search (Arrazola & Bromley, 2018; Hamilton et al., 2017). Empirical work has reported that GBS can double the success probabilities of finding dense subgraphs relative to classical random sampling (Banchi et al., 2020). In terms of applications, it was reported improved success for maximum clique identification, including molecular docking and RNA-folding (Yu et al., 2023). Nevertheless, despite these algorithmic and experimental advances, applications of GBS to real-world multi-omic networks remain limited, particularly in settings where cross-layer directional influences are central to biological interpretation.

In this work, we introduce a GBS-based framework to identify key gene groups by constructing weighted networks from metagenomic (gene potential) and metatranscriptomic (gene activity) data from a rumen microbiome (Wu et al. 2023). Previous analyses considered gene-to-transcript ratios for 17 genes within individual metabolic pathways associated with methane yield, but did not quantify cross-gene interactions. As a result, coordinated cross-layer effects influencing methane production have not been examined. Ruminant livestock is a major source of anthropogenic greenhouse gas emissions (Javier Mater-Sagasta et al., 2017; Marino et al., 2016; Sejian et al., 2017). In the rumen, microbial metabolism regulates carbon catabolism and electron carrier redox cycling through simultaneous anaerobic and aerobic respiratory reactions. This regulation influences the production of organic acid, carbon dioxide and molecular hydrogen, which directly modulate methanogenesis. Cross-layer dense subgraphs may therefore capture subtle differences in coordinated metabolic modules that underlie variations in methane yield. Identifying such key gene clusters can inform mitigation strategies including feed manipulation, additives and probiotics (Javier Mater-Sagasta et al., 2017).

Methodologically, we introduced node duplication to represent directional influence within and across omics layers. We then applied GBS to identify dense subgraphs and benchmarked against a classical Greedy Peel baseline. In addition, we quantified sampling efficiency of GBS via approximation ratios relative to uniform sampling, compare computational costs and runtimes of GBS and Greedy Peel, and perform runtime decomposition to identify bottlenecks that guide future hardware and algorithmic optimisations.

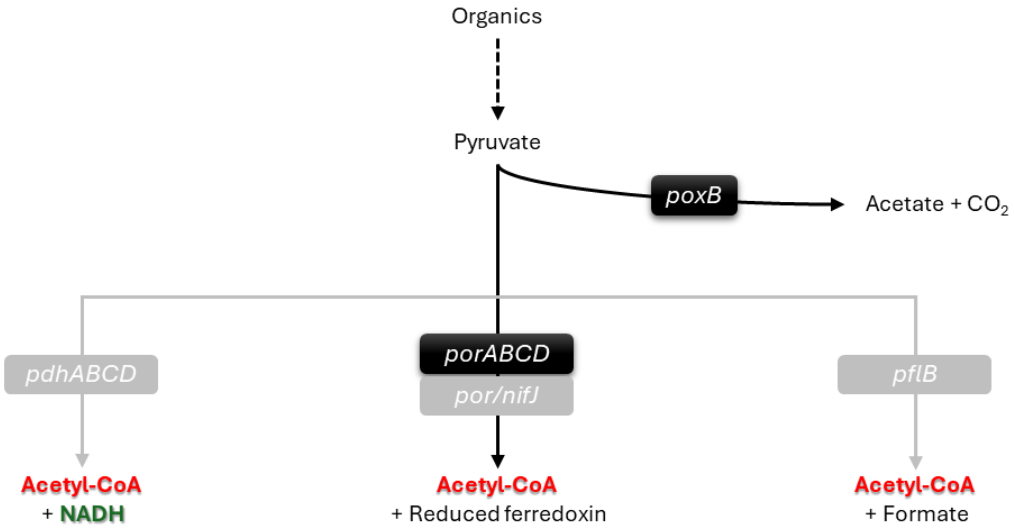
Results and Discussion

Overview

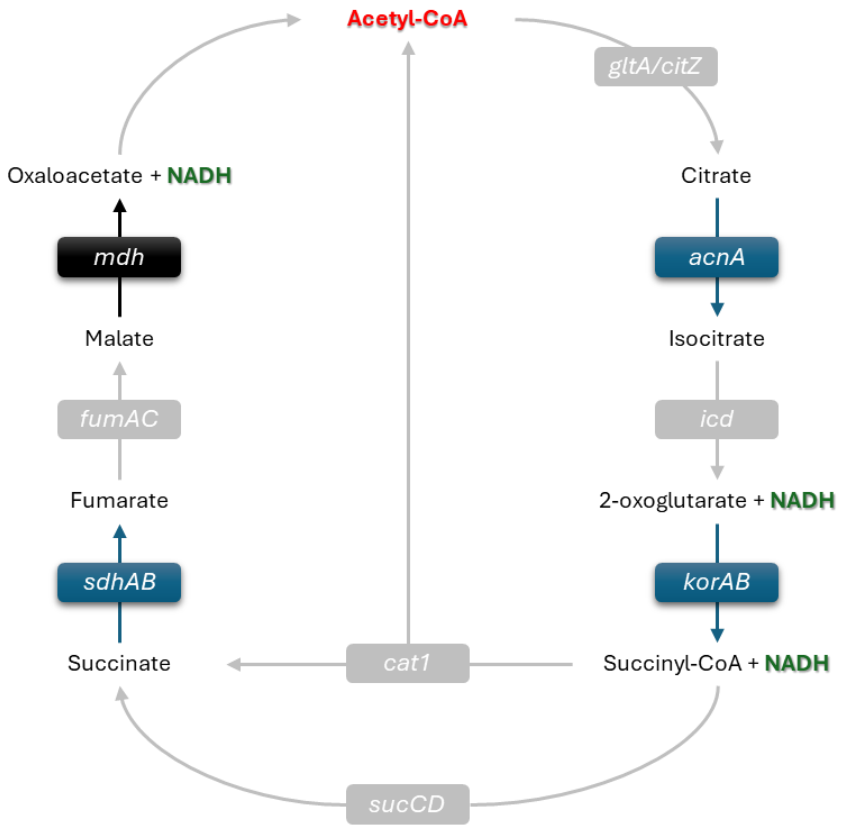
This study analysed eight raw omics datasets derived from sheep rumen samples reported by Wu et al. (2023), comprising duplicate paired metagenomic and metatranscriptomic measurements under contrasting conditions of low methane yield (10.9 ± 0.9 g CH_4 per kg dry matter intake) and high methane yield (16.1 ± 0.2 g CH_4 per kg dry matter intake). Within each omic type and methane condition, paired datasets showed no statistically significant differences and were therefore averaged for downstream analyses as described in the Methods section. Methane yield was treated as the primary biological variable throughout.

All datasets contained measurements for the same set of 17 genes, ensuring biochemical consistency across omics layers and methane yield conditions. These genes were organised into three functional groups, reflecting pyruvate metabolism, tricarboxylic acid (TCA) cycle and energy conservation. Figure 1 illustrated the roles of the genes in metabolism and highlighted the densely connected genes identified by the dense subgraph search algorithms.

Pyruvate metabolism



TCA cycle



Energy production

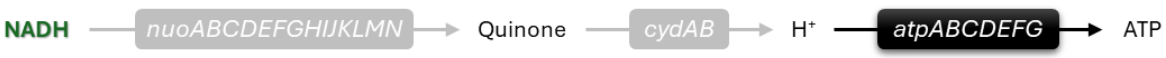


Figure 1 Role of the enzymes in bacterial metabolism encoded by the analysed genes. Genes highlighted in black represent densely connected genes identified under both methane conditions (multi-omics level), whereas genes highlighted in blue represent those appeared only under high methane yield conditions. All genes appeared in the low methane condition also appeared under the high methane condition.

The pyruvate metabolism group regulates the entry of pyruvate into downstream metabolic pathways and determines the generation of reduced electron carriers, including quinol and reduced ferredoxin. This group comprised *poxB*, *por/nifJ*, *porABCD*, and *pdhABCD*, which encode pyruvate:quinone dehydrogenase, pyruvate-ferredoxin/flavodoxin oxidoreductase, pyruvate ferredoxin oxidoreductase, and the pyruvate dehydrogenase complex, respectively.

The TCA cycle group mediates oxidative carbon flux and contributes to the production and recycling of NADH, $FADH_2$ and quinone intermediates. This group included *pflB*, *gltA/citZ*, *acnA*, *icd*, *korAB*, *sucCD*, *cat1*, *sdhAB*, *fumAC*, and *mdh*, which respectively encode formate C acetyltransferase, citrate synthase, aconitate hydratase, isocitrate dehydrogenase, 2-oxoglutarate ferredoxin oxidoreductase, succinyl-CoA synthetase, succinyl-CoA acetate CoA transferase, succinate dehydrogenase, fumarate hydratase, and malate dehydrogenase.

The energy production group links electron transport to proton motive force generation and ATP synthesis, thereby integrating redox carrier balance with cellular energy output. This group consisted of *nuoABCDEFGHIJKLMN*, *cydAB*, and *atpABCDEFGH*, which encode NADH quinone oxidoreductase, cytochrome bd oxidase, and ATP synthase, respectively.

This set of genes and its functional classification were used consistently throughout all network constructions and analyses. For single-omic analyses, two metagenomic and two metatranscriptomic networks were generated using node duplication, each comprising 34 nodes. For multi-omic analyses, two networks were constructed by combining metagenomic

and metatranscriptomic layers, resulting in graphs with 68 nodes. These networks formed the structural basis for dense subgraph identification.

Dense subgraph searches were performed using both the GBS and Greedy Peel algorithms, with Greedy Peel serving as the reference method. These subgraphs were presented in Figure S1. In all cases, the densest subgraphs identified by GBS were identical to those obtained using Greedy Peel. This agreement confirms that GBS produces results consistent with the established Greedy Peel approach, supporting its suitability for this type of network analysis. Moreover, GBS requires fewer evaluations, offering a more computationally efficient alternative while preserving the same solutions.

Multi-Omics Network Under Low Methane Yield

As showed in Figure 2(a), the densest subgraph identified in the combined metagenomic and metatranscriptomic network under low methane yield conditions consisted of *poxB*, *porABCD*, *mdh* and *atpABCDEFGH*. All four genes were present at both the metagenomic and metatranscriptomic levels, indicating coordinated representation across omics layers within the integrated dense module.

This densely connected group of genes centres on quinone reduction via *poxB* and ferredoxin reduction via *porABCD*, with NADH generation through *mdh*. *poxB* catalyses pyruvate oxidation coupled to quinone reduction (Gecse et al., 2024), thereby directly feeding electrons into the quinone pool. *porABCD* enables conversion of pyruvate to acetyl-CoA (Abdel-Hamid et al., 2001; Gecse et al., 2024; Katsyv et al., 2021), producing CO_2 that can serve as substrate for hydrogenotrophic (hydrogen-utilising) methanogenesis. *mdh* participates in oxaloacetate–malate interconversion and contributes to NADH cycling, while *atpABCDEFGH* suggests coupling between electron transport and ATP synthesis. However, the absence of genes

encoding enzymes from the oxidative branch of the TCA cycle indicates a limited expansion of electron carrier generation, particularly $FADH_2$-linked reactions.

In rumen microbial systems, excess electron carriers are re-oxidised partly through hydrogen production, which subsequently increases hydrogenotrophic methanogenesis. Limited carbon flux through the oxidative TCA branch therefore constrains the overall availability of electron carriers for hydrogen formation. Under low methane yield conditions, the redox balance associated by this cohort of genes may limit the electron flow towards hydrogen production, thereby aligning with the reduced methane yield observed.

Within this dense module, *poxB* acted as the sole source gene, while *porABCD*, *mdh* and *atpABCDEFGH* functioned as effect genes (Figure S1(a)). This single-source organisation suggests that upstream pyruvate oxidation represents the dominant initiating reaction within the gene cluster.

Multi-Omics Network Under High Methane Yield

The densest subgraph identified in the multi-omic network under high methane yield conditions consisted of the genes observed under low methane yield conditions, with three additional genes from the oxidative branch of the TCA cycle. As seen in Figure 2(a), *sdhAB* and *acnA* appeared at the transcript level and *korAB* at the genomic level, indicating cross-layer engagement of oxidative metabolism.

sdhAB catalyses succinate oxidation and transfers electrons to the quinone pool, directly contributing to $FADH_2$-linked electron flow. *korAB* mediates oxidative decarboxylation within the TCA cycle, generating NADH. The inclusion of these genes expands the overall pool of electron carriers available for downstream reactions, enabling re-oxidation of excess reducing electron carriers that produces hydrogen. Increased hydrogen availability in turn enhances

hydrogenotrophic methanogenesis. The findings under high methane yield conditions are thus consistent with increased electron carrier turnover and the relatively higher methane yield.

acnA catalyses citrate–isocitrate interconversion within the oxidative branch of the cycle (Z. Li et al., 2023). The occurrence of *sdhAB* and *acnA* is notable, as metabolites associated with each gene have been reported to inhibit the other under certain conditions. *acnA* converts citrate into itaconate as an intermediate, which can inhibit *sdhAB* (Z. Li et al., 2023; O'Neill & Artyomov, 2019; Sun & Chang, 2022), while *sdhAB* converts succinate into fumarate, which can inhibit *acnA* (Guberovic & Frezza, 2024; Ternette et al., 2013).

Within this group, *poxB*, *sdhAB* and *acnA* acted as source genes, while *atpABCDEFGH*, *mdh*, *porABCD* and *korAB* functioned as effect genes (Figure S1 (b)). This multi-source organisation contrasts with the single-source structure observed under low methane yield conditions and reflects distributed initiation of oxidative flux within the gene cluster.

Metagenomic Network Under Low Methane Yield

The densest subgraph of the metagenomic network under low methane yield conditions consisted of *poxB*, *porABCD*, *sdhAB*, *korAB*, *mdh* and *atpABCDEFGH* (Figure 2(b)). Compared with the integrated multi-omics result, the key gene cluster additionally included *sdhAB* and *korAB*, both associated with the oxidative branch of the TCA cycle.

korAB catalyses oxidative decarboxylation within the TCA cycle, while *sdhAB* links succinate oxidation to the quinone pool. Their presence at the metagenomic level indicates that the gene potential for oxidative metabolism is enhanced. However, their absence from the multi-omic dense subgraph suggests that this broader capacity may not be expressed at the transcriptional level under low methane yield conditions. Thus, although genes supporting enhanced quinone reduction and $FADH_2$-associated electron transfer are present, the redox state may not be dominated by oxidative TCA activity. This supports the interpretation that electron carrier

turnover and downstream hydrogen production remain comparatively constrained under low methane yield conditions.

Metagenomic Network Under High Methane Yield

The densest subgraph consisted of the same group of genes identified in the multi-omics dense subgraph (Figure 2(b)). While the multi-omics network demonstrated cross-layer integration of oxidative metabolism, the metagenomic result reflects the underlying genetic capacity for this extended redox engagement. The consistency between the densest subgraphs at meta-genomic and multi-omic levels supports the interpretation that enhanced oxidative TCA participation is structurally embedded in the system under high methane yield conditions.

Metatranscriptomic Under Low Methane Yield

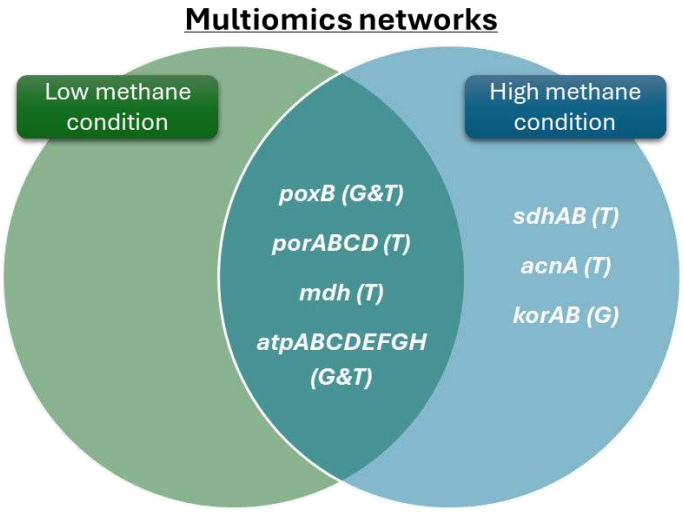
The densest subgraph of the metatranscriptomic network consisted of *poxB*, *porABCD*, *mdh* and *atpABCDEFGH*, identical to that of the multi-omic network (Figure 2(c)). The consistency between these two networks reinforces the interpretation that, under low methane yield conditions, metabolic organisation is confined to relatively narrow redox pathways rather than extended to oxidative TCA pathways. This pattern further supports the observation that limited electron carrier generation may reduce hydrogen availability for methanogenesis.

Metatranscriptomic Network Under High Methane Yield

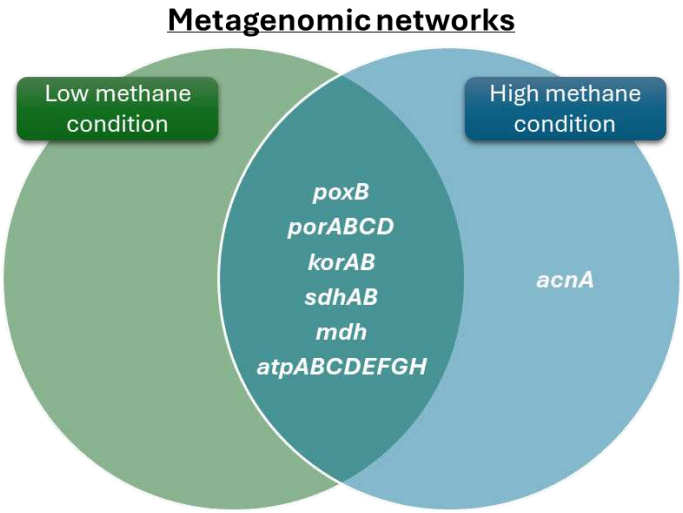
Relative to the multi-omic densest subgraph, *korAB* was not retained in the densest subgraph of the metatranscriptomic network, whereas the remaining genes were preserved (Figure 2(c)). This dense metabolic group captures the core oxidative pathway identified in the multi-omics analysis, particularly through the inclusion of *sdhAB* and *acnA*. This transcriptional alignment with the integrated network reinforces the interpretation that expanded redox carrier turnover contributes to increased hydrogen availability and thereby methane production.

235

a


Multiomics networks
Low methane condition
High methane condition
poxB (G&T)
porABCD (T)
mdh (T)
atpABCDEFGH (G&T)
sdhAB (T)
acnA (T)
korAB (G)

b


Metagenomic networks
Low methane condition
High methane condition
poxB
porABCD
korAB
sdhAB
mdh
atpABCDEFGH
acnA

c

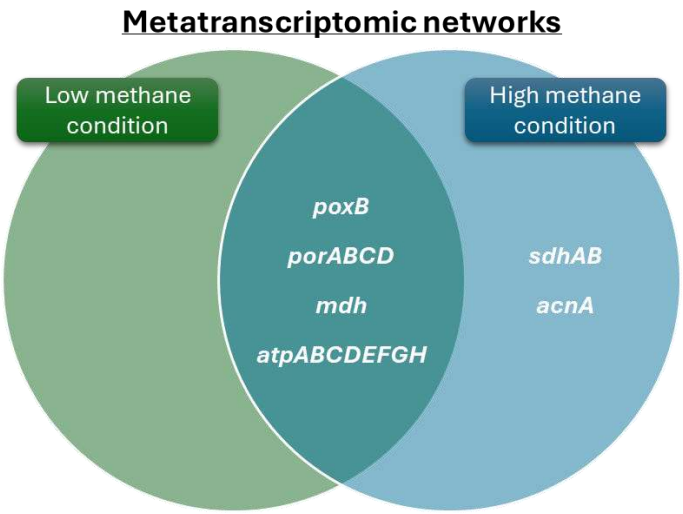

Metatranscriptomic networks
Low methane condition
High methane condition
poxB
porABCD
mdh
atpABCDEFGH
sdhAB
acnA

Figure 2 Comparison of genes identified in the densest subgraphs under low and high methane yield conditions. Venn diagrams show the overlap of genes detected in densest subgraphs for (a) multi-omics networks ('G' denotes genes identified at the metagenomic level and 'T' denotes genes identified at the metatranscriptomic level), (b) metagenomic networks, and (c) metatranscriptomic networks. The overlapping region indicates genes present under both methane yield conditions.

Computational Performance Analysis

Sampling Efficiency

To quantify sampling efficiency, we compared approximation ratios achieved via GBS and uniform random sampling. The approximation ratio was defined as the density of the densest sampled subgraph divided by the density of the optimal subgraph obtained using Greedy Peel.

As shown in Figure S2, GBS consistently achieved an approximation ratio of 1 across different numbers of shots for all networks. In contrast, uniform random sampling fluctuated around approximately 0.4 and did not converge to the optimal density even with increasing shots.

This consistent outperformance indicates that the intrinsic mathematical formulation of GBS biases sampling towards dense subgraphs. Rather than sampling uniformly from the subgraph space, the probability distribution of GBS concentrates on dense regions of the graph, enabling identification of the optimal solution with substantially fewer samples than uninformed random sampling.

Subgraph Evaluation Counts

A brute-force approach evaluates every possible subgraph of a given network. For the single-omic networks, which contain 34 nodes derived from 2×17 genes, this requires evaluation of

2^{34}subgraphs, corresponding to more than 17 billion candidate subgraphs. For the multi-omic networks, which contain 68 nodes, the number of candidate subgraphs increases to 2^{68}, exceeding 295 quintillion evaluations.

The Greedy Peel algorithm has complexity $O(m + n \log n)$ with n the number of nodes and m the number of edges (Gudapati et al., 2021). At the first step, the algorithm already needs to evaluate $\frac{n(n-1)}{2}$ subgraphs obtained by removing each edge from the original complete graph. This corresponds to 561 evaluations for a 34-node network and 2278 evaluations for a 68-node network. Although this is dramatically smaller than the exhaustive enumeration of the brute-force approach, the number of evaluated candidates still increases quadratically with the number of nodes.

In contrast, GBS generates $m \times n$ candidate subgraphs per network, with $m = 5$ chosen based on the approximation ratio analysis. The number of subgraph sampling scales approximately as $O(n)$ under GBS, which is significantly lower than $O(n^2)$ for Greedy Peel and $O(2^n)$ for brute-force. By removing repeated and empty samples, the numerical results are summarised in Figure 3. These findings indicated that GBS substantially reduces the number of subgraph evaluations required, providing a more computationally efficient approach for dense subgraph searches.

a **b**

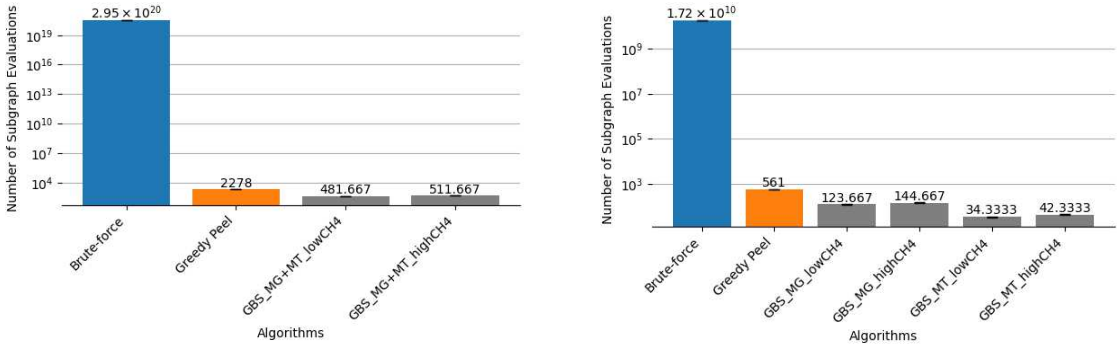


Figure 3 Number of Subgraph Evaluations needed using brute force (blue bars), greedy peel (orange bars) and GBS (grey bars). Bar chart (a) corresponds to the multi-omics network, whereas bar chart (b) corresponds to the metagenomic and metatranscriptomic networks.

Runtime Comparison

Total runtimes of Greedy Peel and GBS are summarised in Table 1. Despite requiring more subgraph evaluations, Greedy Peel exhibited shorter runtimes across all networks. For the 34-node networks, Greedy Peel required less than 0.2 seconds, whereas GBS required more than 6 seconds. For the 68-node networks, Greedy Peel required around 1 second, while GBS required approximately 30 seconds.

Breaking down the GBS runtime shows that the sampling step dominates the total computation time, while the density evaluation step contributes only a negligible fraction. This observation is consistent with prior analyses demonstrating that classical simulation of GBS is dominated by torontonian computation (Quesada et al., 2018).

Therefore, when implementing GBS on a classical simulator, the sampling step constitutes the principal computational bottleneck. The reduction in subgraph evaluations does not directly lead to shorter runtime because sample generation itself is numerically intensive. This finding encourages two broader research directions in the future. First, improving classical simulations

would increase sampling efficiency, thereby shortening overall runtime (Yang & Wang, 2024). Second, advancing the development of real optical hardware for GBS could generate samples physically rather than numerically, thereby reducing the dominant computational cost. Nonetheless, the scalability of photonic GBS hardware to larger datasets highlights its potential to outperform classical benchmarks, demonstrating a clear trajectory toward practical quantum advantage in multi-omic studies.

Table 1 Runtime of GBS (Sampling and evaluations) and Greedy Peel

Datasets		**GBS Runtime (seconds)**			**Greedy Peel Runtime (seconds)**
		Sampling	Evaluation	Total	
Metagenomics	Low-methane	6.41 ± 0.49	0.0035 ± 0.0015	6.42 ± 0.49	0.16
	High-methane	6.82 ± 0.35	0.0055 ± 0.0020	6.82 ± 0.35	0.11
Metatranscriptomics	Low-methane	6.14 ± 0.49	0.0024 ± 0.0020	6.15 ± 0.49	0.11
	High-methane	6.57 ± 1.39	0.0028 ± 0.0020	6.58 ± 1.39	0.15
Metagenomics + Metatranscriptomics	Low-methane	29.28 ± 3.58	0.0096 ± 0.0030	29.29 ± 3.58	1.01
	High-methane	35.69 ± 1.92	0.0127 ± 0.0019	35.70 ± 1.92	0.80

Conclusion

This work developed an identification method for key gene clusters that captures directionality and integrates multiple omics layers systematically via dense subgraph search. The results showed that GBS-based sampling can identify the same dense subgraphs as classical methods while requiring significantly fewer subgraph evaluations, highlighting its potential for large-scale problems. Biological analysis of the identified gene clusters revealed metabolic differences associated with methane yield, where low methane yield networks exhibited limited oxidative TCA activity and a single-source pyruvate oxidation structure, whereas high methane yield networks included additional oxidative TCA genes that expanded electron carrier turnover and supported increased hydrogen availability for methanogenesis. These findings support further investigation of this pipeline as a scalable tool for analysing microbial metabolism in metabolic engineering and environmental biotechnology.

Method

Figure 4 illustrates the methodology overview of this work. All computations were written in Python language. The omics datasets were pre-processed to construct the network graphs through node duplications. The density of subgraph was then defined, followed by the algorithmic developments of Greedy Peel, GBS, and uniform random sampling. The densest subgraphs identified by GBS were compared to that by Greedy Peel to ensure correctness. The findings were also validated with biological knowledge.

To evaluate the technical performance, we measured and compared the runtime between the two algorithms. The number of subgraph evaluations of these two algorithms were then compared alongside that of the brute-force algorithm. Lastly, to investigate if GBS is intrinsically favouring dense subgraph search, we compared the approximation ratio achieved by GBS and uniform sampling at given numbers of shots.

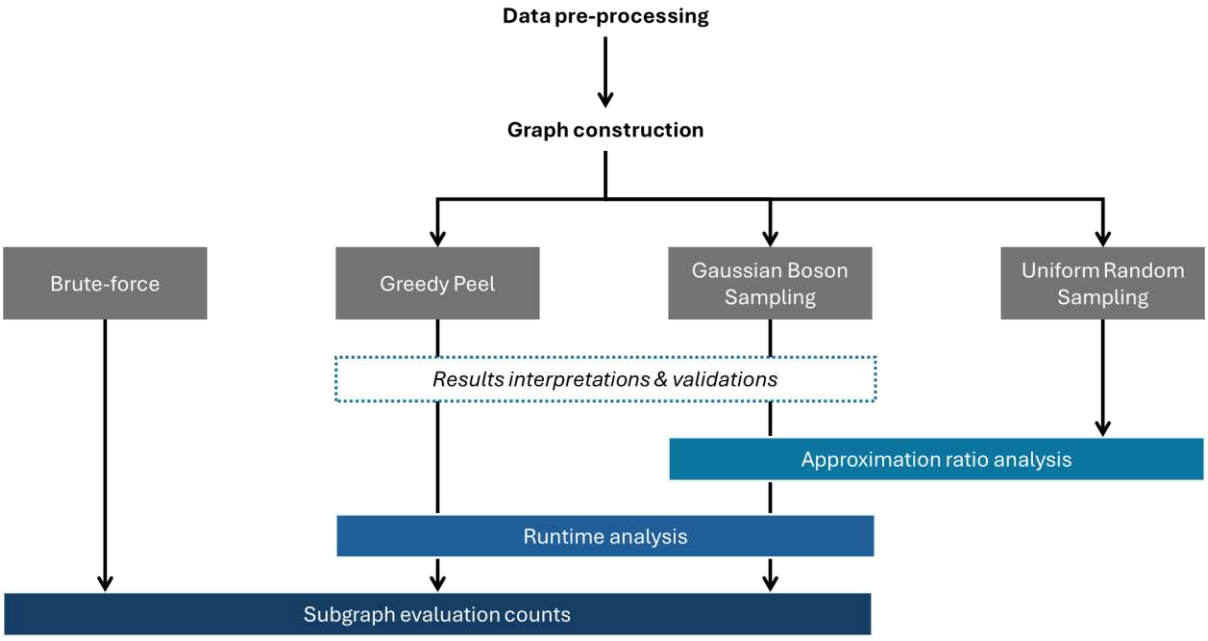


Figure 4 Schematic of the methodology of this work

Data source and pre-processing

The omics data, obtained by Wu et al. (2023), consisted of metagenomic and metatranscriptomic sequences of 8 raw sheep rumen samples, alongside the corresponding methane emission levels (See Table SX). Samples G1 and G2 are replicates for metagenomic measurements at low-methane conditions, while Samples T1 and T2 are replicates for metatranscriptomic measurements at low-methane conditions. Similarly, Samples G3 and G4 are replicates for metagenomic measurements at high-methane conditions, while Samples T3 and T4 are replicates for metatranscriptomic measurements at high-methane conditions.

All omics values were normalised using housekeeping genes. For this study, subunit values of each metabolic were summed to obtain abundance values.

Statistical tests to average replicate measurements

Statistical comparisons were conducted to determine whether replicate measurements within each dataset could be averaged for subsequent analyses. Tests were performed on the metabolic abundance vectors within each replicate pair. A significance threshold of $\alpha = 0.01$ was used throughout.

Metagenomic replicates (G1 vs. G2; G3 vs. G4) showed normality in the Shapiro–Wilk test ($p = 0.047$ and $p = 0.044$, respectively) and were analysed with paired t-tests, resulting in p-values of 0.044 and 0.114. Metatranscriptomic replicates (T1 vs. T2; T3 vs. T4) did not meet normality in the Shapiro–Wilk test ($p < 0.001$ in both cases) and were compared using Wilcoxon signed-rank tests, yielding p-values of 0.015 and 0.035. Since both the metagenomic and metatranscriptomic data did not exceed the $\alpha = 0.01$ significance threshold, each pair of replicates was averaged.

Graph construction and node supplication

Node duplication was used to allow encoding of directionality in enzymatic relationships. For each metabolic i, two distinct nodes were created: a start node $e_{i,s}$ and an end node $e_{i,e}$. This

yields a graph with $2n$ nodes for n genes, permitting the construction of a square $2n \times 2n$ adjacency matrix where directional influence can be captured. In this representation, start and end nodes encode directional interactions between genes through the corresponding symmetric matrices.

Let $x \in R^n$ denote the vector of omics values (averaged across replicates and duplicated to match the $2n$-node representation). The matrix $A \in R^{2n \times 2n}$ is constructed such that each entry A_{ij} represents the interaction strength between metabolic nodes. For each pair of nodes i and j:

$$A_{i,j} = \{\frac{x_j}{x_i}\ , if\ x_i \neq 0\ and\ i = start\ node, j = end\ node\ \ \frac{x_i}{x_j}\ , if\ x_j \neq 0\ and\ j = start\ node, i = end\ node\ \ 0\ , otherwise$$

This ensures all entries are non-negative and interpretable as pairwise expression ratios aligning with Wu et al. (2023). Diagonal entries were set to zero to avoid self-loops.

Density definition

The density $D(S)$ of a subgraph $S \subseteq V$ was defined as the total edge weight per node within that subgraph:

$$D(S) = \frac{1}{|S|}\sum \quad A_{ij} \quad i, j \in S$$

where the summation is taken over all node pairs within S. Since diagonal entries of A were set to zero, self-loops do not contribute to the density. This formula was used to determine the densities of the subgraphs obtained across all algorithms.

Development of Greedy Peel algorithm

The Greedy Peel algorithm began with the complete graph and progressed iteratively. At each iteration, the current subgraph's density was evaluated. The node whose removal resulted in

the smallest decrease in density (equivalently, the highest remaining density after removal) was eliminated, and the procedure was repeated on the reduced subgraph. Iterations continued until no nodes remained. The densest subgraph was identified as the one achieving the maximum density across all iterations. The results of Greedy Peel served as the classical deterministic benchmark and were used as the reference for comparison with GBS.

Development of GBS algorithm

The GBS algorithm consists of three components: squeezing gates, linear interferometer, and photon detectors. The network graph, mathematically represented as an adjacency matrix, was decomposed using Autonne–Takagi factorisation into a diagonal matrix and a unitary matrix. The values in the diagonal matrix specify the parameters of the squeezing gates, whereas the unitary matrix specifies the operation of the linear interferometer. For the photon detectors, threshold detections were used, meaning that instead of counting the exact number of photons in each detector, we only consider whether any photons were detected. Such implementation is known as Torontonian (Quesada et al., 2018).

Sampling was implemented using the torontonian_sample_graph() method from the Walrus library (Quesada et al., 2018), corresponding to the classical simulation of GBS sampling. The total mean photon number was set to be n, the dimension of the corresponding adjacency matrix.

The number of shots required was determined by test-running GBS with different multiples of n. The selected shot number corresponded to the point at which the approximation ratio stabilised at 1 relative to the classical benchmark, and was used for subsequent benchmarking comparisons.

Due to the probabilistic nature of GBS, re-randomisation was carried out to ensure the reliability and robustness of the results. This was achieved by resampling, where the GBS algorithm was run independently with different random seeds (12, 34, 56, 78, and 90).

Subgraph evaluation counts

We determined and compared the number of subgraph evaluations for Brute-force, Greedy Peel, and GBS. A subgraph evaluation is defined here as one computation of subgraph density.

For deterministic algorithms, the number of subgraph evaluations can be calculated theoretically using the number of density computations required to obtain the densest subgraph. This implies that brute-force requires 2^n evaluations corresponding to all possible subgraphs, and that Greedy Peel requires at least $\frac{n(n-1)}{2}$ and at most $\frac{n^2(n-1)}{2}$ evaluations based on the iterative node-removal implementation.

For GBS, the algorithm is probabilistic and hence the number of subgraph evaluations depends on how many subgraphs are sampled. In our case, the number of subgraph evaluations is equivalent to the number of distinct sampled subgraphs, after removing repeated or empty subgraphs.

Runtime measurements for Greedy Peel and GBS

Wall clock timing was performed within the Python scripts to determine the runtime required for Greedy Peel and GBS. Runtime measurements were implementation-dependent and reported separately from theoretical evaluation counts. For GBS, the runtime for sampling and subgraph evaluation were measured separately in order to identify potential bottlenecks of the algorithm.

Development of Uniform Sampling and Approximation Ratio Comparisons

The same number of subgraphs as generated by GBS were produced using uniform sampling. Subgraphs were randomly generated with uniform probability over the subgraph space, and the approximation ratio was then compared to the results from GBS. The same resampling procedure was applied to ensure reliability of the algorithm.

Data availability

The metagenomics and metatranscriptomics data of this work is available in https://doi.org/10.1021/acs.est.2c07813 .

Code availability

The code of this work is available upon request.

Author contributions

Conceptualisation, K.H., P.L.; methodology, K.H., C.L. and P.L.; implementation of the code, K.H.; analysis, K.H., G. Y. A. T., C.L. and P.L.; writing and editing, K.H., G. Y. A. T., C.L. and P.L.; supervision, P.L. All authors reviewed and contributed to the manuscript.

Competing Interests

The authors declare no competing interests.

Funding declaration

443 This research did not receive any specific grant from funding agencies in the public,
444 commercial, or not-for-profit sectors

445

References

Abdel-Hafiz, M., Najafi, M., Helmi, S., Pratte, K. A., Zhuang, Y., Liu, W., Kechris, K. J., Bowler, R. P., Lange, L., & Banaei-Kashani, F. (2022). Significant Subgraph Detection in Multi-omics Networks for Disease Pathway Identification. *Frontiers in Big Data*, *5*, 894632. https://doi.org/10.3389/FDATA.2022.894632/BIBTEX

Abdel-Hamid, A. M., Attwood, M. M., & Guest, J. R. (2001). Pyruvate oxidase contributes to the aerobic growth efficiency of Escherichia coli. *Microbiology*, *147*(6), 1483–1498. https://doi.org/10.1099/00221287-147-6-1483/CITE/REFWORKS

Arrazola, J. M., & Bromley, T. R. (2018). Using Gaussian Boson Sampling to Find Dense Subgraphs. *Physical Review Letters*, *121*(3), 030503. https://doi.org/10.1103/PHYSREVLETT.121.030503/FIGURES/3/THUMBNAIL

Banchi, L., Fingerhuth, M., Babej, T., Ing, C., & Arrazola, J. M. (2020). Molecular docking with Gaussian Boson Sampling. *Science Advances*, *6*(23). https://doi.org/10.1126/SCIADV.AAX1950/SUPPL_FILE/AAX1950_SM.PDF

Deng, Y. H., Gong, S. Q., Gu, Y. C., Zhang, Z. J., Liu, H. L., Su, H., Tang, H. Y., Xu, J. M., Jia, M. H., Chen, M. C., Zhong, H. Sen, Wang, H., Yan, J., Hu, Y., Huang, J., Zhang, W. J., Li, H., Jiang, X., You, L., … Pan, J. W. (2023). Solving Graph Problems Using Gaussian Boson Sampling. *Physical Review Letters*, *130*(19), 190601. https://doi.org/10.1103/PHYSREVLETT.130.190601/FIGURES/4/THUMBNAIL

Everett, L., Wang, L. S., & Hannenhalli, S. (2006). Dense subgraph computation via stochastic search: Application to detect transcriptional modules. *Bioinformatics*, *22*(14). https://doi.org/10.1093/BIOINFORMATICS/BTL260,

Feng, J., Jiang, R., & Jiang, T. (2011). A max-flow-based approach to the identification of protein complexes using protein interaction and microarray data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *8*(3), 621–634. https://doi.org/10.1109/TCBB.2010.78,

Gecse, G., Labunskaite, R., Pedersen, M., Kilstrup, M., & Johanson, T. (2024). Minimizing acetate formation from overflow metabolism in Escherichia coli: comparison of genetic engineering strategies to improve robustness toward sugar gradients in large-scale fermentation processes. *Frontiers in Bioengineering and Biotechnology*, *12*, 1339054. https://doi.org/10.3389/FBIOE.2024.1339054/BIBTEX

Goldberg, A. V. (1984). *Finding a maximum density subgraph*.

Guberovic, I., & Frezza, C. (2024). Functional implications of fumarate-induced cysteine succination. *Trends in Biochemical Sciences*, *49*(9), 775–790. https://doi.org/10.1016/J.TIBS.2024.05.003

Gudapati, N. V. C., Malaguti, E., & Monaci, M. (2021). In search of dense subgraphs: How good is greedy peeling? *Networks*, *77*(4), 572–586. https://doi.org/10.1002/NET.22034

Hamilton, C. S., Kruse, R., Sansoni, L., Barkhofen, S., Silberhorn, C., & Jex, I. (2017). Gaussian Boson Sampling. *Physical Review Letters*, *119*(17), 170501. https://doi.org/10.1103/PHYSREVLETT.119.170501/FIGURES/1/THUMBNAIL

Hayes, C. N., Nakahara, H., Ono, A., Tsuge, M., Oka, S., Hayes, C. N., Nakahara, H., Ono, A., Tsuge, M., & Oka, S. (2024). From Omics to Multi-Omics: A Review of Advantages and Tradeoffs. *Genes 2024, Vol. 15, 15*(12). https://doi.org/10.3390/GENES15121551

489 Hu, H., Yan, X., Huang, Y., Han, J., & Zhou, X. J. (2005). Mining coherent dense subgraphs
490 across massive biological networks for functional discovery. *Bioinformatics*, *21*(suppl_1),
491 i213–i221. https://doi.org/10.1093/BIOINFORMATICS/BTI1049

492 Javier Mater-Sagasta, Sara Marjani Zadeh, & Hugh Turral. (2017). Livestock solutions for climate
493 change. In *Crítica Literaria*. FAO ;
494 https://openknowledge.fao.org/handle/20.500.14283/i8098en

495 Katsyv, A., Schoelmerich, M. C., Basen, M., & Müller, V. (2021). The pyruvate:ferredoxin
496 oxidoreductase of the thermophilic acetogen, Thermoanaerobacter kivui. *FEBS Open Bio*,
497 *11*(5), 1332. https://doi.org/10.1002/2211-5463.13136

498 Kocharovsky, V. V., Kocharovsky, V. V., & Tarasov, S. V. (2022). The Hafnian Master Theorem.
499 *Linear Algebra and Its Applications*, *651*, 144–161.
500 https://doi.org/10.1016/J.LAA.2022.06.021

501 Koutrouli, M., Karatzas, E., Paez-Espino, D., & Pavlopoulos, G. A. (2020). A Guide to Conquer
502 the Biological Network Era Using Graph Theory. *Frontiers in Bioengineering and*
503 *Biotechnology*, *8*, 504360. https://doi.org/10.3389/FBIOE.2020.00034/XML

504 Lanciano, T., Miyauchi, A., Fazzone, A., & Bonchi, F. (2024). A Survey on the Densest Subgraph
505 Problem and its Variants. *ACM Computing Surveys*, *56*(8).
506 https://doi.org/10.1145/3653298;WGROUP:STRING:ACM

507 Li, R., Lee, J. Y., Yang, J. M., & Akutsu, T. (2022). Densest subgraph-based methods for protein-
508 protein interaction hot spot prediction. *BMC Bioinformatics*, *23*(1), 1–12.
509 https://doi.org/10.1186/S12859-022-04996-1/FIGURES/4

Li, Z., Zheng, W., Kong, W., & Zeng, T. (2023). Itaconate: A Potent Macrophage Immunomodulator. *Inflammation 2023 46:4*, *46*(4), 1177–1191. https://doi.org/10.1007/S10753-023-01819-0

Marino, R., Atzori, A. S., D'Andrea, M., Iovane, G., Trabalza-Marinucci, M., & Rinaldi, L. (2016). Climate change: Production performance, health issues, greenhouse gas emissions and mitigation strategies in sheep and goat farming. *Small Ruminant Research*, *135*, 50–59. https://doi.org/10.1016/J.SMALLRUMRES.2015.12.012

O'Neill, L. A. J., & Artyomov, M. N. (2019). Itaconate: the poster child of metabolic reprogramming in macrophage function. *Nature Reviews Immunology 2019 19:5*, *19*(5), 273–281. https://doi.org/10.1038/s41577-019-0128-5

Quesada, N., Arrazola, J. M., & Killoran, N. (2018). Gaussian boson sampling using threshold detectors. *Physical Review A*, *98*(6), 062322. https://doi.org/10.1103/PhysRevA.98.062322

Sejian, V., Bhatta, R., Gaughan, J., Malik, P. K., Naqvi, S. M. K., & Lal, R. (2017). Adapting sheep production to climate change. *Sheep Production Adapting to Climate Change*, 1–29. https://doi.org/10.1007/978-981-10-4714-5_1/TABLES/1

Sun, D. S., & Chang, H. H. (2022). Emerging role of the itaconate-mediated rescue of cellular metabolic stress. *Tzu Chi Medical Journal*, *34*(2), 134–138. https://doi.org/10.4103/TCMJ.TCMJ_79_21

Ternette, N., Yang, M., Laroyia, M., Kitagawa, M., O'Flaherty, L., Wolhulter, K., Igarashi, K., Saito, K., Kato, K., Fischer, R., Berquand, A., Kessler, B. M., Lappin, T., Frizzell, N., Soga, T., Adam, J., & Pollard, P. J. (2013). Inhibition of Mitochondrial Aconitase by Succination in Fumarate Hydratase Deficiency. *Cell Reports*, *3*(3), 689–700. https://doi.org/10.1016/J.CELREP.2013.02.013

Wu, Z., Nguyen, D., Shrestha, S., Raskin, L., Khanal, S. K., & Lee, P. H. (2023). Evaluation of Nanaerobic Digestion as a Mechanism to Explain Surplus Methane Production in Animal Rumina and Engineered Digesters. *Environmental Science and Technology*, *57*(33), 12302–12314. https://doi.org/10.1021/ACS.EST.2C07813/SUPPL_FILE/ES2C07813_SI_005.PDF

Yang, T. Y., & Wang, X. Bin. (2024). Speeding up the classical simulation of Gaussian boson sampling with limited connectivity. *Scientific Reports 2024 14:1*, *14*(1), 7680-. https://doi.org/10.1038/s41598-024-58136-1

Yu, S., Zhong, Z. P., Fang, Y., Patel, R. B., Li, Q. P., Liu, W., Li, Z., Xu, L., Sagona-Stophel, S., Mer, E., Thomas, S. E., Meng, Y., Li, Z. P., Yang, Y. Z., Wang, Z. A., Guo, N. J., Zhang, W. H., Tranmer, G. K., Dong, Y., … Guo, G. C. (2023). A universal programmable Gaussian boson sampler for drug discovery. *Nature Computational Science*, *3*(10). https://doi.org/10.1038/s43588-023-00526-y

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- MSGBSnewrevisedfinalSIv3.docx
- Graphicalabstrac1.docx