TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Stein, J., Ott, D., Nüßlein, J., Bucher, D., Schönfeld, M., & Feld, S. (2023). NISQ-Ready Community Detection Based on Separation-Node Identification. *Mathematics*, *11*(15), Article 3323. https://doi.org/10.3390/math11153323

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

mathematics

MDPI

Article

NISQ-Ready Community Detection Based on Separation-Node Identification

Jonas Stein [1,*], Dominik Ott [1], Jonas Nüßlein [1], David Bucher [2], Mirco Schönfeld [3] and Sebastian Feld [4]

[1] Mobile and Distributed Systems Group, LMU Munich, 80538 Munich, Germany
[2] Aqarios GmbH, 80538 Munich, Germany
[3] Data Modelling & Interdisciplinary Knowledge Generation, University of Bayreuth, 95445 Bayreuth, Germany
[4] Quantum Machine Learning, Delft University of Technology, 2628 CD Delft, The Netherlands
* Correspondence: jonas.stein@ifi.lmu.de

Abstract: The analysis of network structure is essential to many scientific areas ranging from biology to sociology. As the computational task of clustering these networks into partitions, i.e., solving the community detection problem, is generally NP-hard, heuristic solutions are indispensable. The exploration of expedient heuristics has led to the development of particularly promising approaches in the emerging technology of quantum computing. Motivated by the substantial hardware demands for all established quantum community detection approaches, we introduce a novel QUBO-based approach that only needs number-of-nodes qubits and is represented by a QUBO matrix as sparse as the input graph's adjacency matrix. The substantial improvement in the sparsity of the QUBO matrix, which is typically very dense in related work, is achieved through the novel concept of separation nodes. Instead of assigning every node to a community directly, this approach relies on the identification of a *separation-node set*, which, upon its removal from the graph, yields a set of connected components, representing the core components of the communities. Employing a greedy heuristic to assign the nodes from the separation-node sets to the identified community cores, subsequent experimental results yield a proof of concept by achieving an up to 95% optimal solution quality on three established real-world benchmark datasets. This work hence displays a promising approach to NISQ-ready quantum community detection, catalyzing the application of quantum computers for the network structure analysis of large-scale, real-world problem instances.

Keywords: quantum computing; community detection; QUBO; NISQ

MSC: 68Q12

check for updates

Citation: Stein, J.; Ott, D.; Nüßlein, J.; Bucher, D.; Schönfeld, M.; Feld, S. NISQ-Ready Community Detection Based on Separation-Node Identification. *Mathematics* **2023**, *11*, 3323. https://doi.org/10.3390/math11153323

Academic Editor: Jan Sładkowski

Received: 1 July 2023
Revised: 24 July 2023
Accepted: 26 July 2023
Published: 28 July 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In the era of digitization, the amount of collected data is rising rapidly. This poses substantial problems in data analysis as the algorithms employed there typically have superlinear and thus deficient runtime for many relevant datasets. In this article, we investigate a new approach to cope with this problem in the domain of graph structure analysis. Graphs are one of the central data structures used in information theory and find application in a vast range of scientific disciplines [1–3]. The task of identifying the inherent structure of a graph is known as community detection [4]. In practice, the use of corresponding clustering methods allows the discovery of structural information from real-world networks in domains ranging from social science to biology [5–7].

Although no exact definition has been agreed upon, a graph is typically said to inherit a *community structure* if it can be partitioned in a way such that the number of edges within the partitions is higher than the number of edges between the partitions [5]. While some approaches exist that can provably find existing community structures, all of them are NP-hard [8–11]. This indicates a general NP-hardness of community detection and hence poses

a demand for efficient heuristics to acquire sufficiently good solutions in reasonable time. Motivated by recent advancements and promising results in solving NP-hard problems in the field of quantum computing (QC) [12–15], we investigate possible advantages in building such heuristics by utilizing the more powerful algorithmic toolset available in QC.

In general, quantum computers allow the application of quantum mechanical effects to perform computation. Based on the concepts of superposition and entanglement, quantum computers can solve many computational problems provably faster than classical computers [16–18]. In the case of community detection, related work has shown promising results using the popular *modularity maximization* approach [13,19]. Modularity is a measure for the quality of a given partitioning based on comparing the edge distribution of the given graph to the edge distribution of a graph with the same node degree but inheriting no community structure [20]. The more these distributions differ, the higher the modularity, indicating a clearer community structure. While this approach is provably optimal in the sense that no other approach could detect a community structure when modularity maximization cannot [8], its implementation on a quantum computer is cumbersome, especially for current quantum computers.

Present implementations of modularity maximization on quantum computers make use of the quadratic nature of the modularity [13,19]. Simulating the time evolution of a specific quantum physical system, i.e., typically the transverse field Ising Model under adiabatic time evolution, a quantum heuristic solver for quadratic unconstraint binary optimization (QUBO) problems (e.g., modularity maximization) can be implemented on a quantum computer [13,21]. Even though no quantum speedups where proven for solving NP-hard optimization problems with this approach yet, many cases of potential scaling advantages have been identified, with modularity maximization being one of them [13–15].

A critical limitation of the established quantum modularity maximization approach hindering its execution on near-term quantum hardware is the size of the search space in optimization. Scaling linearly in the number of nodes and the number of communities, the required amount of quantum bits (qubits) needed for representing a specific solution quickly exceeds the number of qubits available in present noisy intermediate scale quantum (NISQ) hardware [22].

Motivated by these results, we develop a novel approach to community detection: **C**ommunity **D**etection based on **S**eparation-**N**ode identification (CDSN). This approach is specialized for (quantum heuristic) QUBO solving that uses a smaller search space than the state-of-the-art quantum modularity maximization approach [13]. This objective led to the sociologically inspired approach of defining a community by its extreme ends, similar to, e.g., differentiating political parties by their position on the left–right spectrum. For graphs, we translate this idea to the existence of what we later define as a *bijective set of separation nodes*. The removal of the nodes contained in this set then yields connected components, which represent the "cores" of the communities. We subsequently conduct experiments that indicate that this essentially solves the computationally hard part of the community detection problem, as the community assignment for the separation nodes can typically be obtained using a greedy optimizer.

This idea allows for a quantum–classical hybrid detection of communities while merely using one qubit for every node in the graph with a single call to a QUBO solver. We show empirically that such a set of separation nodes can be found for graphs inheriting community structure and introduce a quantum heuristic approach to find them, constituting a proof of concept.

This paper is structured in the following way: In Section 2, we describe the current state of the art of quantum community detection; in Section 3, the separation node set approach to (quantum) community detection is introduced, which is then evaluated in Section 4, and then the findings are concluded in Section 5.

2. Background

With the advent of quantum optimization heuristics like quantum annealing, possible quantum advantages have been explored for many optimization problems [23]. Easily allowing for a binary encoding of solutions and showing promising performance, community detection quickly became a popular problem in quantum optimization [24].

Representing community detection natively as a QUBO problem in the basic case of partitioning into $k = 2$ communities, modularity maximization was the first approach used in quantum-computing-based community detection [19]. For a given graph $G = (V, E)$, the modularity of a partitioning into $V_0 = \{v_i \in V \mid x_i = 0\}$ and $V_1 = \{v_i \in V \mid x_i = 1\}$ according to $x = \left(x_1, \ldots, x_{|V|}\right) \in \{0,1\}^{|V|}$ is given by

$$\frac{1}{2|E|} \sum_{ij} \left(a_{ij} - \frac{d_i d_j}{2|E|} \right) x_i x_j, \tag{1}$$

where given node degrees $d = \left(d_1, \ldots, d_{|V|}\right)$ and a_{ij} denote the entries of the adjacency matrix A of G. Straightforward calculations yield the resulting QUBO matrix $Q = A - \frac{dd^\intercal}{2|E|}$ which is sufficient to apply practically all currently available quantum optimization heuristics.

This approach to can be generalized to $k > 2$ communities by introducing one-hot encoding [13]. Here, the community assignment of a node $v_i \in V$ is encoded by a k-dimensional bit string $x_i = \left(x_i^{(1)}, \ldots, x_i^{(k)}\right)$ with $x_i^{(l)} = 1$ and $x_i^{(m)} = 0 \; \forall m \neq l$ if the node v_i is assigned to community l. The resulting optimization term is hence given by

$$\frac{1}{2|E|} \sum_{ij} \left(a_{ij} - \frac{d_i d_j}{2|E|} \right) \left(\sum_l x_i^{(l)} x_j^{(l)} \right). \tag{2}$$

In order to formulate this as a QUBO problem, we have to add a suitably weighted penalty term $P(x)$ (for details, see [25]) to the optimization term to indirectly enforce the one-hot encoding by $P(x) = 0$ if every node is assigned to exactly one community and $P(x) > 0$; otherwise,

$$P(x) = \sum_i \left(1 - \sum_l x_i^{(l)} \right)^2. \tag{3}$$

Apart from capitalizing on the inherent QUBO form of modularity maximization, many other quantum-computing-based approaches to community detection like Quantum Genetic Algorithms and Quantum Walks have been proposed in the recent literature [26,27]. A particularly promising approach for near-term application on large graphs is based on exploiting the quadratic nature of regularity checking related to Szemeredi's Regularity Lemma (SRL) [28]. While similar to our approach, as the solved QUBO problems only involve $|V|$ qubits, it works fundamentally differently, as communities are identified iteratively. In essence, the algorithm proposed in [28] executes the following steps:

1. Randomly split the given graph $G = (V, E)$ into two equally sized partitions $A \dot{\cup} B = V$ and delete all edges inside the partitions to yield a bipartite graph.
2. Find subsets $X \subseteq A$ and $Y \subseteq B$ such that $X = \{v_i \in A \mid s_i = 1\}$ and $Y = \{v_j \in B \mid s_j = 1\}$ where $s = \left(s_1, \ldots, s_{|V|}\right)$ is the solution to the quadratic program given by

$$\underset{s \in \{0,1\}^{|V|}}{\arg\min} \sum_{\substack{v_i \in A \\ v_j \in B}} \big(d(A, B) - a_{ij} \big) s_i s_j. \tag{4}$$

Here, $d(V_1, V_2)$ denotes the link density of two disjoint sets V_1, V_2 given by $\frac{e(V_1,V_2)}{|V_1||V_2|}$ and $e(V_1, V_2)$ represents the number of edges connecting V_1 and V_2.

3. Identify $C := X \cup Y$ to be a community and repeat Steps 1 and 2 for the subgraph induced on G by $V \setminus C := \{v \in V \mid v \notin C\}$.

While this approach has a solid graph theoretic foundation, the high number of needed solver calls and the dense QUBO matrix still pose nontrivial hardware execution challenges in the NISQ era.

Aiming to minimize the demands to the QUBO solver, we propose a radically different approach that only needs a single QPU call and whose QUBO matrix is topologically identical to the adjacency matrix of the given graph and is thus equally sparse. The approach presented in this work essentially purifies a solution of a relaxed community detection problem, i.e., the final community structure is represented by the solution of a QUBO problem which is based on classically computed, probabilistic community assignments for each node. While we introduce a particularly efficient approach to calculate the needed input for the QUBO problem, many other approaches to relaxed community detection have been proposed in related work like semidefinite programming or convexification [29–32].

As derived in detail in the next section, our approach requires a solution for a novel relaxation of the community detection problem as input to the QUBO problem formulation. In essence, our approach demands an estimate value for each edge, specifying whether it connects nodes belonging to different or the same communities. While such estimates could in principle be computed based on the output of solvers for the relaxed community detection problem by using, e.g., the KL-divergence of the community affiliations of neighboring nodes, we introduce a specialized estimation method tailored to this task. Notably, metrics like the edge betweenness centrality [33] also do not yield satisfactory results for our approach, as the difference in values between separation and non-separation edges is seemingly too small.

3. Proposed Model

In the following, we explore the idea of performing community detection based on finding a suitable set of nodes separating the communities as defined in Definition 1 in a rigorous mathematical manner. Meeting the demand from the derived QUBO formulation for a separation edge estimator, we subsequently introduce a promising heuristic approach based on the concept of modularity.

3.1. Separation-Node Sets

The approach presented in this paper consists of two steps:

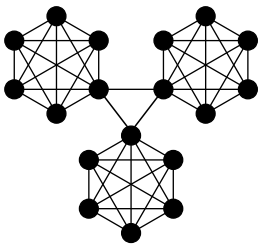
(1) Identifying a set of nodes separating communities and thus revealing the fundamental community structure (see Sections 3.4 and 3.5).
(2) Classifying the community of each separation node to finalize community detection (see Section 3.6).

Using either a trivial, greedy approach introduced in Section 3.6 or a slight adaptation of the well-known QUBO-formulation of modularity maximization [34] to perform (2), the main objective of this paper is the development of a QUBO-approach realizing (1). To provide a more formal definition of (1), we now introduce the concept of *separation-node sets*. In the following, we use $\mathcal{S}$ to denote the set of all separation-node sets.

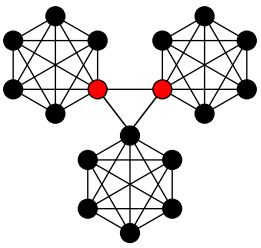
Definition 1. *For a graph $G = (V, E)$ and a ground truth community structure C partitioning V, we call $S \subseteq V$ a set of separation nodes if the connected components $\overline{S}_i$ partitioning the graph induced by $V \setminus S$ are distributed such that $\{\overline{S}_i\}_i$ is a refinement of C.*

Equivalent to this definition, one could also demand the existence of a refinement map $\phi : \mathcal{P}(V) \to \mathcal{P}(V)$ mapping each connected component $\overline{S}_i \subseteq V$ onto a community $\phi(\overline{S}_i) = C_j \in C$ such that $\overline{S}_i \subseteq C_j$. Utilizing the notion of separation-node sets, (1) can be formulated as finding the smallest set of separation nodes whose associated refinement map ϕ is ideally bijective. An example of a set of separation nodes satisfying these conditions is depicted in Figure 1b, which is part of Figure 1 displaying the proposed approach. As it

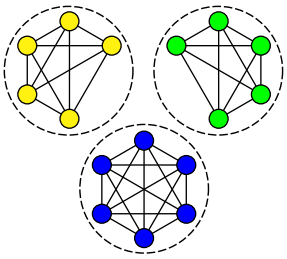
becomes apparent in the evaluation, such well-behaved separation node sets can also be found in real-world datasets.



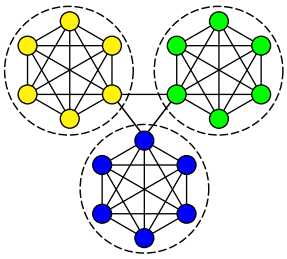
(**a**) Exemplary graph consisting of three connected cliques.



(**b**) Identification of a set of separation nodes (marked in red).



(**c**) Removal of the set of identified separation nodes and identification of the resulting connected components.



(**d**) Classification of the community of all identified separation nodes.

Figure 1. Outline of the workflow for the proposed approach of community detection via separation-node identification. The computationally expensive tasks of identifying a set of separation nodes (**b**) and classifying the communities for these nodes (**d**) are performed using quantum computing, while the computationally cheap tasks of removing the classified separation-nodes and identifying the resulting connected components (**c**) are performed classically.

The surjectivity of ϕ ensures that each community becomes detected and its injectivity ascertains that no communities are split. In the following, we call separation-node sets injective, surjective or bijective if the respective refinement function satisfies these conditions. In order to formulate a QUBO problem where the optimal solution represents the minimal separation-node set, we start by stating an alternate, more convenient definition of minimal separation-node sets.

Theorem 1. *For an adequate penalty term $P : \{0,1\}^{|V|} \to \mathbb{R}_0^+$ ensuring the separation-node set properties, the following equation states an equivalent definition of the set containing all minimal separation-node sets $\mathcal{S}_{min}$:*

$$\mathcal{S}_{min} = \left\{ \bigcup_{\substack{v_i \in V \\ x_i = 0}} v_i \;\middle|\; x = \underset{x \in \{0,1\}^{|V|}}{\arg\min}\, 2P(x) - \sum_{v_i \in V} x_i \right\}. \tag{5}$$

Here, we used $x \in \{0,1\}^{|V|}$ as a 0-flag for separation nodes, a_{ij} to denote the entries of the adjacency matrix, $c : V \to C$ as a mapping of nodes to their ground truth community and the Kronecker delta δ_{xy}. For a penalty term P ensuring the validity of the separation-node set definition by penalizing incident node pairs from strictly different communities where neither node is element of the sought-after separation-node set, the following definition is a viable option:

$$P(x) := \sum_{(v_i, v_j) \in V^2} a_{ij} \Big(1 - \delta_{c(v_i)c(v_j)}\Big) x_i x_j.$$

Proof. See Section 3.2. □

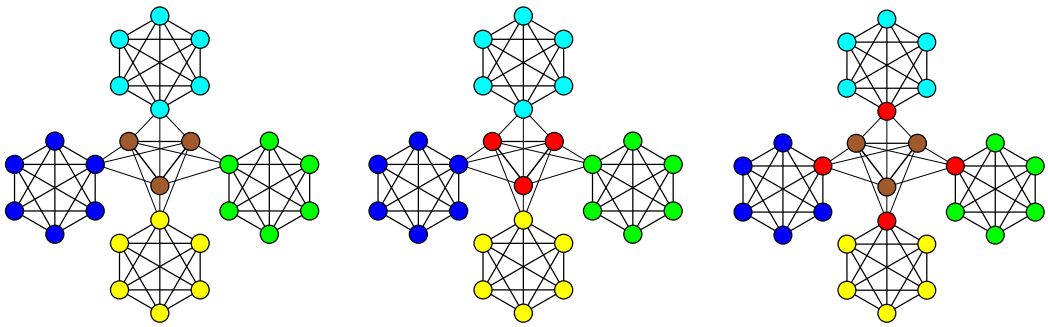
Therefore, the task of finding the smallest set of separation-nodes for any given graph is native to the concept of QUBO. Its formulation can be reduced to approximating $\delta_{c(v_i)c(v_j)}$ for incident node pairs $v_i, v_j \in V$. This can be understood as calculating the probability of an edge being an interconnection of adjacent nodes belonging to different communities, or, more formally, a *separation edge*.

Most interestingly, we can show that solving the QUBO problem stated in Equation (5) is NP-hard for a specific estimator. To see this, we start by observing a substantial similarity of our QUBO formulation with the QUBO formulation of the Max-Clique problem as stated in [35],

$$\underset{x\in\{0,1\}^{|V|}}{\arg\min}\ 2 \sum_{(v_i,v_j)\in V^2} (1-a_{ij})x_i x_j - \sum_{v_i\in V} x_i \tag{6}$$

for a given graph $G = (V, E)$ and its corresponding adjacency matrix A with entries a_{ij}. Choosing the estimator $s : V \times V \to \{0, 1\}$ by $s\big((v_i, v_j)\big) := a_{ij}$, it becomes apparent that the QUBO formulations are identical if we specify the use of a complete graph of size $|V|$ as an input to our QUBO formulation. Leaving an extensive mathematical analysis of the NP-hardness for more realistic estimators to future work, this shows that the problem of finding a minimal separation-node set is NP-hard when treating the estimator as a variable. This result supports the pursuit of the proposed approach of using quantum computing in order to find a minimal separation-node set.

Returning to the initial goal of finding bijective separation-node sets, we now expedite their surjectivity. A significant discovery regarding surjectivity is illustrated in Figure 2, showing no-free-lunch when using Theorem 1 to find surjective separation-node sets. This necessitates the addition of a penalty term to the QUBO formulation in order to ensure surjectivity when building upon Theorem 1. For the formulation of a suitable penalty term, see Section 3.3.



(**a**) A graph and its community structure.

(**b**) Minimal separation-node set consisting of all nodes marked in red.

(**c**) Minimal surjective separation-node set consisting of all nodes marked in red.

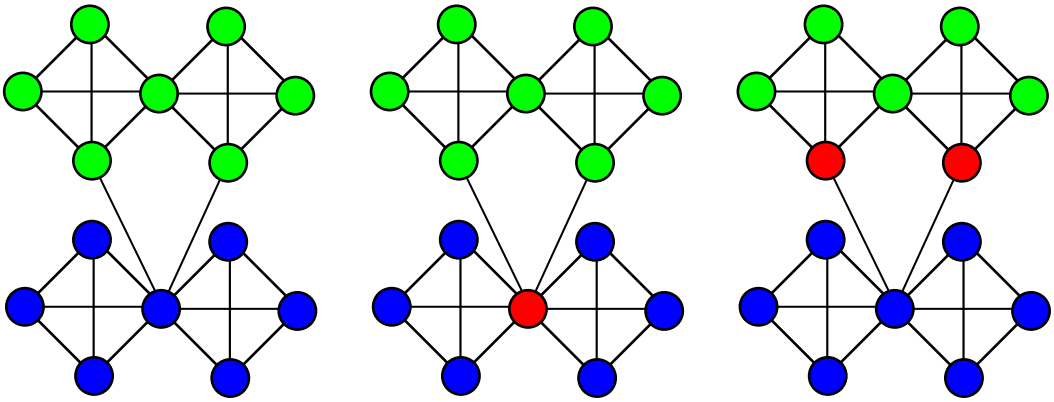
Figure 2. Counterexample proving no-free-lunch when using Theorem 1 to find surjective separation-node sets.

As our formulation results in a PUBO (polynomial unconstrained binary optimization) problem of degree $\mathcal{O}(\log_2 |V|)$, we conjecture that this constraint cannot be realized in a QUBO form without the addition of ancillary variables. Using the standard quadratization approach with the Rosenberg polynomial [36], a QUBO formulation of this term demands superpolynomially many ancillary variables, i.e., $\mathcal{O}\Big(|V|^{2\log_2 \log_2 |V|}\Big)$. In the context of quantum annealing, this scaling beyond a quadratic number of qubits makes the surjective separation-node approach overly complex compared to the standard modularity maximization. In the gate model, the QAOA can be used to solve PUBO problems in principle (see, e.g., [37]), but as current hardware limitations prohibit adequate evaluation, we leave the exploration of the surjectivity constraint to future work.

As a consequence of not enforcing surjectivity, there exists a possibility that the number of communities is incorrect after step (1) of detecting the fundamental community structure by separation-node set identification. Modifying step (2) slightly, this could in principle be compensated by iteratively increasing the number of possible communities until no further improvement of the modularity can be achieved. A clever way to do this could be the elbow method known in clustering [38]. For the alternative greedy approach for the second step (2), the possibility of merging communities could be allowed.

Fortunately, the conducted experiments show that topological structures precluding free lunch for property of surjectivity are scarce in practice. Therefore, further, we omit the explicit demand for surjective separation-node sets.

Analog to the surjectivity, there exist graph topologies like the one displayed in Figure 3 showing no-free-lunch when using Theorem 1 to find injective separation-node sets. Hence, it appears necessary to ensure injectivity explicitly using a penalty term when building upon Theorem 1 in principle, as well. The formulation of such a penalty term also turns out to be rather tedious, as can be seen in Lemma 6 of Section 3.2. In this case, we end up with an even higher-degree PUBO problem for the injectivity than for the surjectivity. Luckily, compared to the surjectivity, the injectivity of a separation-node set is of less importance, as the second step (2) could easily be adapted to cope with this. Analog to the case of surjectivity, we observe such topological structures preventing free lunch quite rarely in conducted experiments, resulting in the analog dismissal of an explicit demand for the separation-node sets to be injective in practice.



(**a**) A graph and its community structure.

(**b**) Minimal separation-node set consisting of all nodes marked in red.

(**c**) Minimal injective separation-node set consisting of all nodes marked in red.

Figure 3. Counterexample indicating no-free-lunch when using Theorem 1 to find injective separation-node sets.

In summary, the apparent infrequence of topological structures preventing free lunch regarding bijectivity renders the QUBO formulation stated in Theorem 1 to be a well-founded starting point for the proposition of QUBO-based community detection via separation-node sets.

While this approach provides exact results for a perfect classification of separation edges, it fully relies on a suitable estimation heuristic. Although many known measures for various edge properties exist (as described in Section 2), none showed to be entirely suitable for detecting separation edges according to pretesting conducted for this paper. Consequently, we now motivate a novel approach tailored for exactly this task based on the concept of modularity.

3.2. Proving Theorem 1

In the following, we provide a proof for Theorem 1, which states equation

$$\mathcal{S}_{min} = \left\{ \bigcup_{\substack{v_i \in V \\ x_i = 0}} v_i \;\middle|\; x = \underset{x \in \{0,1\}^{|V|}}{\arg\min}\, 2P(x) - \sum_{v_i \in V} x_i \right\} \tag{7}$$

where, by definition, we have

$$\mathcal{S}_{min} := \left\{ S \in \mathcal{S} \mid |S| \leq \left|S'\right| \forall S' \in \mathcal{S} \right\}, \tag{8}$$

$$P(x) := \sum_{(v_i, v_j) \in V^2} a_{ij} \Big(1 - \delta_{c(v_i)c(v_j)} \Big) x_i x_j. \tag{9}$$

Aiming to prove "$\subseteq$" and "$\supseteq$" individually, we first prove some lemmata.

Lemma 1. *All $x \in \{0,1\}^{|V|}$ satisfying $P(x) = 0$ represent sets of separation nodes.*

Proof. Let x be a binary vector such that $P(x) = 0$ and let S be the corresponding set of nodes. In order to prove the desired statement by contradiction, assume $S \notin \mathcal{S}$, which is equivalent to the existence of a connected component of the graph induced by $V \setminus S$ not being a subset of one community. Then, at least two nodes $v_i, v_j \in V$ must exist that are connected via a path and belong to different communities. On this path, there must exist two adjacent nodes the belong to different communities with neither of them being an element of S. Therefore, $P(x)$ must be bigger than 0, yielding a contradiction. □

Lemma 2. *The following equation states an alternative definition of the set containing all sets of separation nodes.*

$$\mathcal{S} = \left\{ \bigcup_{\substack{v_i \in V \\ x_i = 0}} v_i \;\middle|\; P(x) = 0 \right\}. \tag{10}$$

Proof. Using Lemma 1 to show "$\supseteq$", we now show "$\subseteq$". Let $S \in \mathcal{S}$ be an arbitrary separation-node set and x the corresponding binary vector 0 flagging the nodes belonging to S. Assuming $P(x) \neq 0$, at least two adjacent nodes $v_i, v_j \in V$ belonging to different communities exist following the definition of P. These nodes subsequently belong to the same connected component $\overline{S}_i$ of the graph induced by $V \setminus S$ implicating that no community can exist that resembles a superset of the nodes inducing the connected component $\overline{S}_i$. Therefore, S cannot be a set of separation nodes as the corresponding refinement map cannot exist, yielding a contradiction and showing $P(x) = 0$. □

Lemma 3. *For every $S \subset V$ satisfying $P(x) > 0$, there exists a superset $\tilde{S} \supset S$ such that $\tilde{x}^T Q \tilde{x} < x^T Q x$, with Q defined such that $x^T Q x = 2P(x) - \sum_{v_i \in V} x_i$ and $\tilde{x}$ corresponding to $\tilde{S}$.*

Proof. Let $S \subset V$ be a set of nodes such that the corresponding penalty term $P(x)$ is bigger than zero. This implies the existence of a pair of incident nodes $v, w \in V$ being part of the same community while neither $v \in S$ nor $w \in S$. Then, set $\tilde{S} := S \cup \{v\}$ (without loss of generality, we could also define $\tilde{S} := S \cup \{w\}$ while achieving the same) has a smaller QUBO value compared to S: with a decrement of at least 4 in the the penalty term (i.e., taking its weighting of 2 into account) and an increment in the cost function (i.e., the sum of the x_i's) of 1, we obtain $\tilde{x}^T Q \tilde{x} \leq x^T Q x - 1$, completing the proof. □

Corollary 1. $x = \arg\min_{x \in \{0,1\}^{|V|}} 2P(x) - \sum_{v_i \in V} x_i \Rightarrow P(x) = 0$.

Proof. This result follows directly from the application of Lemma 3, as $P(x) \neq 0$ would violate the minimality property of x. □

With these lemmata, we are now ready to prove Theorem 1.

Proof. Let $Q \in \mathbb{R}^{|V| \times |V|}$ be defined such that

$$x^T Q x = 2P(x) - \sum_{v_i \in V} x_i. \tag{11}$$

We start by proving "$\subseteq$". Let $S \in \mathcal{S}_{min}$ and x its corresponding 0-flag vector; then, we know by Corollary 1 that $P(x) = 0$. Therefore, $x^T Q x = |S| - |V| =: \overline{s_{min}}$. It is sufficient to show that $\overline{s_{min}} = \min_{x \in \{0,1\}^{|V|}} - \sum_{v_i \in V} x_i + 2P(x)$. For this, we assume that there exists an $\widetilde{x}$ such that $\widetilde{x}^T Q \widetilde{x} < \overline{s_{min}}$. Now, as P maps onto $\mathbb{N}_0$, two possibilities exist:

1. $P(\widetilde{x}) = 0$ and the separation-node set $\widetilde{S}$ is smaller than S;
2. $P(\widetilde{x}) > 0$ and the separation-node set $\widetilde{S}$ is much smaller than S.

As we can see using Lemma 3, we can reduce the latter case to the former case by iteratively eradicating all penalties. Now, using Corollary 1, $\widetilde{S}$ is a separation-node set, and by definition of $\widetilde{S}$, $\left|\widetilde{S}\right| < s$, yielding a contradiction to the minimality of $\mathcal{S}_{min}$ and thereby proving "$\subseteq$".

We now prove "$\supseteq$". Let $x^* := \arg\min_{x \in \{0,1\}^{|V|}} 2P(x) - \sum_{v_i \in V} x_i$ and let S^* be the node set corresponding to x^*. As we can see using Lemma 3, $P(x^*)$ must be zero, otherwise x^* could not be minimal in the sense of satisfying its definition. Therefore, S^* is a separation-node set according to Lemma 2. Assuming $S^* \notin \mathcal{S}_{min}$ yields $|S^*| \neq |S|$ for an arbitrary $S \in \mathcal{S}_{min}$, two cases are possible:

1. $|S^*| < |S|$;
2. $|S^*| > |S|$.

The former yields a contradiction to $\mathcal{S}_{min}$ being minimal and the latter yields a contradiction to the minimality of x^*. □

3.3. Constructing Penalty Terms for the In- and Surjectivity Constraints

In this section, we formulate penalty terms realizing the in- and surjectivity constraints for separation-node sets. Instead of solving this seemingly non-straightforward task directly, we realize the individual constraints with terms that are larger than 0 if the constraint is satisfied and 0 otherwise. Exploiting the possibilities of PUBO, we show that the respective terms can be used to build penalty functions using a moderate amount of ancillary variables.

Lemma 4. *The separation-node set associated with the* 0*-flag vector* $x \in \{0,1\}^{|V|}$ *is surjective if* $\sum_{v_j \in V} \delta_{c(v_i)c(v_j)} x_j > 0$ *for all* $v_i \in V$.

Proof. This equivalence can be observed by recognizing that the sum is restricted to yield non-negative terms, which are nonzero if at least one node belonging to the same community as v_i is not part of the separation node set associated with x for all v_i. □

As the heuristic approaches presented in this work only allow for the estimation of $\delta_{c(v_i)c(v_j)}$ for adjacent node pairs $(v_i, v_j) \in E$, no direct estimation for $\delta_{c(v_i)c(v_j)}$ seems accessible. However, we can use the estimation of $\delta_{c(v_i)c(v_j)}$ for adjacent node pairs to estimate $\delta_{c(v_i)c(v_j)}$ for non-adjacent node pairs:

$$\delta_{c(v_i)c(v_j)} = \operatorname{sgn} \sum_{\overline{p} \in p(v_i, v_j)} \prod_{k=1}^{dim(\overline{p})-1} \delta_{c(\overline{p}_k)c(\overline{p}_{k+1})}. \tag{12}$$

Here, $p(v_i, v_j)$ denotes a function that returns the set of all simple paths $(v_{\pi(1)}, \dots, v_{\pi(l)})$ between v_i and v_j. To allow for this convenient notation, we introduce $\pi : V \to \{1, \dots, |V|\}$ as the projection, mapping the indices of the path entries of the elements of $p(v_i, v_j)$ to their global indices from $\{v_1, \dots, v_n\} \in V$. In practice, these paths could be found using techniques presented in [39].

The value inside the sign function resembles the number of simple, intracommunity paths between v_i and v_j. A general upper bound for the number of paths with these properties is $2^{|V|-2}$, i.e., the number of all subsets of V containing v_i and v_j. For practical purposes, this term clearly is unsuitable, as small errors in the estimation of $\delta_{c(v_i)c(v_j)}$ add up very quickly. Neglecting applicability concerns for reasons described in Section 3.1, we now show how to build a PUBO penalty function associated to the surjectivity term used in Lemma 4.

Lemma 5. *Given function $f : \{0,1\}^n \to \{0, \dots, m\}$ for arbitrary $n, m \in \mathbb{N}$ representing a constraint via $f(x) > 0$, the following penalty terms can be used to ensure that $f(x) > 0$ in PUBO:*

$$\begin{aligned} P_1(x,y) &:= \left(f(x) - \sum_{i=0}^{\lceil \log_2(m) \rceil} 2^i y_i \right)^2 \\ &= \begin{cases} 0 & \text{if } \mathrm{y} := \sum_{i=0}^{\lceil \log_2(m) \rceil} 2^i y_i = f(x) \\ > 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{13}$$

$$\begin{aligned} P_2(y) &:= \prod_{i=0}^{\lceil \log_2(m) \rceil} (1 - y_i) \\ &= \begin{cases} 1 & \text{if } \mathrm{y} = 0 \\ 0 & \text{otherwise, i.e., } \mathrm{y} > 0. \end{cases} \end{aligned} \tag{14}$$

Proof. Clearly, $min_{x,y} P_1(x,y) + P_2(y) = 0 \iff f(x) > 0$ and thus $P_1(x,y) + P_2(y) > 0 \iff f(x) = 0$. □

When denoting the surjectivity constraint from Lemma 4 as $f(x)$, we can see that $f(x) < |V|$ for every $v_i \in V$. Therefore, we can use Lemma 5 to formulate a penalty term for the surjectivity constraint at the expense of at most $|V| \lceil \log_2 |V| \rceil$ ancillary qubits.

With these results, we are now ready to formulate the following PUBO penalty term for injectivity.

Lemma 6. *$\overline{\sigma}_{ij}(x)$ is positive for every $v_i \in V$ and $v_j \in c(v_i) \setminus \{v_i\}$ not contained in the separation-node set if the separation-node set associated with the 0-flag vector $x \in \{0,1\}^{|V|}$ is injective, and 0 otherwise.*

$$\overline{\sigma}_{ij}(x) := \sum_{\overline{p} \in p(v_i, v_j)} \prod_{k=1}^{dim(\overline{p})-1} \delta_{c(\overline{p}_k)c(\overline{p}_{k+1})} x_{\pi(\overline{p}_k)} x_{\pi(\overline{p}_{k+1})}. \tag{15}$$

Proof. Here, $\overline{\sigma}_{ij}(x)$ is positive if a simple path $\overline{p} \in p(v_i, v_j)$ between v_i and v_j exists that consists exclusively of nodes assigned to the community of v_i which are not part of the separation-node set, and 0 otherwise. □

Analogously to Lemma 4, we can observe that $\overline{\sigma}_{ij}(x) \leq 2^{|V|-2}$. Thus, we can use Lemma 5 at the expense of less than $\lceil \log_2 |V| \rceil$ ancillary qubits for every single node

pair $v_i \in V$ and $v_j \in c(v_i) \setminus \{v_i\}$. As injectivity demands the positiveness of $\overline{\sigma}_{ij}(x)$ for all node pairs $v_i \in V$ and $v_j \in c(v_i) \setminus \{v_i\}$, $|V|^2\lceil \log_2|V| \rceil$ ancillary qubits suffice to construct a penalty term for injectivity. The selection of appropriate node pairs $v_i \in V$ and $v_j \in c(v_i) \setminus \{v_i\}$ can be performed using the term $x_i x_j \delta_{c(v_i)c(v_j)}$.

3.4. Modularity-Based Separation Edge Estimation

Motivated by the proven optimality of modularity and by the fact that at its core, modularity is based on essentially estimating whether each node pair is likely to belong to the same or different communities, we start by showing how this idea can be used to estimate $\delta_{c(v_i)c(v_j)}$. For this, recall the definition of the entries of the modularity matrix:

$$m_{ij} := \frac{a_{ij} - \mathbb{E}[J_{ij}]}{|E|}. \tag{16}$$

As before, a_{ij} denotes the entries of the respective adjacency matrix while $\mathbb{E}[e_{ij}]$ denotes the expected value of the number of edges between v_i and v_j, J_{ij}. Upon closer inspection, we observe two main cases:

- $m_{ij} > 0$, if less connectivity between v_i and v_j was to be expected, indicating that v_i and v_j likely belong to the same community;
- $m_{ij} < 0$, if more connectivity between v_i and v_j was to be expected, indicating that v_i and v_j likely belong to different communities.

As the matrix entries m_{ij} are normalized to the interval of $[-1, 1]$ by the division with $|E|$, we can see that using proper rescaling to the interval of $[0, 1]$, i.e., via $2(m_{ij} + 1)$, this allows for an estimation of the term $\delta_{c(v_i)c(v_j)}$ in principle.

In practice, however, this approach yields extremely bad estimations, as only the entries m_{ij} of the modularity matrix are relevant that correspond to a given edge $(v_i, v_j) \in E$. For these, it quickly becomes apparent that m_{ij} is typically larger than 0, making this exact approach infeasible in practice. However, these considerations motivate an adaptation of modularity for the estimation of separation edges as proposed further.

3.5. Separation Edge Estimation Based on Edge Neighborhood Connectivity

Exploiting the mathematical structure of modularity for a straightforward separation edge estimation, we now introduce a promising generalization of the previous approach which we name the *neighborhood connectivity* of an edge. Instead of merely taking the direct connection between two nodes into account (i.e., an edge), the neighborhood connectivity of an edge considers connections between the neighborhoods of the nodes. In this context, the neighborhood $N_r(v)$ of a node $v \in V$ is defined as the set of nodes with the shortest path of length r to v.

Based on this idea, we can rephrase the basic case of our generalization, i.e., modularity, as merely counting the number of unique edges on paths of length 1 between the 0-neighborhoods $N_0(v_i)$ and $N_0(v_j)$ of the respective nodes v_i and v_J. The here-proposed generalization introduces the following two new notions:

(1) Consider connections between r-neighborhoods with radius $r \geq 0$;
(2) Consider paths of length 2.

Stating this more precisely in mathematical form, we now define the neighborhood connectivity $\nu_r^{(l)}$ of an edge given a path length l, and a neighborhood size r:

$$\nu_r^{(l)} := \frac{a_r^{(l)} - \mathbb{E}\left(a_r^{(l)}\right)}{n_r^{(l)}}. \tag{17}$$

In this definition, $a_r^{(l)}$ denotes the number of unique edges contained in paths of length l connecting the r-neighborhoods of the given nodes which do not involve nodes or edges

contained by the $(r-1)$-neighborhoods (as this would result in possible double counting of edges). Analogously to the definition of modularity, $\mathbb{E}\left(a_r^{(l)}\right)$ denotes the expected value corresponding to $a_r^{(l)}$, and $n_r^{(l)}$ acts as a normalization factor denoting the highest possible number $a_r^{(l)}$ can assume. Note that we use brackets to differentiate our superscript notation $x^{(y)}$ from the standard notation for exponents x^y.

These values can be calculated based on a simple breadth-first search with depth r iterating of the neighborhood layers while choosing v_i and v_j as starting nodes. As for the expected value calculation, the configuration model [40] has shown to be an adequate choice (which is in line with modularity). For details on this, we refer to our implementation freely available upon request to the authors.

Our preferred method of combining the results into the neighborhood connectivity ν of a given edge based on all $\nu_r^{(l)}$ is the dot product with a weight vector w with entries $w_r^{(l)} \in \mathbb{R}_0^+$ such that their sum equals 1:

$$\nu := \sum_{r=1}^{d} w_r^{(1)} \nu_r^{(1)} + \sum_{r=0}^{d-1} w_r^{(2)} \nu_r^{(2)}. \tag{18}$$

As we know that the standard modularity value is of little use, we choose $w_0^{(1)} = 0$. We consider the remaining weights as hyperparameters, for which $w_0^{(2)} = 0.5 = w_1^{(1)}$ have proven to be suitable values according to a conducted hyperparameter search.

3.6. Assigning the Separation Nodes to Communities

As stated in Section 3.1, we propose two different approaches to assigning the separation nodes to communities, i.e., (1) a greedy strategy and (2) modularity maximization. In the experiments conducted in this paper, the greedy strategy was employed for most experiments. It works as follows:

(1) Count the number of edges to every identified community for each separation node.
(2) Assign the node with the most edges to a single community to that community. In case of a tie, the community that reached the highest number of edges first during the iteration over all adjacent nodes is selected.
(3) Update the counts for every neighboring separation node.
(4) Repeat steps two and three until every separation node is properly assigned to a community.

This algorithm has a runtime of the number of separation nodes S times the number of communities $|C|$, $\mathcal{O}(S \cdot |C|)$ and hence runs very efficiently.

As the results calculated based on the edge neighborhood connectivity did not always show reasonable quality to use this greedy optimizer sensibly, we chose to use standard modularity maximization for these special cases. Fortunately, the well-known QUBO approach to this [19] can be easily adapted to our situation, i.e., by *clamping* the values of the known community assignments, where clamping is to be understood in the same way as it is used in quantum Boltzmann machines [41]. This yields a QUBO problem of size $\mathcal{O}(S \cdot |C|)$, which often can be solved a lot quicker than the original problem, as $S < |V|$ in practice.

4. Evaluation

The evaluation aims at the examination of the validity of the following two claims:

(1) The assignment of separation nodes to their communities is computationally easy given a good enough estimator, i.e., it can be executed accurately in linear runtime with respect to the number of communities for each separation node.
(2) Neighborhood connectivity constitutes a suitable estimator for separation edges, i.e., it can be employed to identify an adequate separation-node set in the here-proposed approach to conduct community detection in practice.

As we show in the following, both claims appear to be valid according to the conducted experiments. Subsequently, we evaluate the concept of edge neighborhood connectivity and the proposed separation-node assignment approaches individually to explore the performance of the employed subroutines in greater detail.

In the following sections, we measure performance using the following units:

- Modularity. For the comparability between different datasets, we use the approximation ratio based on the best known solution. This yields values between 0 (bad) and 1 (good).
- NMI score. The NMI score is used to compare the community assignments with known ground truth. It yields values between 0 (bad) and 1 (good).
- R^2 score. This score is used to estimate predictive performance of the separation edge classification. It yields values between 0 (bad) and 1 (good).

Extending this summary, we provide an overview of all utilized datasets in Table 1.

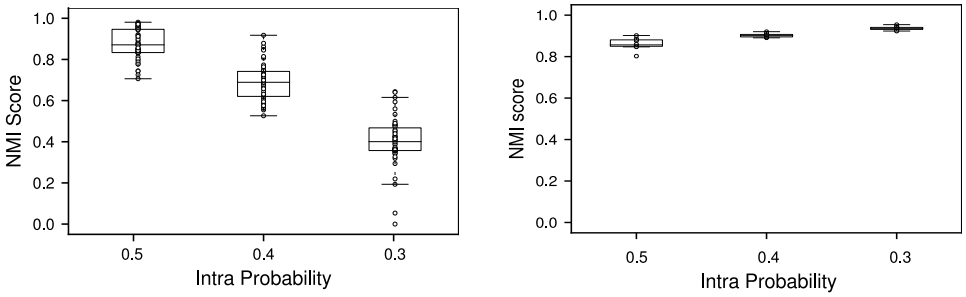
Table 1. Summary of all employed datasets. Note that five different SBM graphs were utilized, each with the same number of nodes and communities, but with varying probabilites of edges being inside communities. For details on all datasets, see the following sections.

	No. of Nodes	**No. of Communities**	**Intra Prob**
SBM graphs	250	7	$[0.75, 0.625, 0.5, 0.4, 0.3]$
Karate Club	24	2	cannot be specified
Dolphins	62	4	cannot be specified
Miserables	77	5	cannot be specified
Protein	83	9	cannot be specified
Books	105	3	cannot be specified

4.1. Evidence That Separation-Node Assignment is Computationally Cheap

For investigating claim (1), we propose to check if the greedy separation-node assignment as described in Section 3.6 is sufficient to assign the nodes of well-behaved separation nodes to the correct communities. If this approach is indeed sufficient to obtain (nearly) perfect solutions, we reason that the claim is most likely valid.

In order to eliminate the possibility of an insufficient separation-node set, we use a synthetic dataset with a known community structure, allowing for the use of a perfect estimator for the separation edges. To find a sufficiently good separation-node set, we utilize a simple simulated annealing approach to solve the associated QUBO as defined in Equation (5). Regarding the synthetic dataset, we choose the stochastic block model (SBM) [42] which is a widely used tool for benchmarking in the realm of community detection. As described in [42], the SBM allows us the generation of ground truth data for a given number of nodes, communities and a specified probability of an edge being inside a community, rather than between different communities. These parameters fully determine the difficulty of the problem instance according to the phase transition of the community detection problem [10]. Aiming to achieve realistic results, we use a graph of size $|V| = 250$ structured into seven equally sized communities with varying intra- and interconnections between the communities resembling three different difficulties according to the phase transition of community detection on SBMs (for details on the phase transition, see [10]). As it becomes apparent in the corresponding Figure 4, the greedy separation-node assignment indeed yields optimal or at least close to optimal results, indicating the validity of claim (1).



(**a**) Our approach. (**b**) SA-based modularity maximization.

Figure 4. This figure shows the Normalized Mutual Information (NMI) score of the presented approach for 50 different graphs each based on ground truth and a perfect separation edge estimator coupled with the greedy separation-node assignment. The NMI score as defined in [43,44] was used, as it resembles a well-proven measure for the accuracy of a community given the ground truth [45]. The different probabilities for intra-community edges in the chosen SBM model resemble different difficulties according to the phase transition known for this model. The lower the stated probability, the harder the problem. The probabilities were chosen such that the hardest graphs barely differed from a null model inheriting no measurable structure up to the hardest that still allowed perfect NMI scores. For this dataset, the phase transition can be calculated to be at a probability of 0.2865 for the intra-community edges. As modularity maximization has been shown to perform very well up until the sharp phase transition (which is not reached here), the constantly good results for the SA based approach appear to be reasonable. Additional tests show a sharp performance drop off to NMI values at around 0.5 for smaller intra-probabilities such as 0.23.

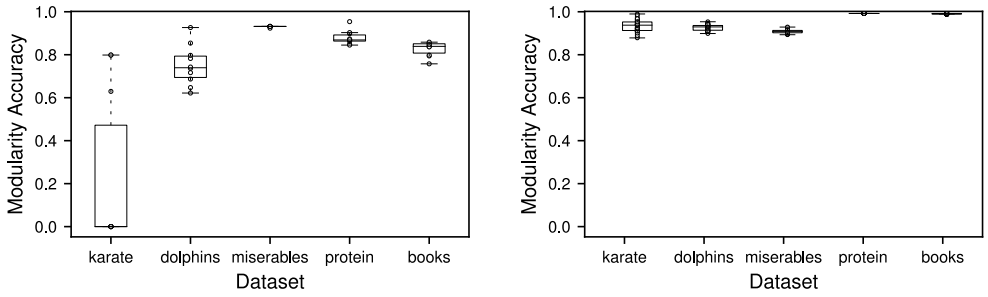
4.2. Neighborhood Connectivity Constitutes a Suitable Estimator for Separation Edges

Having seen solid results for the optimal estimator, we now want to investigate the performance of the here-presented "neighborhood connectivity" approach for real-world data and hence explore claim (2). For this, we choose the greedy separation-node assignment so that the separation-node identification displays the only non-trivial task capable of solving the problem instances. By choosing standard real-world benchmark graphs of varying size, we can observe stable results for most datasets in Figure 5, while often achieving 90 to 95% optimal results. Choosing the state-of-the-art approach of modularity maximization (in this case, based on simulated annealing) as a baseline, we can find that while our approach typically performs slightly worse, there also exist problem instances such as the "miserables" graph in which our approach matches the chosen baseline.

4.3. Evaluating the Performance of Edge Neighborhood Connectivity

Motivated by these proof-of-concept results, we now investigate the performance of the proposed estimator (edge neighborhood connectivity) in order to explore its optimal mission scenario. For this, we again resort to equally formed SBM benchmark graphs with slightly higher intra-community connection probabilities as they offer the comparison with ground truth information. Concretely, we choose these probabilities to be 0.75 for the easy case, 0.625 for the medium case and 0.5 for the hard case, which was the easiest case for the experiments conducted with the perfect estimator (and picked previously as the hardest case to yield perfect results still).

Analogously to the perfect estimator, the identified separation-node sets were all valid and bijective in a small test run on 10 graphs. Switching to the main optimization goal, we now examine the size of the identified separation-node sets for graphs of different difficulty, as displayed in Figure 6. Here, we can see that the sizes of the separation-node sets found are substantially larger than the best-known solution. This becomes especially apparent for easy problem instances. Interestingly, the performance quality increases for harder problems in relative perspective, showing promising scaling behavior.



(**a**) Our approach with greedy separation-node assignment.

(**b**) Simulated annealing-based modularity maximization.

Figure 5. This box plot displays the fraction of the achieved modularity score by the best known solution for selected standard benchmark datasets: (1) the social network of a karate club [46], (2) the social interactions between dolphins [47], (3) the collectively appearing characters in the book "Les Miserables" [48], (4) protein–protein interactions [49] and (5) jointly bought political books [50]. Each graph was analyzed 10 times using simulated annealing. Our approach clearly does not work well for the karate club network. Closer inspections yield the result signifying that the connected components resulting from the found separation-node sets often only consist of single nodes, indicating suboptimality in using neighborhood connectivity for this relatively small dataset.

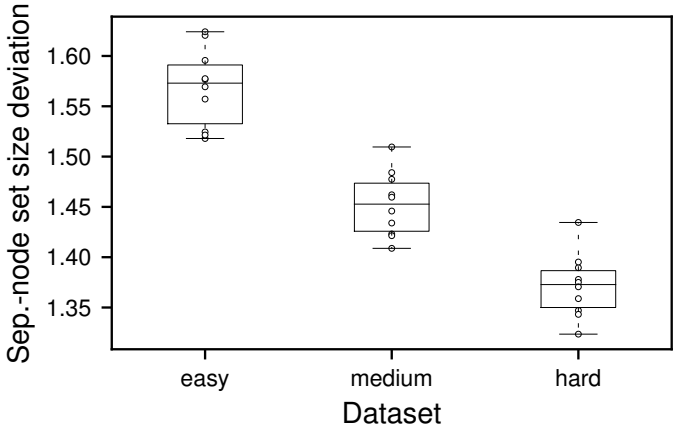


Figure 6. The *y*-axis depicts the deviation factor from the best-known separation-node set in size. Notably, the absolute sizes of the identified separation-node sets are typically similar over the different difficulties, while they rise slightly for larger graphs.

In order to put the results of the developed separation edge estimator based on edge neighborhood connectivity into perspective with an optimal estimator, we now investigate its R^2 score in Figure 7. Interestingly, the worse performance for larger datasets has no impact on the validity and bijectivity of the subsequently identified separation-node set, which is very promising regarding scaling.

4.4. Evaluating the Separation-Node Assignment

Although the separation-node sets found are well behaved, the combination with the greedy separation-node assignment to communities does yield substantially worse results than the perfect estimator. Detailed results to this are displayed in Figure 8.

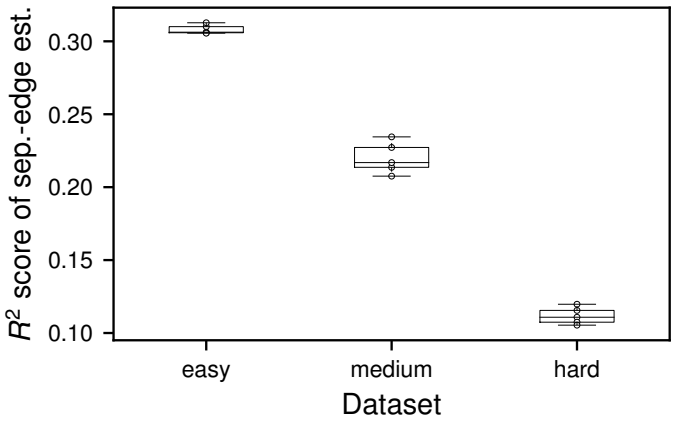
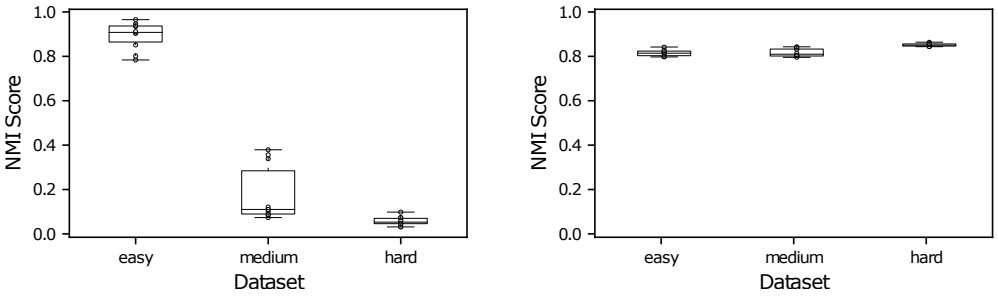


Figure 7. R^2 score of the edge-neighborhood-connectivity-based separation edge estimator. In practice, an R^2 score of 30% implies that merely 30% of the variability of the ground truth has been accounted for. A strict trend towards worse results for harder datasets is clearly visible. This shows that the performance of the estimator decreases for harder problem instances as to be expected while still yielding somewhat accurate results.



(**a**) Our approach with greedy separation-node assignment.

(**b**) Simulated annealing-based modularity maximization.

Figure 8. This figure depicts the normalized mutual information score of the selected SBM benchmark graphs using the greedy assignment of separation nodes to communities. A substantial drop off in performance can be observed for the harder datasets. Meanwhile, as all problem instances are significantly above the phase transition for modularity maximization in these datasets (an intra-prob of 0.2865), our classical baseline easily identifies close to optimal solutions. Notably, however, it is promisingly slightly outperformed by our approach in the case of the easiest dataset.

Subsequent experiments show that the performance of the medium and hard datasets can be improved significantly by exchanging the greedy approach of a simulated annealing-based one, as depicted in Figure 9.

As described in the caption of Figure 9, the employed simulated annealing approach using the QUBO formulation described in Section 3.6 seems to be a suboptimal choice to assign separation nodes to communities. We suspect that the reason for this resides in the large size of the search space for the given problem instances due to the employed one-hot encoding. As identified separation-node sets are typically sized up to 200 nodes (compared to the roughly 120 nodes for the perfect estimator), the search space for the problem instances thus contains roughly $(200 \cdot 7)! = 1400!$ possible solutions, as seven different communities exist. While this case study does not generally rule out the effective applicability of simulated annealing for this problem, it displays first evidence of its suboptimal performance for an important class of problem instances. Analogously to the behavior of other meta heuristics for arbitrary problem instances, one can expect simulated annealing to yield better results for smaller datasets and worse results for larger datasets

in principle. We suggest the development of a more sophisticated heuristic to solve this problem, as the proposed greedy heuristic already shows viable performance.

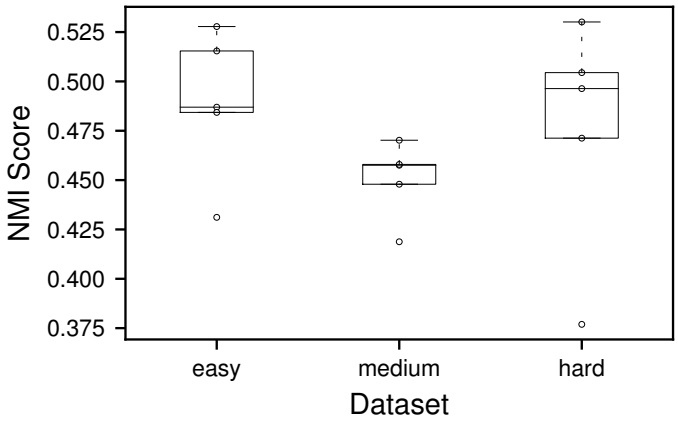


Figure 9. This figure depicts the normalized mutual information score of the selected SBM benchmark graph using a simulated annealing-based approach of assigning the separation nodes to communities. The worse performance for the easy dataset clearly indicates that the chosen simulated annealing approach based on the QUBO as described in Section 3.6 is suboptimal in general.

5. Conclusions

Having set out with the goal of developing a quantum community detection approach that allows for the analysis of large graphs in the NISQ era, we introduced the idea of identifying communities via their borders. The derived separation-node-set-based approach CDSN was shown to yield (close to) optimal results depending on the accuracy of the classical separation edge estimator. The therefore proposed an heuristic approach based on the introduced concept of *edge neighborhood connectivity* enabled for proof-of-concept results on real-world data. In particular, as our approach merely requires $|V|$ qubits and as the corresponding QUBO is as sparse as the input graph $G = (V, E)$, separation-node-based community detection resembles the least hardware demanding quantum computing approach to community detection to the best of our knowledge. The underlying trade-off necessary for this accomplishment is the more demanding classical part of this hybrid approach (i.e., the separation edge estimation). We firmly encourage future work on this heuristic while conjecturing the incorporation of solutions to the relaxed community detection problem as highly beneficial. Expanding the exploration of the proposed approach, a more extensive investigation of the chosen neighborhood setting and the corresponding hyperparameters should help in improving the employed heuristic assumptions and clarify the sensitivity of neighborhood connectivity to different hyperparameters. Furthermore, the exploration of adaptations of similar known metrics like edge betweenness centrality [51] also seem very interesting. Finally, we conclude our approach to be highly promising for accelerating the possibility of solving real-world community detection problems using quantum computers and thus opening up a path towards network structure analysis in big data.

Author Contributions: Conceptualization, J.S., M.S. and S.F.; methodology, J.S., J.N. and D.B.; software, J.S. and D.O.; validation, J.S., M.S. and S.F.; formal analysis, J.S.; investigation, J.S., J.N. and D.B.; resources, J.S.; data curation, J.S. and D.O.; writing—original draft preparation, J.S. and D.O.; writing—review and editing, J.S., D.O., J.N., D.B., M.S. and S.F.; visualization, J.S. and D.O.; supervision, M.S. and S.F.; project administration, J.S.; funding acquisition, J.N. All authors have read and agreed to the published version of the manuscript.

Funding: This publication was created as part of the Q-Grid project (13N16179) under the "quantum technologies—from basic research to market" funding program supported by the German Federal Ministry of Education and Research.

Data Availability Statement: The data presented in this study are openly available in on github at https://github.com/jonas-stein/qcd/graphs, accessed on 23 July 2023.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bondy, J.A.; Murty, U.S.R. *Graph Theory with Applications*; Elsevier: New York, NY, USA, 1976.
2. Mashaghi, A.R.; Ramezanpour, A.; Karimipour, V. Investigation of a protein complex network. *Eur. Phys. J. B Condens. Matter Complex Syst.* **2004**, *41*, 113–121. [CrossRef]
3. Shah, P.; Ashourvan, A.; Mikhail, F.; Pines, A.; Kini, L.; Oechsel, K.; Das, S.R.; Stein, J.M.; Shinohara, R.T.; Bassett, D.S.; et al. Characterizing the role of the structural connectome in seizure dynamics. *Brain* **2019**, *142*, 1955–1972. [CrossRef] [PubMed]
4. Fortunato, S. Community detection in graphs. *Phys. Rep.* **2010**, *486*, 75–174. [CrossRef]
5. Girvan, M.; Newman, M.E.J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826. [CrossRef]
6. Fani, H.; Bagheri, E. Community detection in social networks. *Encycl. Semant. Comput. Robot. Intell.* **2017**, *1*, 1630001. [CrossRef]
7. Vilenchik, D. Simple Statistics Are Sometime Too Simple: A Case Study in Social Media Data. *IEEE Trans. Knowl. Data Eng.* **2020**, *32*, 402–408. [CrossRef]
8. Nadakuditi, R.R.; Newman, M.E.J. Graph Spectra and the Detectability of Community Structure in Networks. *Phys. Rev. Lett.* **2012**, *108*, 188701. [CrossRef]
9. Brandes, U.; Delling, D.; Gaertler, M.; Goerke, R.; Hoefer, M.; Nikoloski, Z.; Wagner, D. Maximizing Modularity is hard. *arXiv* **2006**, arXiv:physics/0608255.
10. Decelle, A.; Krzakala, F.; Moore, C.; Zdeborová, L. Inference and Phase Transitions in the Detection of Modules in Sparse Networks. *Phys. Rev. Lett.* **2011**, *107*, 065701. [CrossRef]
11. Newman, M.E.J. Equivalence between modularity optimization and maximum likelihood methods for community detection. *Phys. Rev. E* **2016**, *94*, 052315. [CrossRef]
12. Arute, F.; Arya, K.; Babbush, R.; Bacon, D.; Bardin, J.C.; Barends, R.; Biswas, R.; Boixo, S.; Brandao, F.G.S.L.; Buell, D.A.; et al. Quantum supremacy using a programmable superconducting processor. *Nature* **2019**, *574*, 505–510. [CrossRef] [PubMed]
13. Shaydulin, R.; Ushijima-Mwesigwa, H.; Safro, I.; Mniszewski, S.; Alexeev, Y. Network Community Detection on Small Quantum Computers. *Adv. Quantum Technol.* **2019**, *2*, 1900029. [CrossRef]
14. Denchev, V.S.; Boixo, S.; Isakov, S.V.; Ding, N.; Babbush, R.; Smelyanskiy, V.; Martinis, J.; Neven, H. What is the Computational Value of Finite-Range Tunneling? *Phys. Rev. X* **2016**, *6*, 031015. [CrossRef]
15. Albash, T.; Lidar, D.A. Demonstration of a Scaling Advantage for a Quantum Annealer over Simulated Annealing. *Phys. Rev. X* **2018**, *8*, 031016. [CrossRef]
16. Grover, L.K. A fast quantum mechanical algorithm for database search. In Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing—STOC'96, Association for Computing Machinery, Philadelphia, PA, USA, 22–24 May 1996; pp. 212–219. [CrossRef]
17. Shor, P.W. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM J. Comput.* **1997**, *26*, 1484–1509. [CrossRef]
18. Lloyd, S. Universal Quantum Simulators. *Science* **1996**, *273*, 1073–1078. [CrossRef]
19. Ushijima-Mwesigwa, H.; Negre, C.F.A.; Mniszewski, S.M. Graph Partitioning Using Quantum Annealing on the D-Wave System. In Proceedings of the Second International Workshop on Post Moores Era Supercomputing, Denver, CO, USA, 12–17 November 2017; PMES'17, Association for Computing Machinery: New York, NY, USA, 2017; pp. 22–29. [CrossRef]
20. Newman, M.E.J.; Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **2004**, *69*, 026113. [CrossRef]
21. Kadowaki, T.; Nishimori, H. Quantum annealing in the transverse Ising model. *Phys. Rev. E* **1998**, *58*, 5355–5363. [CrossRef]
22. Preskill, J. Quantum Computing in the NISQ era and beyond. *Quantum* **2018**, *2*, 79. [CrossRef]
23. Dalyac, C.; Henriet, L.; Jeandel, E.; Lechner, W.; Perdrix, S.; Porcheron, M.; Veshchezerova, M. Qualifying quantum approaches for hard industrial optimization problems. A case study in the field of smart-charging of electric vehicles. *EPJ Quantum Technol.* **2021**, *8*, 12. [CrossRef]
24. Akbar, S.; Saritha, S.K. Towards quantum computing based community detection. *Comput. Sci. Rev.* **2020**, *38*, 100313. [CrossRef]
25. Zahedinejad, E.; Crawford, D.; Adolphs, C.; Oberoi, J.S. Multiple Global Community Detection in Signed Graphs. In *Proceedings of the Future Technologies Conference (FTC) 2019*; Arai, K., Bhatia, R., Kapoor, S., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 688–707.
26. Sedghpour, A.S.; Nikanjam, A. Overlapping Community Detection in Social Networks Using a Quantum-Based Genetic Algorithm. In Proceedings of the Genetic and Evolutionary Computation Conference Companion, Denver, CO, USA, 12–17 November 2017; GECCO '17, Association for Computing Machinery: New York, NY, USA, 2017; pp. 197–198. [CrossRef]
27. Mukai, K.; Hatano, N. Discrete-time quantum walk on complex networks for community detection. *Phys. Rev. Res.* **2020**, *2*, 023378. [CrossRef]

28. Reittu, H.; Kotovirta, V.; Leskelä, L.; Rummukainen, H.; Räty, T. Towards analyzing large graphs with quantum annealing. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 2457–2464. [CrossRef]
29. Chan, E.Y.K.; Yeung, D.Y. A Convex Formulation of Modularity Maximization for Community Detection. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, IJCAI'11, Barcelona, Spain, 16–22 July 2011; AAAI Press: Menlo Park, CA, USA, 2011; Volume 3, pp. 2218–2225.
30. Chen, Y.; Li, X.; Xu, J. Convexified Modularity Maximization for Degree-Corrected Stochastic Block Models. *Ann. Stat.* **2018**, *46*, 1573–1602. [CrossRef]
31. Abdalla, P.; Bandeira, A.S. Community detection with a subsampled semidefinite program. *Sampl. Theory Signal Process. Data Anal.* **2022**, *20*, 6. [CrossRef]
32. Li, W. Visualizing network communities with a semi-definite programming method. Security and privacy information technologies and applications for wireless pervasive computing environments. *Inf. Sci.* **2015**, *321*, 1–13. [CrossRef]
33. Brandes, U. A faster algorithm for betweenness centrality. *J. Math. Sociol.* **2001**, *25*, 163–177. [CrossRef]
34. Negre, C.F.; Ushijima-Mwesigwa, H.; Mniszewski, S.M. Detecting multiple communities using quantum annealing on the D-Wave system. *PLoS ONE* **2020**, *15*, 227538. [CrossRef]
35. Chapuis, G.; Djidjev, H.; Hahn, G.; Rizk, G. Finding Maximum Cliques on the D-Wave Quantum Annealer. *J. Signal Process. Syst.* **2019**, *91*, 363–377. [CrossRef]
36. Rosenberg, I.G. Reduction of bivalent maximization to the quadratic case. *Cah. Cent. D'Etudes Rech. Oper.* **1975**, *17*, 71–74.
37. Stein, J.; Chamanian, F.; Zorn, M.; Nüßlein, J.; Zielinski, S.; Kölle, M.; Linnhoff-Popien, C. Evidence that PUBO outperforms QUBO when solving continuous optimization problems with the QAOA. *arXiv* **2023**, arXiv:2305.03390.
38. Thorndike, R.L. Who belongs in the family? *Psychometrika* **1953**, *18*, 267–276. [CrossRef]
39. Sedgewick, R. *Algorithms in c, Part 5: Graph Algorithms*, 3rd ed.; Addison-Wesley Professional: Hoboken, NJ, USA, 2001.
40. Van Der Hofstad, R. *Random Graphs and Complex Networks*; Cambridge University Press: Cambridge, UK, 2009; Volume 11, p. 60. Available online: https://www.win.tue.nl/~rhofstad/NotesRGCN.pdf (accessed on 23 July 2023).
41. Amin, M.H.; Andriyash, E.; Rolfe, J.; Kulchytskyy, B.; Melko, R. Quantum Boltzmann Machine. *Phys. Rev. X* **2018**, *8*, 021050. [CrossRef]
42. Holland, P.W.; Laskey, K.B.; Leinhardt, S. Stochastic blockmodels: First steps. *Soc. Netw.* **1983**, *5*, 109–137. [CrossRef]
43. Fred, A.L.N.; Jain, A.K. Robust data clustering. In Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, 27 June–2 July 2004; Volume 2.
44. Kuncheva, L.; Hadjitodorov, S. Using diversity in cluster ensembles. In Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583), Hague, The Netherlands, 10–13 October 2004; Volume 2, pp. 1214–1219. [CrossRef]
45. Danon, L.; Díaz-Guilera, A.; Duch, J.; Arenas, A. Comparing community structure identification. *J. Stat. Mech. Theory Exp.* **2005**, *2005*, P09008. [CrossRef]
46. Zachary, W.W. An Information Flow Model for Conflict and Fission in Small Groups. *J. Anthropol. Res.* **1977**, *33*, 452–473. [CrossRef]
47. Lusseau, D.; Schneider, K.; Boisseau, O.J.; Haase, P.; Slooten, E.; Dawson, S.M. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behav. Ecol. Sociobiol.* **2003**, *54*, 396–405. [CrossRef]
48. Knuth, D.E. The Stanford GraphBase: A Platform for Combinatorial Algorithms. In Proceedings of the Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics, SODA '93, Austin, TX, USA, 25–27 January 1993; pp. 41–43.
49. Palla, G.; Derényi, I.; Farkas, I.; Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **2005**, *435*, 814–818. [CrossRef]
50. Newman, M.E.J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8577–8582. [CrossRef]
51. Li, Y.; Li, W.; Tan, Y.; Liu, F.; Cao, Y.; Lee, K.Y. Hierarchical Decomposition for Betweenness Centrality Measure of Complex Networks. *Sci. Rep.* **2017**, *7*, 46491. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.