Noname manuscript No.
(will be inserted by the editor)

A heuristic approach for dividing graphs into bi-connected components with a size constraint

Raka Jovanovic · Tatsushi Nishi · Stefan Voß

Received: date / Accepted: date

Abstract In this paper we propose a new problem of finding the maximal bi-connected partitioning of a graph with a size constraint (MBCPG-SC). With the goal of finding approximate solutions for the MBCPG-SC, a heuristic method is developed based on the open ear decomposition of graphs. Its essential part is an adaptation of the breadth first search which makes it possible to grow bi-connected subgraphs. The proposed randomized algorithm consists of growing several subgraphs in parallel. The quality of solutions generated in this way is further improved using a local search which exploits neighboring relations between the subgraphs. In order to evaluate the performance of the method, an algorithm for generating pseudo-random unit disc graphs with known optimal solutions is created. Computational experiments have also been conducted on graphs representing electrical distribution systems for the real-world problem of dividing them into a system of fault tolerant interconnected microgrids. The experiments show that the proposed method frequently manages to find optimal solutions and has an average error of only a few percent to known optimal solutions. Further, it manages to find high quality approximate solutions for graphs having up to 10,000 nodes in reasonable time.

Keywords Bi-connected Graphs · 2-connected · Breadth First Search · Growth Algorithm · Graph Partitioning · Heuristic

Raka Jovanovic
Qatar Environment and Energy Research Institute, Hamad bin Khalifa University, PO Box 5825, Doha, Qatar
Tel.: +974 4454 7148
Fax: +97444541528
E-mail: rjovanovic@qf.org.qa

Tatsushi Nishi
Graduate School of Engineering Science, Osaka University, Osaka, Japan

Stefan Voß
Institute of Information Systems, University of Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany
and Escuela de Ingenieria Industrial, Pontificia Universidad Católica de Valparaíso, Chile

1 Introduction

The bi-connectivity of graphs is essential for many real-world applications ranging from power distribution systems, communication networks and many others. The reason for this is that such topologies provide a certain level of resistance to failures, and as a result more robust and reliable systems (Zhang et al, 2009; Moraes and Ribeiro, 2013; Goldschmidt et al, 1996; Hu et al, 2010). These benefits are a direct consequence of the fact that such graphs have two disjoint paths connecting any two nodes in the graphs. This property is often called 2-connectivity of graphs. In literature there is differentiation between vertex and edge 2-connected graphs, where the two paths are vertex or edge disjoint, respectively. In practice this means that the vertex/edge 2-connected graph stays connected if any single vertex/edge is removed from it. Vertex 2-connectivity is a stronger property in the sense that every such graph is also 2-edge connected but the reverse is not necessarily true. The term bi-connected graph is used for 2-vertex connected graphs.

Testing if a graph is bi-connected can be done efficiently using the standard algorithm for finding articulation points of a graph based on depth first search in linear time (Hopcroft and Tarjan, 1973a). A similar approach has also been used for finding a 3-connected partitioning of a graph (Hopcroft and Tarjan, 1973b). Other examples of algorithms for testing bi-connectivity of a graph exploit the fact that bi-connected graphs have an ear (Robbins, 1939) or chain (Schmidt, 2013) decomposition. There are several types of mixed integer programs based on multi-commodity flow constraints that are used for exploring bi-connectivity, but such models generally have a large number of variables (Morgan and Grout, 2008). There are alternative formulations containing a lower number of variables but having an exponential number of constraints (do Forte et al, 2013; Buchanan et al, 2015).

For many graph optimization problems, like the weighted vertex cover and the independent set, it is sufficient to solve the problem separately on each of the bi-connected components (Hochbaum, 1993). The most commonly used method for partitioning graphs into bi-connected components is Tarjan's algorithm, which accomplishes this task in linear time (Tarjan, 1972). There are several interesting variations of the original algorithm with similar computational times (Pearce, 2016). The problem with these types of algorithms is that it is hard to modify them to a setting where the subgraphs need to satisfy some additional constraints. The proposed work is focused on developing a method that can partition a graph into bi-connected subgraphs with a maximal allowed size, but the general concept can be adapted to other interesting constraints. In current literature there are many practical problems which are modeled using the problem of partitioning graphs into connected subgraphs. Some examples are applications in surveillance systems (Borra et al, 2015), data clustering (Shafique, 2004) and education (Matic and Bozic, 2012). An interesting group of applications comes from the satisfactory graph partitioning problem, like finding communities within social or biological networks, defense alliances, artificial intelligence development, etc. (Bazgan et al, 2010).

Partitioning problems defined on supply/demand graphs have proven to be essential in modeling systems of interconnected microgrids (Arefifar et al, 2012, 2013; Jovanovic et al, 2015b; Popa, 2013; Jovanovic and Voß, 2016; Morishita and Nishizeki, 2013; Ito et al, 2012).

For many of them it would be reasonable to substitute the constraint of connectivity with bi-connectivity, producing the benefit of higher reliability of the system. This type of extension has frequently been applied to modeling real-world systems based on general graph problems including connectivity. Some examples are power optimization in ad hoc wireless networks (Moraes and Ribeiro, 2013) and facility layout problems (Goldschmidt et al, 1996). A common approach for solving such problems is to start from an approximate solution that is only connected and extend it with additional nodes to achieve bi-connectivity, like in the case of the problem of constructing a 2-connected virtual backbone in wireless networks (Wang et al, 2009).

In case of graph problems containing connectivity constraints, a standard approach is to grow a partial solution by adding neighboring nodes. One example is the method used for finding the minimal connected dominating set of a graph (Jovanovic and Tuba, 2013). The concept of growth has also been extended to problems where a graph is divided into connected components. In case of such problems the general approach is to grow several subgraphs in parallel with the constraint that no vertex can be added to more than one of them. The effectiveness of this type of method is well presented on the problem of partitioning supply/demand graphs into connected subgraphs (Jovanovic et al, 2015b,a, 2016b). It is important to note that the concept of growing a solution gives a high level of flexibility of the method, in the sense of potential applications. Another advantage of growth based algorithms is that they can easily be improved by their extension to metaheuristics like the ant colony optimization (Dorigo and Blum, 2005), the GRASP algorithm (Feo and Resende, 1995) and the variable neighborhood search (Hansen et al, 2010).

In this paper we introduce a new problem of finding the maximal bi-connected partitioning of a graph with a size constraint (MBCPG-SC). For the newly defined problem we show NP-hardness. Because of this a growth based algorithm is developed for MBCPG-SC for finding approximate solutions. More precisely, it is solved using a heuristic procedure that exploits the fact that each bi-connected graph has an ear decomposition. The algorithm is based on a breadth first search (BFS) that also tracks additional properties of the nodes in the BFS tree. The additional information makes it possible to have an efficient way to "grow" a bi-connected subgraph by expanding it with suitable ears. The concept of growing a bi-connected graph using an ear decomposition is also used in case of constructing a fault-tolerant connected set cover problem (Zhang et al, 2009). In this algorithm the best open ears for extending the current solution are found using the idea of shortest cycles in the original graph. A similar approach is used for the minimum 2-connected r-hop dominating set problem (Li and Zhang, 2010). These two approaches compute new open years having some specific properties in polynomial time. In contrast to them, the proposed method computes a new open ear by simply

checking for back edges and comparing pre-calculated values. The new method consists in growing several subgraphs in parallel, with some auxiliary corrections applied to the corresponding BFS trees. To improve the performance of the basic algorithm several methods of randomization are developed. The quality of found solutions is further enhanced using a local search procedure.

The practical objective of the proposed graph problem and the corresponding solution method is its application to real-world problems in the field of smartgrids and ad hoc wireless networks. The proposed graph problem is closely related to the clustering scheme for hierarchical control of wireless networks (Banerjee and Khuller, 2001; Chang et al, 2006). One part of the conducted computational experiments is dedicated to unit disc graphs which are frequently used to model ad hoc wireless networks, although this is not always the best representation (Kuhn et al, 2003). To be able to evaluate the method, an extensive effort has been dedicated to developing an algorithm for generating problem instances with known optimal solutions. Our computational results show that the proposed method is able to find optimal solutions for small graphs. In case of large graphs (10,000 nodes) the method manages to find solutions within a few percent of error, in reasonable time. In the second group of computational experiments our focus is on fault tolerant systems of interconnected micro-grids for which we have solved problem instances based on real-world electrical distribution systems.

The paper is organized as follows. In the second section we give the definition for MBCPG-SC and a proof of NP-hardness. In the following section we present the method for growing a bi-connected subgraph. The third section gives details of the algorithm for parallel growth of several subgraphs and randomization. The next section describes the proposed local search mechanism. In Section 5, we provide details of the data generation mechanism for problem instances with known optimal solutions. Moreover, we discuss results of our computational experiments and provide some conclusions and ideas for future research.

2 Maximal bi-connected partitioning of a graph with a size constraint

The problem is defined on a graph $G(V, E)$, where V is the set of nodes and E is the set of edges. We also define a set $R \subset V$, whose elements will be called root nodes. The aim of the problem is to divide the graph G into a set of subgraphs $\mathit{\Pi} = \{\bar{S}_1, \bar{S}_2, ..., \bar{S}_n\}$, where $n = |R|$, satisfying the constraints given in the following text. The notation S_i will be used for the set of nodes that induces subgraph $\bar{S}_i$. Each of the S_i must contain only one distinct root node $r \in R$. A node $v \in V$ can be an element of at most one S_i. The number of nodes in each of the subgraphs $|S_i|$ is less or equal to some constant M. The last constraint is that each of the subgraphs $\bar{S}_i$ is bi-connected. The goal of the problem is to find $\mathit{\Pi}$ containing the maximal number of nodes in all the

subgraphs. More formally, we wish to maximize the sum:

$$\sum_{i=1..n} |S_i| \tag{1}$$

where each of the subgraphs S_i satisfies

$$|R \cap S_i| = 1 \tag{2}$$
$$|S_i| \leq M \tag{3}$$
$$S_i \cap S_j = \emptyset, i \neq j \tag{4}$$
$$\bar{S_i} \text{ is bi-connected.} \tag{5}$$

In the definition we use notation $|S|$ to indicate the number of nodes in a graph. An illustration of a problem instance and solution for the MBCPG-SC is given in Figure 1. It is important to note that the MBCPG-SC does not produce a partitioning in the strict sense, since some nodes may not be included in any S_i.

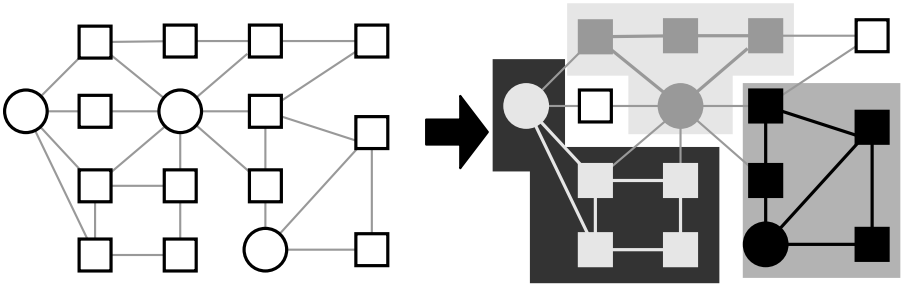

Fig. 1 Example of a problem instance (left) and solution (right) for the MBCPG-SC with maximal allowed size $M = 5$. The circle nodes represent root nodes. Different shades of gray are used for different subgraphs.

The MBCPG-SC is a hard optimization problem in the sense that it is NP-hard. We give a proof by reduction to a restriction of the problem of maximal partition of supply/demand graphs (MPGSD) (Ito et al, 2008). It has been shown that the MPGSD is NP-hard in the case of a graph containing only one supply node and having a star structure (Ito et al, 2008). For a more comprehensive presentation we first give the definition of the MPGSD for star graphs, which is a simplified version of the original problem. It is defined for an undirected star graph $G = (V, E)$ with a set of nodes V and a set of edges E. The center node of the graph s is called a supply node and it has a corresponding supply value *sup*. All the other nodes $u \in D = V \setminus \{s\}$ are called demand nodes, and they have a corresponding demand value which is a positive integer $dem(v)$. The aim is to find a set of nodes $S \subset D$ which satisfies the constraint that the supply *sup* must be greater or equal to the total demand of nodes in S. The goal is to maximize the fulfillment of demands.

In the following text we prove that MBCPG-SC is NP hard by reducing the MPGSD for star graphs to it. Let MPGSD be defined for a star graph G' having a central supply node s with a supply value sup connected to n demand nodes d_i having demand values dem_i. Let us transform this problem to MBCPG-SC in the following way. The parameter M of MBCPG-SC will be set to $sup + 3$ and there will be only one root node r. Let us convert the star graph G' to $\hat{G}(\hat{V}, \hat{E})$ used in MBCPG-SC. First, we include the root node r in $\hat{V}$. The node r is connected to two nodes s (corresponding to the supply node in G') and node e which will be used for some extra edges. Let us remember that the edge set E' of G' consists of edges (s, d_i). For each demand node d_i we will add a path $(s, n_{i,1}, \cdots, n_{i,dem_i})$. Next, we add an additional edge (e, d_{i,dem_i}) for each demand node d_i. An illustration of this conversion is given in Figure 2.

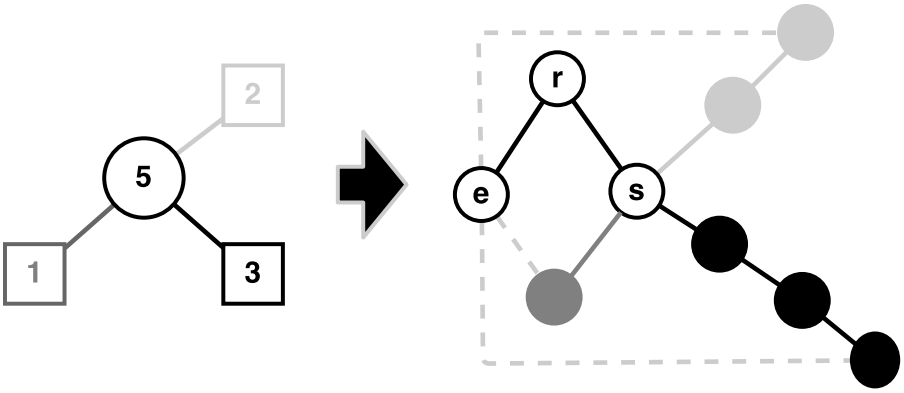

Fig. 2 Illustration of a conversion of a MPGSD for a star graph (left) to MBCPG-SC (right). In the graphic representation for the MPGSD the circle node indicates the supply node and the corresponding supply value. The square nodes indicate the demand nodes and corresponding demand values. Different shades of gray are used to show the changes from nodes (MPGSD) to paths in the MBCPG-SC. In case of the illustration for MBCPG-SC r represents the root node, s the supply node from MPGSD and e the auxiliary node. Dashed lines are used for the auxiliary edges. The size limit MBCPG-SC is $M = 8$.

Let us make a few observations on the solution of a MBCPG-SC on the converted graph G'. First, any solution must contain nodes e, s since there must be at least two disjoint paths from r to any other node. Second if any node $n_{i,j}$ is included in the solution $\hat{S}$ all the nodes $n_{i,1}, \cdots, n_{i,dem_i}$ must be included in it. Let us assume the opposite that node $n_{i,k}$ is not included in the solution and that the nodes $n_{i,j} \in \hat{S}$ for $j \neq k$ are. In such a case the removal of node e from $\hat{S}$ results in all the nodes $n_{i,l}$, where $l > k$, being disconnected from the graph induced by $\hat{S}$. In the same way, the removal of node s from $\hat{S}$ results in all the nodes $n_{i,l}$, where $l < k$, being disconnected from the graph induced by $\hat{S}$. As a consequence any $n_{ik} \in \hat{S}$ is connected to r by two disjoint paths (containing only elements in $\hat{S}$) $(n_{i,k-1}, n_{i,k-2}, \cdots, s, r)$ and by path

($n_{i,k+1}$,$n_{i,k+2}$, $\cdots$, e, r). Note that nodes r, s, $n_{i,1}$,..., n_{i,dem_i}, e form a cycle and as a consequence two disjoint paths exist among any two of them.

We prove that MPGSD on G' can be reduced to MBCPG-SC on $\hat{G}$ by showing that any solution S' of MPGSD has a corresponding solution $\hat{S}$ of the MBCPG-SC such that $\sum_{u \in S'} dem(u) = |\hat{S}| - 3$ and and vice versa. Let us first prove the direction MPGSD to MBCPG-SC, by constructing the solution $\hat{S}$ from S'. $\hat{S}$ will consist of nodes r, e, s and all nodes $n_{u,k}$ were $u \in S'$ and $k = 1, .., dem_u$. First, by construction we have $\sum_{u \in S'} dem(u) = |\hat{S}| - 3$. The only additional case that needs to be considered to prove bi-connectivity of $\hat{S}$ is for two nodes $n_{u,i}$ and $n_{v,j}$ such that $u \neq v$. To be more precise we need to prove that there are two disjoint paths connecting them. From the previous observation we have that each of these two nodes is connected to nodes e, s by two disjoint paths. So, nodes $n_{u,i}$ and $n_{v,j}$ are connected with disjoint paths $n_{u,i} - s - n_{v,j}$ and $n_{u,i} - e - n_{v,j}$.

The construction of the solution S' of MPGSD from a solution $\hat{S}$ of the MBCPG-SC is trivial and is based on the construction of graph $\hat{G}$ and the fact, shown in the previously made observation, that any solution $\hat{S}$ consists of r, e, s and all nodes n_u, k where $k = 1, \ldots, dem_u$ for some nodes $u \in S' \subset D$.

The main motivation for defining the MBCPG-SC is its application to systems of interconnected microgrids (Hatziargyriou et al, 2007). It has been shown that optimization of self-adequacy of individual microgrids in such systems can be well modeled using the MPGSD. The proposed problem can also be understood as a partitioning of a power supply network in which all supply nodes (elements of R) have a value M and demand nodes (elements of $V \setminus R$) have value 1. The problem with using the MPGSD for this type of systems is that it does not address the problem of failure resistance. In MBCPG-SC we exploit the fact that by strengthening the constraint of connectivity to bi-connectivity, compared to MPGSD, we are able to have a model that produces more robust subsystems.

Another potential application of the proposed problem is on the hierarchical clustering for wireless networks. In the work of (Banerjee and Khuller, 2001), a network is divided into clusters that can be hierarchically controlled. The basic properties of a cluster are that it has a single control node, the number of nodes it can contain is limited, and the corresponding subgraph needs to be connected. It is natural to extend this formulation with an additional constraint of bi-connectivity of subgraphs to enhance reliability.

3 Growing bi-connected subgraphs

The proposed algorithm is based on the property that a bi-connected graph has an open ear decomposition. Therefore, we start with its definition before presenting the algorithmic outline and the algorithm itself.

3.1 Definition of open ear decomposition

An open ear decomposition of a graph G is defined as a series of paths $\bar{P}_0, \bar{P}_i, \ldots, \bar{P}_n$ called ears. The term path is used for an ordered sequence (v_1, v_2, …, v_m) such that all edges $(v_i, v_{i+1}) \in E$. In the following text we will use the notation P for the set of nodes in $\bar{P}$. The notation $(v_1^j, \ldots, v_{m^j}^j)$ is used for a sequence of nodes in ear $\bar{P}_j$. All the ears $\bar{P}_i$, $i = 1 \ldots n$ in the decomposition satisfy $v_j^i \neq v_k^i$ if $j \neq k$. The exception is the first ear $\bar{P}_0$ which is a cycle $v_1^0 = v_n^0$. For $j > 0$, we have that for each of the two terminating nodes $v_1^j, v_{m^j}^j \in P_j$ exists $k, l < j$ such that $v_1^j \in P_k$, $v_{m^j}^j \in P_l$ where it is not necessarily $l \neq k$. We wish to point out that $v_1^j \neq v_{m^j}^j$, so each P_j is an open ear. Except for such terminating nodes, there is no $v \in P_i \cap P_j$ if $i \neq j$. Finally, for each $v \in G$ there exists at least one i such that $v \in P_i$. If such a decomposition exists for graph G then G is bi-connected. An illustration of an open ear decomposition of a graph is given in Figure 3.

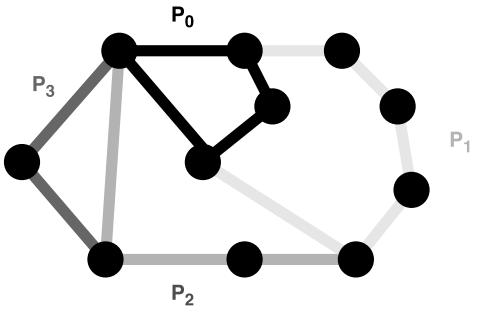

Fig. 3 Examples of an open ear decomposition for a bi-connected graph. Different shades of gray are used for separate open ears.

3.2 Algorithm outline

In the proposed algorithm the idea is to grow a bi-connected subgraph S, starting from an initial cycle and extending it with adequate ears. More formally, we will be iteratively generating a sequence of subgraphs $\bar{S^0} \subset \bar{S^1} \subset \bar{S^2} \subset \cdots$, where $S^{i+1} = S^i \cup P^i$ and P^i is an open ear for S^i. It is evident that if a subgraph is generated in this way it will always have an open ear decomposition. Although it is possible to develop such an algorithm using a depth first search (DFS) the use of BFS is more suitable since it gives us more control of the size $|P_i|$ of the ear that will be added.

In the subsequent text we will be using the following notation.

- $N(u)$ the set of adjacent/neighboring nodes to node u in graph G.
- $BFStreeV/BFStreeE$ are the set of nodes/edges in the BFS tree.

- $ch(u)$ is the set of children of node $u \in BFStree$ in the BFS tree.
- $par(u)$ is the parent node of $u \in BFStree$ in the BFS tree.
- $desc(u)$ is the set of all descendants of node $u \in BFStree$ in the BFS tree.
- $\bar{p}_b[u, v]$ is defined for nodes $u, v \in BFStree$. It represents the path (u, w_1, w_2, $\cdots$, w_n, v) connecting u, v such that all $w_i \in BFStree$. Brackets "("/"[" are used to indicate if u, v are included/excluded in the path. The notation $p_b[u, v]$ is used for the corresponding set of nodes.
- $root(u)$ is node v which is the first ancestor of u in the BFS tree, such that $v \in S$. In case $u \in S$, then $root(u) = u$.
- $root(u, P)$ is node v which is the first ancestor of u in the BFS tree, such that $v \in P$. In case $u \in P$, then $root(u) = u$.
- $d(u)$ is the length of the path (number of nodes) $p_b[u, root(u))$. Note, that it does not include node $root(u)$. In case $u = root(u)$, we have $d(u) = 0$.

It is well known that BFS can be used to find cycles which we exploit in the proposed method. Let us assume that we start the BFS from some initial node r. As we expand the BFS tree, or in other words, new nodes are visited, the first time we encounter a back-edge $(u, v) \in E \setminus BFStreeE$ an initial cycle S is found. More precisely, S is acquired by connecting three segments: path from r to u, the back edge u, v and the path from v to r. In the proposed notation $S = (p_b[r, u], p_b[v, r))$ and it has all its nodes in the BFS tree.

In a similar way we can find new open ears. Let us assume the BFS tree is further expanded and a new back-edge $(s, t) \in E \setminus BFStreeE$ has been found, and that at least one of the nodes s, t is not in S. It is obvious that if the following is satisfied,

$$root(s) \neq root(t) \tag{6}$$

then the sequence

$$\bar{P}(s, t) = (\bar{p}_b[root(s), t], \bar{p}_b[s, root(t)]) \tag{7}$$

will produce a new open ear connected to S. As a consequence $S \cup P(u, v)$ will also be a bi-connected subgraph. We will use the notation $P(u, v)$ for the set of nodes corresponding to $\bar{P}(u, v)$. The same procedure can be used to further expand S with new open ears.

The second constraint that exists in the MBCPG-SC is that $|S| < M$. While growing S, it can easily be maintained if we only allow adding ear $P(u, v)$ if the following equation is satisfied

$$|S| + d(u) + d(v) \leq M \tag{8}$$

If we adapt the BFS search in a way to always explore node u having the lowest value of $d(u)$ the length $|P|$ of the newly found open ear will be among the shorter ones (except in rare cases). This is due to the method of construction in which newly found ears will always have one terminating node equivalent to the currently visited node in the BFS. If $d(M)$ has the largest value of all nodes in the BFS tree which are not in S then the maximal difference between $|P|$ and the shortest ear will be $d(M)$. In case of adapting the BFS in this

way some additional work will be necessary to update the values of $d(u)$ and $root(u)$ as new ears are added to S. It is important to note that after such updates it is possible that several back edges (u, v) that have been previously tested may now satisfy the constraints given in Eqs. (6),(8).

3.3 Algorithm

The algorithm for growing a bi-connected subgraph based on the idea presented in the previous section will start the BFS from some root node r. As the first ear in the decomposition P_0 differs from the rest as it is a cycle, some special initialization needs to be done for r and its neighbors $N(r)$. For all these nodes u we will initially set $root(u) = u$. In the adaptation of BFS for growing a bi-connected graph the distance in the BFS tree will have a different meaning. Instead of following the distance of a node u from the root node r we will track the distance from u to the already generated bi-connected subgraph $\bar{S}$. At the initial step we will consider $S = \{r\}$. Let us assume that we have found two nodes u, v that satisfy the constraints given in Eqs. (6), (8), then the first ear P_0 can be constructed as

$$P_0 = P(u, v) \cup \{r\} \tag{9}$$

As previously stated in case a new ear P is added to S the values of $root(u)$, $d(u)$ will need to be updated for some elements of the BFS tree. It will be necessary to update these values for all $u \in P \cup desc(P)$ in the following way.

$$root(v) = root(v, P) \tag{10}$$

$$d'(v) = \begin{cases} d'(v) - d'(root(v, P)) & v \notin P \\ 0 & v \in P \end{cases} \tag{11}$$

It is important to note that the proposed correction for functions $d(u)$ will produce approximations to the exact distance $d'(u)$. For the function d' we have that $d'(u) \geq d(u)$ since it is possible to have an alternative path to S using some back edges which is shorter. A consequence of this is that the constraint defined in Eq.(8) will never give false positives if function d' is used instead of d. By using this approximate approach it is possible to have a simpler and less computationally expensive implementation.

In the standard BFS there is no change in the distance for visited nodes and no node is re-visited. In the proposed adaptation of BFS such changes can occur and some revisits are necessary. It is possible to use a heap or similar structure instead of a queue to always test the node u with the lowest value $d'(u)$. In practice this is not necessary especially since we are only using an approximation to $d(u)$. On the other hand the need for retesting some nodes further increases the complexity of implementation. Both of these issues are

addressed simultaneously using the following approach. First, nodes will be re-added to the queue as a new ear is added to S and their re-evaluation is needed. Since it is possible for the same node to be added multiple times to the queue due to the addition of multiple ears, an additional value will be used to track if an evaluation is needed. The algorithm for growing a bi-connected subgraph is better understood by observing Algorithm 1.

Algorithm 1 Pseudo code for growing a bi-connected subgraph

```
procedure BFSGrowBiConnected(G, r)
    For all u ∈ G initialize Dist, Eval, Parent
    Initialize all u ∈ r ∪ N(R)
    Add all N(R) to Q
    while Q is not empty do
        current = Q.dequeue()                          ▷ Using Queue (FIFO) structure
        if current.Eval ∧ (M − |S| ≤ current.Dist) then
            for all u ∈ N(current) do
                if (u, current) is BackEdge) then
                    if u, v produce an open ear satisfying Eq. (6), (8) then
                        Set P(u, current) based on Eq. (7) or Eq. (9)
                        S = S ∪ P(u, current)
                        Update(P, Q)
                        Exit procedure if S = M
                    end if
                else
                    u.[Root, Dist] = [current.Root, current.Dist + 1]
                    Update parent, child relations for u, current
                    Q.enqueue(u)
                end if
            end for
            current.Eval = false
        end if
    end while
end procedure
```

The proposed algorithm starts with a standard BFS initialization of the distance, parents and descendants for all the nodes with the additional property of the need for evaluation. Initially all the values of $Eval$ will be set to *true*. An auxiliary structure is used to store all the properties of individual nodes, which can be accessed and updated using the node id. Next, we initialize the root r and all its neighbors $N(r)$ as previously described and all nodes in $N(r)$ are added to the queue Q. The main loop is executed for each node $current$ in Q until $Q = \emptyset$. For each such node we first check if an evaluation is needed and if so all its neighbors $N(current)$ are evaluated. For each $u \in N(current)$ we check if $(current, u)$ is a back-edge. In case it is not we add u to the Q as in the BFS, and we set $root(u) = root(current)$. In case u is a back-edge we check if $P(u, current)$ is an open ear connected to S. If this is true the subgraph S is extended with $P(u, current)$ and necessary updates are performed using procedure $Update(P, Q)$. After all the elements of

$N(current)$ are visited the evaluation of node *current* is complete and we set $current.Eval = false$.

The update procedure for a newly found open ear, $Update(P, Q)$, is used to change the state of Q and nodes based on the $P(u, current)$. The details of the procedure are given in Algorithm 2. In it, for all $u \in P$ the distance is set

Algorithm 2 Update procedure for adding an ear to subgraph S.

```
procedure UPDATE (P,Q)
    for all (u ∈ P) do
        u.[Root, Dist] = [u, 0]
        UpdateBFSBranch(u, u, 0)
        Q.Enqueue(u)
    end for
end procedure
procedure UPDATEBFSBRANCH(u, root, dist)
    for all (v ∈ ch(u)) do
        if v ∉ S ∧ v ∈ BFStree then
            u.[Root, Dist, Eval] = [root, dist + 1, true]
            UpdateBFSBranch(v, root, Dist + 1)
            Q.Enque(u)
        end if
    end for
end procedure
```

to $d.Dist = 0$. Each node u now becomes a root of a new potential ear, so we set $u.root = u$. For each node $u \in P$ we wish to update the branch of the BFS tree whose root is u.

This is done using a recursive procedure $UpdateBFSBranch(u, root, dist)$. In it we go through all the BFS descendants v of u that are not already in S and set the $v.root = root$ and $v.dist = dist+1$. By doing so the node properties are set to the values that correspond to the new state of S. For each such node re-eavaluation is needed to check if new open ears have been created so we set $v.Eval = true$. The addition to the queue Q is done after the recursive call $UpdateBFSBranch(v, root, dist + 1)$ for updating the descendants. This order is important since nodes will be added to the queue in reverse order of their distance and as a consequence shorter potential ears are checked first. By doing so S is more gradually grown.

4 Algorithm for MBCPG-SC

When solving the MBCPG-SC multiple subgraphs $\bar{S_1}, .., \bar{S_n}$ should be grown together. The idea of the algorithm is to randomize this process. The randomization is done on two levels. First, the growth of individual subgraphs $\bar{S_i}$ should be randomized. This can simply be done by adding an additional parameter $p_0 \in [0, 1]$ and a corresponding random variable $p \in [0, 1]$ which will be used to decide if a valid open ear is added to S or not. It is important to note that by not adding an open ear P_k at iteration k does not necessarily

exclude the nodes inside P_k from the corresponding subgraph $\bar{S}$. This is due to the fact that they can be a part of some ear P_l that will be added to S at a later iteration.

It is evident that the growth of subgraphs $\bar{S_1}, \cdots, \bar{S_n}$ is interdependent since a node u can only be an element of a unique S_i. On the other hand, the growth of some S_i will also effect the direction of expansion of the BFS tree of neighboring subgraphs. Because of this the second type of randomization should effect the speed and order in which all the S_i will be grown. The proposed method can be better understood through the pseudocode given in Algorithm 3.

Algorithm 3 Randomized method for generating a solution S for MBCPG-SC.

```
procedure GenerateMBCPG-SC(roots)
    For all i Set root_i^j.Unavalable = true for all S_j where i ≠ j
    Initialize all S_i for roots_i
    while Can Expand Some S_i do
        Select Random Expansion Length MaxL ∈ (2, MaxExpLength)
        Select Random S_i from Expandable
        cLength = 0
        repeat
            P = BFSGrowSingleEarBiConnected(S_i, G)
            Apply UpdateBFSTree(S_j, P) for all j ≠ i
            cLength = cLength + |P|
        until (cLength ≥ MaxL) ∧ CanExpand(S_i))
    end while
end procedure
```

In the proposed algorithm separate BFS trees and corresponding auxiliary structures exist for each of the S_i. The auxiliary structure for tracking node properties is extended with a property *Available* for nodes, which is used to indicate if a node has been added to some of the other subgraphs. Initially this value is set to *true* for all nodes in all subgraphs. The first step, is setting $root_i^j.Available = false$ for all the S_j where $i \neq j$. Here we use the notation u_i^j for node u_i in the auxiliary structure of subgraph $\bar{S_j}$. Next, the $BFStree$ is initialized for each subgraph $\bar{S_i}$ using the corresponding $root_i$ as described in the previous section. The main loop is repeated until no subgraph can be further expanded. At each iteration a random subgraph $\bar{S_i}$ is selected for expansion of a maximal allowed size $MaxL$. $MaxL$ is a random variable from some range $[2, MaxExpLength]$. The goal of the inner loop is to extend S_i with at least $MaxL$ nodes, or until it is not possible to extend it. In this loop an adaptation of the procedure presented in the previous section is used in the form of function $BFSGrowSingleEarBiConnected(S_i, G)$ which only adds a single ear P to S_i. This function does not consider nodes having $Available = false$. After an ear P is added to S_i, the BFS tree and corresponding structures need to be updated for all S_j where $i \neq j$ using procedure $UpdateBFSTree(S, P)$.

The update procedure needs to perform several tasks. First, all the nodes $u \in P$ need to be made unavailable for the growth of S_j, where $j \neq i$, which can be done by setting $u.Available = false$ in the corresponding auxiliary structures. Secondly, all the nodes in $v \in desc(u)$ which have been cutoff from the BFS tree (for some S_j, $i \neq j$), by the removal of u, need to be re-initialized so they can potentially be re-added to the BFS tree. This is done by setting $v.Distance = INF$ and $v.Eval = true$. Although there is a potential that a node $v \in desc(u)$ may be reached by continuing the growth of the BFS tree, there is no guarantee for this. The question is which nodes need to be re-evaluated to make this possible. It is obvious that if a back-edge (w, v) existed for the BFS tree that node w can be reconnected to v. On the other hand if such a back edge existed and $root(w)! = root(v)$ then the corresponding open ear would have been already added to S. The only potentially disregarded back edges are of the type $root(w) = root(v)$. Because of this the nodes in $desc(root(u)) \setminus desc(u)$ should be re-evaluated if $a.Eval = false$, since they can establish a connection with v. The details of the update procedure for deleting a node are given in Algorithm 4.

Algorithm 4 Update procedure that is applied after removing a node from the BFS Tree.

```
procedure UpdateBFSTree(j, P)
    for all u ∈ P do
        u.Available = False
        Update Parent, Descent relation for u
        for all a ∈ desc(root(u)) \ desc(u) do
            if a.Eval = false then
                a.Eval = true
                Q.enqueue(a)
            end if
        end for
        for all v ∈ desc(u) do
            v.[Eval, Dist] = [false, INF]
            Clear parent, child relations
        end for
    end for
end procedure
```

5 Computational Complexity

In this section we give the computational complexity for the proposed algorithm for growing bi-connected subgraphs of size M. We will start with a simple upper bound. The computational complexity of the BFS is $O(|V|+|E|)$. In the proposed algorithm a BFS is used to find an initial cycle which is calculated in $O(|V| + |E|)$. In the worst case the update procedure will need to recalculate all the visited nodes, and the update procedure can at most be

applied M times so the computational complexity of the proposed algorithm is $O(M(|V| + |E|))$

A much better upper bound can be acquired if we observe the distance d and the average "branching factor" b of the graph (the average out-degree) of vertices. The computational complexity of BFS in this form is $O(b^{d+1})$ in both time and memory. Let us assume that we are generating a bi-connected subgraph having M nodes, for simplicity let $M = 2k - 1$. Further we only consider $b > 1$ which is true for any connected graph having more than two vertices. In case that the proposed algorithm initially finds a cycle containing M nodes the computational complexity will be $O(b^{k+1})$ since we have performed a BFS search of depth k. This is due to the fact that each of the two BFS paths the cycle consists of have k nodes (the root is in both paths). Let us assume that the initial cycle was found after a BFS tree of depth $x + 1 = k$ was explored. The number of nodes that are visited and the maximum number of nodes that could be recalculated, which corresponds to the upper bound of computational complexity, is give by the following formula

$$b^{x+1}+2\sum_{i=0}^{x-1} b^i = b^{x+1}+2\frac{b^x-1}{b-1} = \frac{b^{x+2}+2b^x-b^{x+1}-2}{(b-1)} < \frac{2b^{x+2}}{(b-1)} < \frac{2b^{k+2}}{(b-1)^2} \tag{12}$$

In formula (13) the first term represents the number of nodes visited in the BFS search. The sum represents the number of nodes that could be recalculated. It is equal to the number of nodes in the two subtrees containing the two paths that create the cycle. For $x+t = k$, by applying (13) t times, this inequality will have the form $\frac{2b^{k+2}}{(b-1)^{2t}}$. It is obvious that a bi-connected subgraph containing m nodes will be acquired in less than the following number of visits:

$$\sum_{i=1}^{k-1} \frac{2b^{k+2}}{(b-1)^{2i}} < \frac{2b^{k+2}}{(b-1)^2-1} \approx O(b^{\frac{M}{2}+2}) \tag{13}$$

From this we can conclude that the computational complexity of the proposed algorithm for growing a bi-connected subgraph is $O(b^k)$. It is important to note that a tight bound will be lower since we do not need to sum $i = 1, k-1$, but only have the sum of elements equal to k.

In case of the algorithm for solving the MBCPG-SC we need to grow N bi-connected subgraphs containing at most M nodes. The computational complexity of growing N separate bi-connected subgraphs is $O(Nb^{k+2})$. As previously stated in case of solving the MBCPG-SC, the parallel growth of these subgraphs results in the necessity of revisiting some nodes and edges. This is a consequence of the removal of nodes from existing BFS trees. Each such removal can result in revisiting a subtree of the BFS tree which has at most $O(b^{k-1})$ nodes. When a node a is add to some subgraph A it must be removed from BFS trees of the rest of the N subgraphs. In total there are at most NM nodes that are added to subgraphs. From this we can conclude that the computational complexity of solving the MBCPG-SC is $O(N^2Mb^{k-1} + Nb^{k+2})$. It is important to note that this is an upper bound for the worst case computational

complexity and in practical applications it will be significantly lower. Another important aspect, as in the case of the standard BFS, is that the proposed algorithm has a significant worst case space complexity which is similar to the computational one. A consequence of this is that in case of finding very large 2-connected subgraphs in very large graphs this issue should be addressed in a similar way as for the original BFS.

6 Local Search for MBCPG-SC

The algorithm in the previous section gives us a method for generating a single solution for MBCPG-SC. A simple way of finding higher quality solutions is to perform multiple runs with different random seeds and selecting the best one. The problem with this approach is that no experience is gained from previously generated solutions and as a consequence a very high number of them needs to be generated to get good quality ones. An alternative approach is to develop a local search procedure which improves sections of already generated solutions. The basic idea of the proposed local search is to regrow only a subsection $I \subset \mathit{\Pi}$ of the previously best found solution. Ideally, we wish to do this only for subsections for which an improvement is possible. The growth procedure presented in the previous section can easily be adapted to such a setting.

Let us define the set of non-located nodes as

$$NonLoc = V \setminus \bigcup_{i=1..n} S_i \tag{14}$$

It is obvious that I should contain at least one $\bar{S}_i$ such that $|S_i| < M$. The other requirement is that I has a potential to expand to some $u \in NonLoc$. For simplicity let us assume that there is only one non-located node $u \in NonLoc$, and there is only one $\bar{S}_i \in I$ such that $|S_i| < M$. Since $\bar{S}_i$ must be bi-connected there must be a $path(root_i, u)$ having only nodes in I. We can define the set of neighboring nodes of a subgraph in the following way

$$N_s(\bar{S}) = \{u \mid (u \in V) \wedge (u \notin S) \wedge (\exists (v \in S)((u, v) \in E)\} \tag{15}$$

It is evident that for an appropriate $path(root_i, u)$ to exist, u must satisfy $u \in N_s(S_j)$ for some $\bar{S}_j \in I$. Using this logic, we can specify I containing m subgraphs based on this necessary condition. Select one $\bar{S}_i \in \mathit{\Pi}$ having $|S_i| < M$ and $\bar{S}_j \in \mathit{\Pi}$ such that it $N_s(\bar{S}_j) \cap NonLoc \neq \emptyset$. Select $m-2$ random subgraphs from $\mathit{\Pi} \setminus \{\bar{S}_i, \bar{S}_j\}$. We will call this method for selecting the elements of I growth random ($GrowR$).

This method for specifying I satisfies the necessary condition but further constraints can be added to make the local search more efficient. Let us observe a graph $G'(V', E')$ induced by $\mathit{\Pi}$ in the following way. Each node in V' corresponds to a subgraph $\bar{S}_i \in \mathit{\Pi}$. An edge $(\bar{S}_i, \bar{S}_j)$ is in E' if there exist nodes $u \in S_i$, $v \in S_j$ such that $(u, v) \in E$. If we are re-growing only two subgraphs $\bar{S}_i, \bar{S}_j$ there is a potential that nodes may be exchanged between them if $(\bar{S}_i, \bar{S}_j) \in E'$, and more generally S_i, S_j will influence each other's

growth. Transitively, the influence will extend to all subgraphs in a connected subgraph $S' \subset G'$.

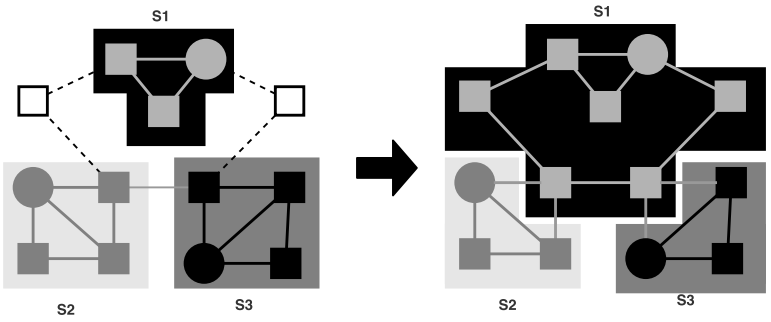

Fig. 4 An illustration of an exchange between subgraphs through non-located nodes. Different shades of gray are used to indicate different subgraphs. Dashed lines represent non-located paths (left side). The edge sets used for defining G' and $\hat{G}$ are $E' = \{(S_2, S_3)\}$ and $\tilde{E} = \{(S_1, S_2), (S_1, S_3)\}$, respectively.

An exchange can occur between subgraphs $\bar{S}_i, \bar{S}_j$ in the same setting even if $(\bar{S}_i, \bar{S}_j) \notin E'$ through nodes in $NonLoc$. Let us assume that for some $u \in S_i$, $v \in S_j$ there is a connecting $path(u, v)$ which only contains nodes in $NonLoc$. In such a case there is a chance that node v can be added to S_i using this path. We give an illustration in Figure 4. Let us define $\tilde{E}$ as a set of edges, where $(\bar{S}_i, \bar{S}_j) \in \tilde{E}$ only if a connecting path exists containing only non-located nodes. Let us now define a new graph $\hat{G}(V', \hat{E})$ using the following edge set

$$\hat{E} = E' \cup \tilde{E} \tag{16}$$

Now, we can say that for the same I an exchange between subgraphs $\bar{S}_i$, $\bar{S}_j$ can only occur if and only if $(\bar{S}_i, \bar{S}_j) \in \hat{E}$ and more generally S_i, S_j will influence each other's growth. Transitively, the same applies to any connected subgraph $I \subset \hat{G}$. A consequence of this is that regrowing should only be considered for a connected (in $\hat{G}$) subsection I, since each isolated section can be treated separately.

Using graph $\hat{G}$ we can define a method for generating a set $|I| \geq m$, where m indicates the number of subgraphs in G, suitable for regrowth in the following way. First select an initial random node $I = \{\bar{A}_0\}$ such that $|A_0| < M$. Note as a reminder that $\bar{A}_0$ represents a subgraph in Π. Iteratively, we add a random node $\bar{A}_i$ such that it is connected (in $\hat{G}$) to at least one node $\bar{A}_j \in I$. When $|I| = m$, we check if at least one node $\bar{A}_j \in I$ satisfies $N_s(\bar{A}_j) \cap NonLoc \neq \emptyset$. If this is true I is accepted as the set that will be regrown. In case this is not true a new set I is generated in the same way. If after N attempts no such I is generated m is increased by one. This process is repeated until I satisfying the necessary constraints is generated. We will call this method for selecting the elements of I growth neighbor ($GrowN$).

The method based on the local search for the MBCPG-SC is given in Algorithm 5. The methods starts with generating an initial solution using procedure

Algorithm 5 Local search based algorithm for finding a high quality solutions for MBCPG-SC.

```
Generate solution Π using GenerateMBCPG − SC
BestSolution = Π
Calculate Ĝ(V′, Ê) for BestSolution
while Not Termination Condition do
    Randomly select m from [2, RegrowSize]
    Randomly select S̄_i such that |S_i| < M
    Generate I based on S_i, m and Ĝ
    Π′ = RegrowPartial(I, BestSolution)
    if |Π′| ≥ |BestSolution| then
        BestSolution = Π
        Calculate Ĝ(V′, Ê) for BestSolution
    end if
end while
```

$GenerateMBCPG - SC$ which is set as the $BestSolution$. For this solution we generate the graph $\hat{G}$ as described in the previous text. In the main loop we first select a random subgraph $\bar{S}_i$ such that $|S_i| < M$ and a random number $m \in [2, RegrowSize]$ which indicates how many subgraphs will be regrown. Next, we generate a random subgraph $I \subset \hat{G}$ based on the node S_i containing m nodes using one of the two methods $GrowR$ or $GrowN$. A new solution Π' is acquired by regrowing I using procedure $RegrowPartial(I, BestSolution)$. This procedure is the same as $GenerateMBCPG - SC$ with the following changes. First, for all nodes $u \notin NonLoc \cup I$ we set $u.Available = false$. Next, only subgraphs in I are grown and considered in the update procedure. The remaining subgraphs are left unchanged. The next step in the main loop consists in testing if $|\Pi'| \geq |BestSolution|$. In case this is true the best found solution is updated and the graph $\hat{G}$ is recalculated. Note that we have allowed the update of the best solution even if it has the same quality as Π to diversify the search. The main loop is repeated until a maximal number of iterations or no further improvement can be achieved.

The proposed method is an initial approach for solving the MBCPG-SC and may be improved in several ways. Firstly, by extending the greedy method to standard metaheuristic approaches like tabu search in which extensive recalculation of the same neighborhoods is avoided, or by using simulated annealing to escape local minima. It is important to note that we have defined a neighborhood for solutions. A more effective local search may be developed by adapting some of the standard methods used in graph partitioning problems by substituting the use of nodes with open ears.

7 Computational Experiments

In this section we present the results of the computational experiments used to evaluate the performance of the proposed method. We have compared the basic randomized growth algorithm with its extension using the local search procedure with and without exploiting the subgraph neighborhoods. All the algorithms have been implemented in C# using Microsoft Visual Studio 2015. The source code and the execution files have been made available at (Jovanovic, 2017). The calculations have been done on a machine with Intel(R) Core(TM) i7-2630 QM CPU 2.00 Ghz, 4GB of DDR3-1333 RAM, running on Microsoft Windows 7 Home Premium 64-bit. The computational experiments have been divided into two sections. The first group applies the proposed method on synthetically generated problem instances and the second one is dedicated to real-world problems.

7.1 Test instance generation

In the evaluation of the proposed algorithm our focus is on unit disc graphs due to their close relation with wireless networks. As stated above the proposed problem can be used to find clustering schemes for hierarchical control of such systems with enhanced fault tolerance. Due to the fact that this is a newly defined problem there are no standard benchmark instances available for comparison. Therefore, it was necessary to also generate problem instances with known optimal solutions. In the generation procedure our goal is to generate graphs that are "as random as possible". To be more precise we attempt to have the node positions close to the uniform probability distribution and to have the number of edges that are adjacent to each node relatively balanced. To achieve this we use the following procedure.

Let us say that we are generating a problem instance having n root nodes and M is the maximal allowed number of nodes in a subgraph, having a solution in which Π contains all the nodes. For problem instances of this form we generate unit disc graphs inside of a box B with dimensions 1, 1. The value for the length for establishing edges is $d = 1/(\sqrt{\alpha n M})$. α is a parameter to vary the density of the graph used in specifying different data sets. The first step is generating n random bi-connected unit disc graphs having M nodes. As our aim is to have the nodes of graph G spread over the entire area of the box B having an area equal to 1. We want to have each of the subgraphs G_i covering an area of close to $1/n$. To achieve this for each such graph we first define a box with dimensions $1/(\sqrt{n}R)$, $\sqrt{n}R$ where R is randomly selected from the interval $(0.5, 1)$ used to vary the shape of G_i. For this box a random bi-connected graph G_i is generated. This is done by repeatedly generating $M \cdot \delta$ random points and checking if the resulting unit disc graph contains a bi-connected subgraph G_i containing M nodes. To increase the probability of having nodes of the graph covering an area close to $1/n$, four nodes have been forced to random positions at each edge of the box. To achieve this a high

number of graphs needs to be generated, especially in the case of a high value of M. This is the reason for the additional points, specified using parameter $\delta = 1.1$, in the box, since it greatly reduces the number of graphs that are generated. Out of all the nodes in G_i one would be randomly selected as the $root_i$.

The next step in generating a problem instance with a known solution is combining graphs G_i. Let us remember that by the method of construction for each subgraph G_i contains M nodes having some values of coordinates (x, y). As a consequence we can easily translate G_i by a vector (t_x, t_y) by changing all node positions to $(x + t_x, y + t_y)$. The basic idea of combining the graphs G_i is to randomly position (translate) them in a way that all the nodes fit in a box having dimensions 1, 1. As previously stated our goal is to have a relatively balanced number of edges for all the nodes. Secondly, we want to distribute the nodes over the entire box B. This is achieved through the following iterative procedure. Initially, the graph G is set to G_0. At each iteration a new graph G_i is set at a random position and added to G, but some constraints must be satisfied. First, all the points in $G \cup G_i$ must fit into a box having dimensions 1,1. The number of new edges ne connecting G and G_i, has to satisfy $ne \geq MinCon$. To achieve this, multiple random positions have to be tested for each subgraph G_i. The parameter $MinCon$ is used to bound the allowed number of new edges which is related to the number of edges $|E_i|$ in the graph as $MinCom = max(3, \gamma|E_i|)$. Empirical tests have shown that values $\gamma = 0.2$ produce satisfactory results. For each graph G_i a 1000 positions are tested and the one satisfying the constraints and having the lowest number of $min_{j=1..i}|E_j|$ would be chosen for adding to G. The last step is randomly enumerating all the nodes in G.

7.2 Experiments on Synthetic Graphs

To have an extensive evaluation of the proposed algorithm, we have generated a wide range of problem instances for different numbers of root nodes n (5-100) and maximal allowed number of nodes in a subgraph M (5-100). For each pair n, M 40 problem instances have been generated using the algorithm presented in the previous subsection with different values for the seed of the random number generator. We have generated two problem sets varying in the level of connectivity by setting $\alpha = 1.5, 2$. For each of the problem instances a single run of the growth based algorithms is performed. This is done for the two local search methods based on $GrowR$ and $GrowN$ for generating the set of subgraphs I which will be regrown.

The same set of parameters for specifying the algorithm is used for all the problem sizes. The value of the parameter for accepting an ear is $p_0 = 0.6$ in the growth algorithm. The parameter specifying the maximal expansion of a subgraph at one step is $MaxExpLength = 12$. The value $RegrowSize = 9$ was used for the the upper bound on the value m which is used for generating the size $|I|$ of the set of subgraphs that will be regrown. These values have

been chosen through extensive testing on the method $GrowN$ for a wide range of parameters. In the general case the method is not very sensitive to parameter values which has been confirmed using the Wilcoxon significance test. An important conclusion that has been made is that low values of $p_0 < 0.5$ have a statistically significant worse performance in case of $M = 50, 100$. The parameter $RegrowSize$ was tested for the range $4 - 13$. The chosen value of $RegrowSize = 9$ gave the best overall average results but this was not statistically significant in the majority of cases when compared to lower or higher values. It is important to note that the computational time when comparing the minimal and maximal values of $RegrowSize$ was around five times. The method was least sensitive to changing the value of $MaxExpLength$, which was tested for values between $7 - 16$, for which it always had a similar performance. In all of the computational experiments the algorithm would execute until 10,000 iterations (generated solutions) have been reached or more than 2000 iterations have been completed without an improvement to the best found solution.

The results of the computational experiments focus on the quality of found solutions and the computational cost which can be seen in Tables 1, 2. The evaluation of solution quality is done using the average normalized error of the found solutions compared to the known optimal ones, for each of the used methods. More precisely, for each of the 40 test instances, for each pair (n, M), the normalized error is calculated by $(Optimal - found)/Optimal \cdot 100$, and we show the average values. To have a better comprehension of the performance we have also included the standard deviation, maximal errors and the number of found optimal solutions (hits). Further, we have observed the statistical significance using the Wilcoxon significance test with $p = 0.05$. The computational cost is analyzed through the average execution time needed to find the best solution. To have a better understanding of these execution times, we have also included the average number of iterations (number of generated solutions) to find the best solution and the corresponding standard deviation in Tables 1, 2.

As it can be seen in Tables 1, 2 the local search based on the use of subgraphs $GrowN$ manages to significantly outperform $GrowR$. As expected, $GrowN$ produces a higher level of improvement for problem instances having a higher number of roots, or in other words when the graph is divided into a greater number of subgraphs. In case of problems having 100 subgraphs the average error is nearly halved. For problem instances with lower values of n this improvement is smaller but consistent. In the 50 problem sets, in only 7 $GrowR$ produces slightly better results. This only occurred for problems having five subgraphs, in which the use of neighborhoods is not essential. The improved performance of $GrowN$ was consistently significant for problem instances having more than 10 root nodes. Overall both methods had a good performance, with an average error going up to 9%/4.2% and 5%/3.5% for problem instances generated using $\alpha = 2/1.5$ for $GrowR$ and $GrowN$, respectively. In practice the problem instances having the maximal number of allowed nodes $M = 5$ proved to be the hardest. The methods managed to find better

Table 1 Comparison of the performance of the $GrowR$ and $GrowN$ methods for unit disc graphs with $\alpha = 2$. Statistically significant improvements are underlined.

Roots X M	$Avg(Stdev)$		Max		$Hits$		$AvgIter$		$AvgTime(ms)$	
	R	N	R	N	R	N	R	N	R	N
5X5	0.40(1.50)	0.50(2.04)	8.00	12.00	37	37	92(27)	83(41)	3	4
5X10	0.15(0.94)	0.20(0.98)	6.00	6.00	39	38	270(185)	218(113)	28	25
5X25	0.62(1.38)	0.64(1.36)	7.20	7.20	27	28	974(298)	669(1150)	278	212
5X50	0.89(1.35)	0.89(1.35)	5.20	5.20	16	15	1469(424)	1055(1332)	894	681
5X100	0.65(0.63)	0.60(0.75)	2.80	3.20	4	9	1555(1124)	1191(160)	2201	1693
10X5	0.75(1.71)	0.55(1.48)	6.00	6.00	32	34	551(358)	325(216)	42	27
10X10	1.00(1.36)	0.45(1.02)	7.00	5.00	16	31	1042(659)	766(613)	175	161
10X25	1.50(1.40)	1.08(1.32)	4.80	4.80	8	15	1823(870)	1408(494)	795	676
10X50	1.18(1.04)	0.95(0.92)	4.20	3.80	7	7	2780(1598)	2356(1076)	2777	2504
10X100	1.37(1.07)	1.07(0.90)	4.00	3.30	0	4	2670(677)	3174(935)	6322	7639
25X5	4.10(3.50)	1.74(2.42)	10.40	8.80	11	22	1799(1093)	1373(629)	228	174
25X10	1.61(1.00)	0.77(0.78)	4.40	4.00	1	9	2552(974)	1734(696)	587	610
25X25	1.86(0.97)	1.61(0.76)	4.48	3.36	0	0	4022(1004)	3832(1435)	2363	2198
25X50	1.74(0.72)	1.76(0.70)	3.60	3.76	0	0	6367(963)	4469(1592)	8851	5981
25X100	1.90(0.85)	1.63(0.78)	4.08	4.32	0	0	6742(4185)	5790(523)	23472	19437
50X5	5.74(2.27)	3.00(1.96)	10.00	7.20	2	5	3608(368)	2709(736)	881	572
50X10	2.29(1.10)	1.29(0.74)	5.80	3.00	0	1	5236(2227)	3750(1364)	2142	1392
50X25	2.75(0.88)	1.82(0.75)	5.20	3.92	0	0	7605(1586)	5972(1509)	7640	4901
50X50	2.51(0.82)	1.78(0.63)	4.12	2.92	0	0	8936(820)	7512(3122)	21091	14435
50X100	2.39(0.61)	1.90(0.51)	3.76	3.42	0	0	9053(1557)	8885(502)	53818	45608
100X5	8.39(1.72)	4.11(1.39)	11.80	7.40	0	0	6154(699)	5833(3363)	3639	2547
100X10	3.34(0.95)	1.42(0.49)	6.80	2.90	0	0	8352(2267)	6367(2437)	8122	4163
100X25	3.95(0.74)	2.03(0.51)	5.56	3.00	0	0	9326(402)	9231(606)	23091	14310
100X50	3.89(0.64)	2.23(0.37)	5.58	3.10	0	0	9689(219)	9456(325)	57973	35392
100X100	3.68(0.64)	2.29(0.45)	5.36	3.39	0	0	9807(178)	9709(430)	141864	94561

solutions for graphs with higher edge densities ($\alpha = 1.5$). We believe that the main reason for this is that such problem instances are in practice easier to solve since there is a greater number of potential bi-connected subgraphs. The hardest instances to solve are the ones in which the maximal allowed size of a graph is the smallest and the number of subgraphs is the highest. For a notable number of test instances $GrowN$ manages to find optimal solutions, in case of $\alpha = 2/1.5$ it is close to 20%/30%. It is important to note that optimal solutions are only found for graphs having up to 500 nodes. $GrowN$ has a robust behavior in the sense that the maximal error for a problem set is 12%/5.6% but in the majority of the cases it is less than 5%/3% for $\alpha = 2/1.5$.

When we observe the computational speed of the proposed methods the additional cost for calculating the neighborhoods in $GrowN$ is overall neglectable when compared to $GrowR$. In many cases the use of neighborhoods even decreases the calculation time due to a lower number of generated solutions. Overall both methods prove to be computationally very efficient having taken 140/94 and 160/91 seconds to generate close to 10 000 solutions in case of graphs having 10 000 nodes and $\alpha = 2/1.5$ for $GrowR$ and $GrowN$, respec-

Table 2 Comparison of the performance of the *GrowR* and *GrowN* methods for unit disc graphs with $\alpha = 1.5$. Statistically significant improvements are underlined.

Roots X M	*Avg*(*Stdev*)		*Max*		*Hits*		*AvgIter*		*AvgTime*(*ms*)	
	R	*N*	*R*	*N*	*R*	*N*	*R*	*N*	*R*	*N*
5X5	0.20(0.87)	0.40(1.50)	4.00	8.00	38	37	178(167)	84(74)	8	4
5X10	0.00(0.00)	0.00(0.00)	0.00	0.00	40	40	240(78)	195(1023)	26	24
5X25	0.22(0.54)	0.14(0.40)	2.40	1.60	33	35	647(489)	434(1341)	188	122
5X50	0.06(0.19)	0.07(0.27)	0.80	1.60	36	36	850(636)	508(479)	546	329
5X100	0.14(0.23)	0.18(0.37)	1.00	1.80	25	27	1594(1079)	1025(768)	2289	1378
10X5	0.85(1.78)	1.30(2.12)	6.00	6.00	31	28	542(132)	314(162)	49	27
10X10	0.33(0.57)	0.18(0.44)	2.00	2.00	29	34	761(970)	394(447)	125	79
10X25	0.60(1.00)	0.46(0.78)	4.40	3.20	21	26	1944(758)	1122(1553)	891	524
10X50	0.41(0.53)	0.34(0.50)	2.80	2.00	12	20	2169(1246)	1444(815)	2286	1563
10X100	0.53(0.77)	0.43(0.63)	3.30	2.60	7	14	3259(530)	2211(253)	8131	5213
25X5	3.16(1.70)	2.46(1.64)	6.40	5.60	1	7	1503(933)	1341(556)	205	183
25X10	0.85(0.83)	0.59(0.62)	3.60	2.00	9	16	2299(832)	1675(583)	567	567
25X25	1.01(0.73)	0.90(0.81)	3.04	3.36	2	5	4179(4020)	2894(2215)	2610	1840
25X50	1.03(1.00)	0.57(0.46)	5.44	1.76	1	3	5237(1068)	4387(2178)	7629	5936
25X100	1.05(0.81)	0.88(0.73)	3.44	3.04	1	0	6161(226)	5144(3478)	22790	16628
50X5	4.17(1.31)	2.81(1.10)	6.80	6.00	0	0	3028(2680)	2456(1293)	805	563
50X10	1.27(0.75)	0.63(0.50)	3.60	2.40	2	3	4321(831)	3305(3095)	1858	1983
50X25	1.29(0.53)	0.81(0.40)	2.56	1.84	0	0	6806(1835)	5117(1452)	7371	4586
50X50	1.32(0.41)	0.93(0.43)	2.12	1.80	0	0	8405(1313)	7235(3364)	20658	13846
50X100	1.47(0.56)	1.01(0.42)	2.92	2.30	0	0	8599(1271)	7959(4155)	55918	38379
100X5	4.78(0.91)	3.15(0.80)	7.80	5.00	0	0	5754(1915)	4531(559)	3836	2200
100X10	2.14(0.66)	0.82(0.35)	3.50	1.80	0	0	7737(1524)	5948(2552)	7862	4238
100X25	2.18(0.50)	0.99(0.37)	3.24	1.88	0	0	9537(329)	8688(288)	26109	14141
100X50	2.19(0.52)	1.18(0.30)	3.80	1.82	0	0	9685(114)	9163(675)	60910	34018
100X100	2.24(0.53)	1.20(0.36)	3.54	2.49	0	0	9785(336)	9775(215)	159973	91749

tively. It is important to note that for the largest instances in case of both methods the execution of the algorithm was terminated before the stagnation occurred (no improvements of the best found solution). The scaling of the methods is good in the sense that the increase of average execution time from instances having 100 nodes to 10,000 was around 10,000 times. The number of iterations performed for different problem instances for the same pair n, M would highly vary which can be seen from the standard variation. We wish to point out that *GrowN* manages to find high quality solutions with a low number of generated solutions when compared to the solution space.

The reason for the better performance of *GrowN* compared to *GrowR* comes from the definition of neighborhoods and the way they are explored. Both methods in essence randomly explore the neighborhood, and as a consequence the larger it is the higher number of solutions needs to be generated to find an optimal one. On the other hand if a neighborhood is small it is much easier to find an optimal solution in the sense of a lower number of tests. As previously discussed *GrowN* only regrows subgraphs which are interconnected, and as a consequence dependent. In case of *GrowR* this is not necessarily the

case but independent subsections are explored jointly. In practice this means that *GrowR* often explores unnecessary large neighborhoods which consist of a combination of independent ones.

7.3 Application of the method to the problem of fault tolerant systems of interconnected microgrids

In this section we present the results of applying the proposed method for issues arising in systems of interconnected microgrids. In the proposed model the goal is to divide the electrical distribution grid, presented as a graph, into microgrids which correspond to connected subgraphs. As previously mentioned bi-connectivity is often used for providing fault tolerance in systems, in our case of individual microgrids. Microgrids are very suitable for applying demand response management for lowering the peak load (Nunna and Doolla, 2013) which is the problem of interest in the proposed application of the model. The main reason for this is the fact that demand side management corresponds to a variety of hard optimization problems which cannot, due to computational complexity, be solved on large systems but can in the case of smaller ones like microgrids. It has been previously shown that the benefits, in the sense of lowering peak load, of increasing of the subsystems become neglectable after a certain size, which gives us justification for applying the size constraint in the proposed model (Jovanovic et al, 2016a).

In our numerical experiments we have applied the proposed method on graphs generated on power flow data from Poland, France, EU and standard IEEE benchmark problem instances. To be more precise we have used the data provided by MatPower (Zimmerman et al, 2011). Since previous research has shown that having systems with more than 200 households does not provide additional benefits for peak load shaving using demand side management (Jovanovic et al, 2016a) we have set the maximum subgraph size $M = 200$ for problem instances having more than 500 nodes, and $M = 50$ in case of small problem instances. The number of subgraphs was $N = \lceil TotalNumberOfNodes/M \rceil$. Root nodes have been selected at random with the constraint that they must be a part of at least one cycle. In total 18 different instances are tested having between 100-10000 nodes.

It is important to note that in these test instances not all nodes are parts of cycles and as that they can not be a part of a bi-connected subgraph. In this group of test we also analyze the robustness of the methods in the sense of dependence of the quality of found solutions and the seed of the random number generator. This has been done by executing 10 independent runs of the algorithm using different seeds for the random number generator for each problem instance. In Table 3 we show the best, worst, average and standard deviation of found solutions. We also show the effect of the local search compared to the basic method for solving the MBCPG-SC. Further we have tested statistical significance using the student t-test with unequal

variance with $p = 0.05$ for comparing $GrowR$ and $GrowN$. The parameter setting was the same as in the case of synthetically generated graphs.

Table 3 Comparison of the performance of the $GrowR$ and $GrowN$ on test instances generated based on real-world data. Statistically significant improvements are underlined.

Instance	$Avg(Stdev)$		Max		$AvgTime(ms)$	
	R	N	R	N	R	N
case118	97.50(0.50)	97.30(0.46)	98	98	61	14
case145	101.00(0.00)	101.00(0.00)	101	101	13	13
case_illinois200	102.00(0.00)	102.00(0.00)	102	102	5	6
case1888rte	718.20(10.24)	724.30(10.01)	735	741	7342	15895
case1951rte	763.70(7.93)	763.80(7.65)	775	773	8018	9752
case2383wp	1602.70(9.02)	1597.40(13.11)	1617	1616	9494	13414
case2736sp	2123.20(21.51)	2138.90(19.71)	2146	2165	12246	12779
case2737sop	2139.70(22.97)	2158.10(10.90)	2171	2177	13509	15209
case2746wop	2052.60(19.46)	2058.00(18.10)	2075	2085	11736	17320
case2746wp	1804.40(1.80)	1804.90(0.30)	1805	1805	2450	2847
case2848rte	1114.50(9.04)	1118.60(6.89)	1126	1127	8522	20206
case2868rte	1124.10(7.01)	1131.30(5.51)	1133	1139	8711	23651
case3012wp	2058.50(21.00)	2065.60(13.96)	2087	2088	12026	11809
case3375wp	2093.40(25.57)	2111.10(16.50)	2140	2140	8099	17335
case6468rte	2899.10(58.16)	2988.80(42.97)	3039	3023	41488	58700
case6470rte	2959.70(40.47)	3032.50(40.20)	3021	3084	42763	63537
case6495rte	2995.30(37.57)	3009.80(24.98)	3042	3045	41141	51348
case9241pegase	5479.30(76.11)	5653.90(57.29)	5595	5734	64445	76112

In case of the real-world based test instances the behavior of the two algorithms is similar as in the case of synthetically generated data. $GrowN$ has found better average solutions in all but two cases. Statistical significance has been confirmed in around 30% of the cases. Although the optimal solutions for this group of problem instances was not known the impression is that the improvement is lower than in case of problems defined on unit disc graphs. We believe the main reason for this is that the existence of nodes that cannot be a part of bi-connected subgraphs tends to mislead the local search. It is important to note that such behavior can be avoided with the use of preprocessing.

8 Conclusion

In this paper we have introduced a new problem of finding the maximum number of nodes contained in a set of disjoint bi-connected subgraphs of a graph with the additional constraint on the maximal size of a subgraph. This type of problem can be potentially applied for many practical problems. One example is the partitioning of electrical grids into a system of interconnected microgrids with a high level of resistance to failure. For solving the MBCPG-SC, a novel computationally efficient method for growing bi-connected subgraphs

has been introduced. This method has been adapted to the setting of growing multiple graphs in parallel to generate solutions for MBCPG-SC. The quality of solutions generated in this way was further improved using a local search method exploiting neighboring relations between subgraphs. The proposed method managed to acquire approximate solutions having an average error of up to 5% when compared to known optimal solutions. Further, the method is highly computationally efficient in the sense that it manages to find such solutions within two minutes for graphs having 10,000 nodes.

In the future we plan to extend the current research in several directions. First, we aim to explore the weighted version of the problem which would be more suitable for problems occurring in electrical distribution systems. Once the weighted version is solved we may also relate the problem towards multi-depot vehicle routing problems and the assignment of costumers to routes of different depots. Secondly, we shall extend the method to metaheuristic approaches like ant colony optimization, GRASP or variable neighborhood search which appear as good options. Finally, we shall explore the potential of applying the proposed growth procedure for problems like the bi-connected dominating set and bi-connected vertex cover problem.

References

Arefifar SA, Mohamed YARI, EL-Fouly TH (2012) Supply-adequacy-based optimal construction of microgrids in smart distribution systems. IEEE Trans Smart Grid 3(3):1491–1502

Arefifar SA, Mohamed Y, EL-Fouly TH (2013) Comprehensive operational planning framework for self-healing control actions in smart distribution grids. IEEE Trans Power Syst 28(4):4192–4200

Banerjee S, Khuller S (2001) A clustering scheme for hierarchical control in multi-hop wireless networks. In: INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, IEEE, vol 2, pp 1028–1037

Bazgan C, Tuza Z, Vanderpooten D (2010) Satisfactory graph partition, variants, and generalizations. Eur J Oper Res 206(2):271 – 280

Borra D, Pasqualetti F, Bullo F (2015) Continuous graph partitioning for camera network surveillance. Automatica 52:227 – 231

Buchanan A, Sung JS, Butenko S, Pasiliao EL (2015) An integer programming approach for fault-tolerant connected dominating sets. INFORMS J Comput 27(1):178–188

Chang YC, Lin ZS, Chen JL (2006) Cluster based self-organization management protocols for wireless sensor networks. IEEE Trans Consum Electron 52(1):75–80

Dorigo M, Blum C (2005) Ant colony optimization theory: A survey. Theor Comput Sci 344(2):243–278

Feo T, Resende M (1995) Greedy randomized adaptive search procedures. J Global Optim 6(2):109–133

do Forte VL, Lucena A, Maculan N (2013) Formulations for the minimum 2-connected dominating set problem. Electronic Notes in Discrete Mathematics 41:415 – 422

Goldschmidt O, Takvorian A, Yu G (1996) On finding a biconnected spanning planar subgraph with applications to the facilities layout problem. Eur J Oper Res 94(1):97 – 105

Hansen P, Mladenović N, Pérez JAM (2010) Variable neighbourhood search: methods and applications. Ann Oper Res 175(1):367–407

Hatziargyriou N, Asano H, Iravani R, Marnay C (2007) Microgrids. IEEE Power Energy Mag 5(4):78–94

Hochbaum DS (1993) Why should biconnected components be identified first. Discrete Appl Math 42(2):203 – 210

Hopcroft J, Tarjan R (1973a) Algorithm 447: efficient algorithms for graph manipulation. Commun ACM 16(6):372–378

Hopcroft JE, Tarjan RE (1973b) Dividing a graph into triconnected components. SIAM J Compu 2(3):135–158

Hu B, Leitner M, Raidl GR (2010) The generalized minimum edge-biconnected network problem: Efficient neighborhood structures for variable neighborhood search. Networks 55(3):256–275

Ito T, Demaine ED, Zhou X, Nishizeki T (2008) Approximability of partitioning graphs with supply and demand. J Discrete Algorithms 6(4):627 – 650

Ito T, Hara T, Zhou X, Nishizeki T (2012) Minimum cost partitions of trees with supply and demand. Algorithmica 64(3):400–415

Jovanovic R (2017) Benchmark data sets for the problem of the maximal bi-connected partitioning of a graph with a size constraint. URL `http://mail.ipb.ac.rs/~rakaj/home/mbcpg_sc.htm`

Jovanovic R, Tuba M (2013) Ant colony optimization algorithm with pheromone correction strategy for the minimum connected dominating set problem. Comp Sci Inform Syst pp 133–149

Jovanovic R, Voß S (2016) A mixed integer program for partitioning graphs with supply and demand emphasizing sparse graphs. Opt Lett 10(8):1693–1703

Jovanovic R, Bousselham A, Voss S (2015a) A heuristic method for solving the problem of partitioning graphs with supply and demand. Ann Oper Res 235(1):371–393

Jovanovic R, Bousselham A, Voss S (2015b) Partitioning of supply/demand graphs with capacity limitations: an ant colony approach. J Comb Optim , doi:10.1007/s10878-015-9945-z

Jovanovic R, Bousselham A, Bayram IS (2016a) Residential demand response scheduling with consideration of consumer preferences. Appl Sci 6(1):16

Jovanovic R, Tuba M, Voss S (2016b) An ant colony optimization algorithm for partitioning graphs with supply and demand. Appl Soft Comput 41:317 – 330

Kuhn F, Wattenhofer R, Zollinger A (2003) Ad-hoc networks beyond unit disk graphs. In: Proceedings of the 2003 joint workshop on Foundations of mobile

computing, ACM, pp 69–78
Li X, Zhang Z (2010) Two algorithms for minimum 2-connected r-hop dominating set. Inform Process Lett 110(22):986–991
Matic D, Bozic M (2012) Maximally balanced connected partition problem in graphs: Application in education. Teach Math 15(2):121–132
Moraes RE, Ribeiro CC (2013) Power optimization in ad hoc wireless network topology control with biconnectivity requirements. Comput Oper Res 40(12):3188 – 3196
Morgan M, Grout V (2008) Finding optimal solutions to backbone minimisation problems using mixed integer programming. In: Seventh International Network Conference (INC 2008), Plymouth, UK, July 8-10, 2008. Proceedings, pp 53–63
Morishita S, Nishizeki T (2013) Parametric power supply networks. In: Du DZ, Zhang G (eds) Computing and Combinatorics, Lecture Notes in Computer Science, vol 7936, Springer, Berlin, pp 245–256
Nunna HK, Doolla S (2013) Energy management in microgrids using demand response and distributed storagea multiagent approach. IEEE Trans Power Del 28(2):939–947
Pearce DJ (2016) A space-efficient algorithm for finding strongly connected components. Inform Process Lett 116(1):47 – 52
Popa A (2013) Modelling the power supply network - hardness and approximation. In: Chan TH, Lau L, Trevisan L (eds) Theory and Applications of Models of Computation, Lecture Notes in Computer Science, vol 7876, Springer, Berlin, pp 62–71
Robbins HE (1939) A theorem on graphs, with an application to a problem of traffic control. Am Math Mon 46(5):281–283
Schmidt JM (2013) A simple test on 2-vertex-and 2-edge-connectivity. Inform Process Lett 113(7):241–244
Shafique KH (2004) Partitioning a graph in alliances and its application to data clustering. PhD thesis, University of Central Florida Orlando, Florida
Tarjan R (1972) Depth-first search and linear graph algorithms. SIAM J Compu 1(2):146–160
Wang F, Thai MT, Du DZ (2009) On the construction of 2-connected virtual backbone in wireless networks. IEEE Trans Wireless Commun 8(3):1230–1237
Zhang Z, Gao X, Wu W (2009) Algorithms for connected set cover problem and fault-tolerant connected set cover problem. Theor Comput Sci 410(810):812 – 817
Zimmerman RD, Murillo-Sánchez CE, Thomas RJ (2011) Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. IEEE Trans Power Syst 26(1):12–19