Algorithms and Complexity Results for Finding Graphs with Extremal Randić Index

Rex K. Kincaid
Department of Mathematics, College of William and Mary, Williamsburg, Virginia 23187-8795

Sarah J. Kunkler
Microsoft Corporation, 1 Microsoft Way, Redmond, Washington 98052

Michael Drew Lamar
Department of Biology, College of William and Mary, Williamsburg, Virginia 23187-8795

David J. Phillips
Department of Mathematics, United States Naval Academy, Annapolis, Maryland 21401

We show that finding a subgraph realization with the minimum generalized Randić index for a given base graph and degree sequence is solvable in polynomial time by formulating the problem as the minimum weight perfect *b*-matching problem of Edmonds (J Res Natl Bur Stand 69B (1965), 125–130). However, the realization found via this reduction is not guaranteed to be connected. Approximating the minimum weight perfect *b*-matching problem subject to a connectivity constraint is shown to be NP-hard. For instances in which the optimal solution to the minimum Randić index problem is not connected, we describe a heuristic to connect the graph using pairwise edge exchanges that preserves the degree sequence. Although we focus on finding graph realizations with minimum Randić index, our results extend to finding graph realizations with maximum Randić index as well. Applications of the Randić index are provided to synchronization of neuronal networks controlling respiration in mammals and to normalizing cortical thickness networks in diagnosing individuals with dementia. **NETWORKS, Vol. 000(00), 000–000 2016**

Keywords: generalized Randić index; network realization; degree sequence; minimum weight perfect *b*-matching; connectivity constraint; synchronization; cortical networks

1. INTRODUCTION

Let an undirected graph $G = (N, E)$ be given and for $i \in N$, denote the degree of node i by $d(i)$. For a given $\alpha \in \mathbb{R}$, the generalized Randić index of G is

$$R_\alpha(G) = \sum_{(i,j)\in E} (d(i) \cdot d(j))^\alpha.$$

Let $b = (b_i)_{i\in N}$ be a sequence of positive integers, and let S_b denote the subset of subgraphs of G so that the degrees are specified by b. By this we mean that for $H \in S_b$ and $i \in N$, the degree of node i is exactly b_i. The *generalized minimum Randić index* problem is to find a subgraph $H \in S_b$ so that

$$R_\alpha(H) = \min\{R_\alpha(F) \mid F \in S_b\}.$$

In what follows, we focus on the generalized minimum Randić index problem, but our results extend to the case of the generalized maximum Randić index problem as well. Because of this, we refer to the *generalized Randić index problem* when describing results that apply to either the minimum or maximum Randić index problem. In this article, we examine algorithms and complexity results for variations on the generalized Randić index problem. In particular, we have the following results:

- We describe the first strongly polynomial algorithm to solve the generalized minimum Randić index problem. Although we typically consider the case where $\alpha = 1$, our results and algorithms apply to any non-zero α.
- We define the variant where S_b is restricted to connected subgraphs as the *connected generalized Randić index problem*, an important property for many applications [21].We prove that the connected generalized Randić index problem is APX-hard, that is, it is even NP-hard to approximate.
- We define the variant where G is the complete graph on n vertices as the *generalized Randić index degree sequence*

Received June 2014; accepted March 2016
Correspondence to: D. J. Phillips; e-mail: dphillip@usna.edu
DOI 10.1002/net.21680
Published online in Wiley Online Library (wileyonlinelibrary.com).

problem. We conjecture that the connected generalized Randić index degree sequence problem is NP-hard even if α is restricted to one. However, we also conjecture that an approximation algorithm exists.
- We describe an $O(n^3)$ heuristic for the connected generalized Randić index degree sequence problem, with n the number of nodes.
- Using our heuristic, we provide computational results to investigate our conjectures concerning the connected generalized Randić index degree sequence problem.
- We provide two applications of our algorithm and heuristic as well as computational results.

The Randić index of a graph was originally defined in chemistry by Randić [27] to estimate the degree of *branching* in molecule graphs of hydrocarbons using the specific cases of $\alpha = -1/2$ and $\alpha = -1$. Subsequent to [27], Bollobás et al. [5] generalized the Randić index to general $\alpha \in \mathbb{R} - \{0\}$. The Randić index has become a widely used graph invariant to analyze the connectivity and assortivity properties of a graph, for example, see Li et al. [22] and Beichl and Cloteaux [3]. A survey of results for the Randić index can be found in [23]. Thus, the generalized Randić index problem and its variants are significant and important.

To the best of our knowledge, no polynomial time algorithm to solve the generalized Randić index problem has been described. Additionally, the complexity of the variants of the generalized Randić index problem are not known. Previous attempts to solve the generalized Randić index problem have focused on heuristics without a polynomial time guarantee. Kincaid et al. [20] propose a method based on a tabu search. Li et al. [22] describe a heuristic for the generalized Randić index degree sequence problem and the specific case of $\alpha = 1$ but do not consider the complexity questions, although they do consider the connected case. Aouchiche et al. [1] describe a heuristic that uses a local search to find graphs with specific degree properties that minimize or maximize a graph invariant and present computational results on, among other invariants, the generalized Randić index problem with specific values for α. The heuristic in [1] represents a generalization of the heuristic described in [15]. However, in both [1] and [15], the set of feasible graphs are not restricted in degree.

In what follows, we describe our algorithms and heuristics in section 2. We show that the connected generalized Randić index problem is APX-hard in section 3. We provide computational results investigating the complexity of the connected generalized Randić index degree sequence problem in section 4. Finally, we provide two applications of our algorithms and heuristics in section 5.

1.1. Notation and Definitions

We assume the reader to have a knowledge of graph theory (see, e.g., [33]). We let $\mathbb{R}$ and $\mathbb{Z}$ denote the sets of real numbers and integers, respectively. We consider an undirected graph $G = (N, E)$, which consists of nodes $N = \{1, \ldots, n\}$ and edges E. We assume that the graph is simple, that is, there are no self-loops and no multiedges. The degree of node i is defined as $d_i(G) = |\{j \mid (i,j) \in E\}|$. The degree sequence is the list of the degrees of all the nodes in G, which we represent as $d(G) = (d_1(G), d_2(G), \ldots, d_n(G))$. When the particular graph is clear from context, we omit G in the previous definitions. Let $b \in \mathbb{Z}^n$ be a given positive integer vector. We define a perfect b-matching of G as follows.

Definition 1.1 (Perfect b-Matching for Undirected Graphs)**.** *For a simple undirected graph $G = (N, E)$ and positive integer vector $b \in \mathbb{Z}^n$, a perfect b-matching of G is a subgraph $H = (N, M)$ where $M \subseteq E$ and $d_i(H) = b_i$ for all $i \in N$.*

We will also refer to the subset of edges $M \subseteq E$, as opposed to the subgraph H, as a perfect b-matching. A sequence of non-negative integers is considered *graphic* if it is the degree sequence of a graph. Degree sequences can correspond to more than one adjacency matrix or graph. We call these graphs different realizations of the degree sequence. We use the Havel–Hakimi algorithm [13, 16] to determine if a degree sequence is graphic. For nodes $u \in N$ and $v \in N$, we say u and v are *connected* if there exists a path from u to v, and that a graph is connected if every pair of distinct nodes are connected. We use the following theorem of Chen [6] to determine whether a graphical degree sequence has a connected graph realization.

Theorem 1.2 (Chen [6], Corollary 6.20)**.** *Let $n \geq 2$ be a given positive integer and let $(a_1, a_2, \ldots, a_n)$ be a given sequence of positive integers known to be graphical. Then, the sequence corresponds to the degrees of the nodes of a connected graph if and only if*

1. $\sum_{i=1}^{n} a_i$ *is even, and*
2. $\sum_{i=1}^{n} a_i \geq 2(n-1)$.

2. ALGORITHMS

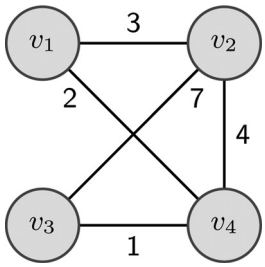
Our primary goal is to devise an algorithm to solve the generalized Randić index problem. In this section, we first describe an algorithm to solve this problem and then provide a heuristic for the connected generalized Randić index problem. Finally, we provide a technique to extend our problem to the directed case.

2.1. The Perfect b-Matching Problem

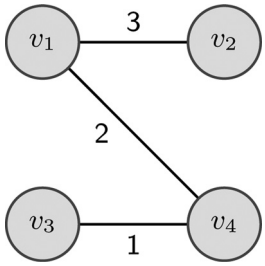
We first describe the perfect b-matching problem of Edmonds [7] which could also be called the f-factor problem of Tutte [31]. Consider a graph $G = (N, E)$, a positive integer vector $b = (b_1, \ldots, b_n) \in \mathbb{Z}^n$ and $M \subseteq E$, a perfect b-matching. For a given b-matching M, the subgraph induced by M is written (N, M). We denote the set of perfect b-matchings of a graph G by $\mathcal{P}\_b(G)$. For edge weights $w : E \to \mathbb{R}$, the *minimum weight perfect b-matching problem* is that of finding a perfect b-matching with minimum weight, that is, to calculate

$$M^*(G) = \arg\min \left\{ \sum_{e \in M} w(e) \mid M \in \mathcal{P}_b(G) \right\}. \quad (1)$$

For example, let G be the undirected graph with edge weights as given below



and let $b = (2, 1, 1, 2)$ for nodes v_1, v_2, v_3, and v_4 respectively. We want to select b_i edges incident with each node v_i that will produce the overall minimum weight. Therefore, a minimum weight perfect b-matching induces the graph G' below.



Note that for this example the solution G' is the only perfect b-matching for G.

Theorem 2.1. *The generalized Randić index problem can be solved in strongly polynomial time.*

Proof. Let an instance of the generalized Randić index problem be given with a positive integer vector $b \in \mathbb{Z}^n$ and graph $G = (N, E)$. Then all graphs induced by perfect b-matchings are feasible subgraph realizations for the minimum Randić index problem. Also, the edge sets of the feasible subgraph realizations of the minimum Randić index problem are perfect b-matchings. Thus, the set of feasible perfect b-matchings on G is identical to the set of feasible subgraph realizations on G for the minimum Randić index problem.

If our instance is a minimum Randić index problem, we create a minimum weight perfect b-matching problem with the following edge weights. For $(i, j) \in E$, set

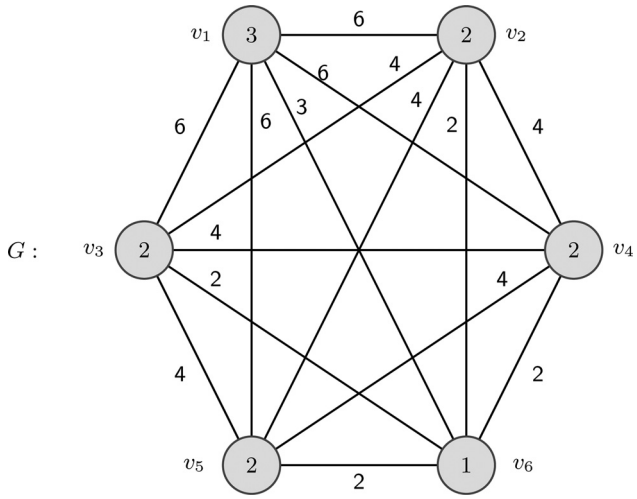
$$w_{ij} = (b_i \cdot b_j)^{\alpha}, \quad (2)$$

and if we are maximizing, set $w_{ij} = -(b_i \cdot b_j)^{\alpha}$. Therefore, we can create an instance of a minimum weight perfect b-matching problem to solve the generalized Randić index problem. As the perfect b-matching problem can be solved in strongly polynomial time [29], the generalized Randić index problem can also be solved in strongly polynomial time. ∎

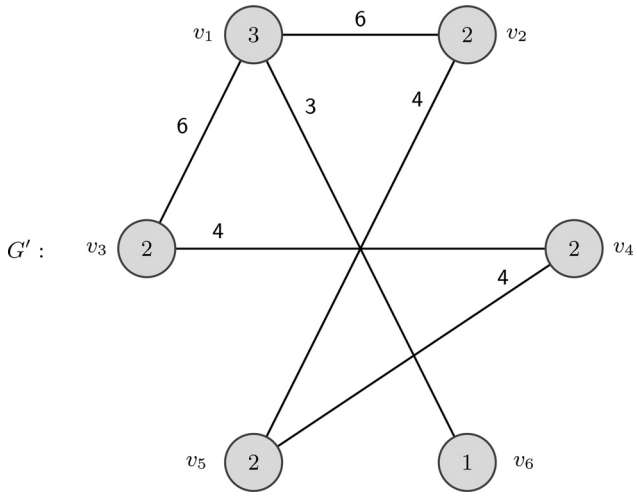
Note that Theorem 2.1 does not apply to the connected generalized Randić index problem. We note that the generalized Randić index degree sequence problem can be solved using the complete graph for G.

2.2. *Example: Generalized Randić Index Degree Sequence Problem*

Suppose we are given the degree sequence $d = (3, 2, 2, 2, 2, 1)$, and wish to find a graph realization with the minimum Randić index and degree sequence d. We let nodes $v_1, v_2, v_3, v_4, v_5, v_6 \in N$ with $b = (3, 2, 2, 2, 2, 1)$. Now we can form the complete graph G, with weights corresponding to $b_i \cdot b_j$ for every node $v_i, v_j \in N$.



Now we solve the minimum weight perfect b-matching for G and obtain G':



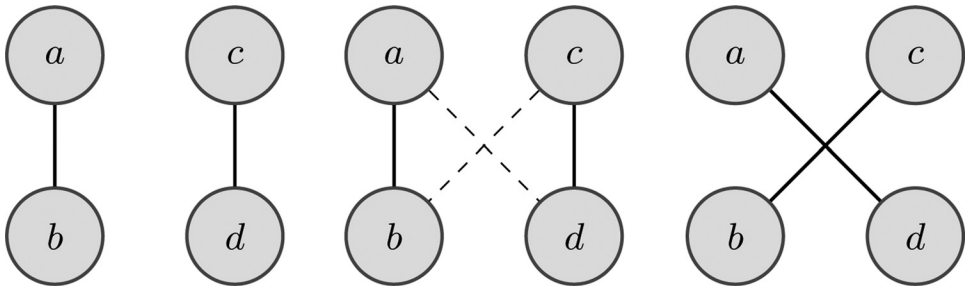
G' is a solution for the minimum weight perfect b-matching. The sum of the weights $R(G') = 6 + 6 + 4 + 4 + 4 + 3 = 27$ is the minimum Randić index. Note that there are other solutions to the matching that will produce the minimum Randić index and a different realization, that is, the solution is not unique.

2.3. *Heuristic for Connecting Disconnected Realizations*

In this section, we present a local search heuristic using *two-switches* to connect disconnected realizations. As we

show below in Theorem 3.1, even approximating the connected generalized Randić index problem is NP-hard. We therefore focus on the connected generalized Randić index degree sequence problem. However, our heuristic can be used for the connected generalized Randić index problem, although we cannot prove that the heuristic terminates successfully for that problem. We note that our heuristic is guaranteed by Theorem 2.3 below to terminate for the connected generalized Randić index degree sequence problem.

We describe a two-switch, with an example as follows:



When doing a two-switch, we examine two edges, $(a,b),(c,d) \in E$. If $(a,d) \notin E$ and $(b,c) \notin E$ then we can remove edges (a, b) and (c, d) and create edges (a, d) and (b, c). This is not a unique move, as we could also use (a, c) and (b, d) if $(a,c) \notin E$ and $(b,d) \notin E$. Two-switching is an easy way to obtain a different graph with the same degree sequence after a graph is created. In the case of the connected generalized Randić index degree sequence problem, the graph could be initially created using any method, for example, using the Havel–Hakimi algorithm [13, 16] or by finding a perfect b-matching on a complete graph. We can construct a *metagraph* of a degree sequence, where the metagraph is an undirected graph with each node representing a graph realization of a degree sequence and each edge representing a two-switch. The following theorem shows that the metagraph is always a connected graph, as shown by Hakimi [14].

Theorem 2.2 ([14])**.** *Given graphs G and G' such that $d(G) = d(G')$, there exists a sequence of two-switches going from G to G'.*

Note that we can use the Havel–Hakimi algorithm to discard any nongraphic degree sequence. The heuristic sequentially performs a two-switch between pairs of connected components of G until all the components are connected:

Two–switch Heuristic

Inputs: G, a disconnected graph with degree sequence d
Outputs: G', a connected graph with degree sequence d

Use Theorem 1.2 to ensure d has a connected graph realization.
Let $G' := G$.
while the number of connected components in G' is ≥ 2
    do a two-switch with two components to connect them using an edge from a cycle in one component and a random edge from another component
return G'

Note that the method to connect the disconnected realizations may not produce graphs with an optimized Randić index. Also note that we do not need to check whether the randomly chosen edges are adjacent or not as they are in separate connected components.

Theorem 2.3. *The two-switch heuristic terminates in $O(n^3)$ time.*

Proof. Suppose we have k connected components. At least one connected component must have a cycle. If not, then each component is a tree, and every possible two-switch keeps the number of connected components the same at k. But this contradicts the existence of a connected realization, as Theorem 2.2 states that any realization can be reached through a series of two-switches. Now choose an edge on a cycle in one of the connected components and perform a two-switch with any random edge in one of the other connected components. This is guaranteed to connect the two components, giving us $k-1$ connected components. By recursion, we have the result. As cycles can be detected using a depth-first search algorithm in $O(n+m)$ time, where m is the number of edges, and there are $k-1 < n$ iterations of the heuristic, a connected realization can be achieved in $O(n^3)$ time. ∎

2.3.1. Extension to Directed Graphs In this section, we extend our algorithm to directed graphs with no loops, that is, no edges from a node to itself. Multiedges are, however, allowed. In particular, we first define a directed Randić index, and then show how a directed perfect b-matching problem can be formulated as an undirected perfect b-matching problem on an associated bipartite graph. Our extension is similar to [24] in which the maximum matching problem is extended to directed graphs for a network controllability problem.

Consider the directed graph $\overrightarrow{G} = (N, E)$ with node set $N = \{v_1, \ldots, v_n\}$ and edge set $E = \{(v_i, v_j) \mid v_i \rightarrow v_j\}$. As stated, we assume $\overrightarrow{G}$ has no loops, that is, for all $(v,u) \in E$, $v) \overline{=} u$. The degree sequence for $\overrightarrow{G}$ is a non-negative integer-pair sequence $d = \{(d_i^+, d_i^-) \mid i = 1, \ldots, n\}$, where

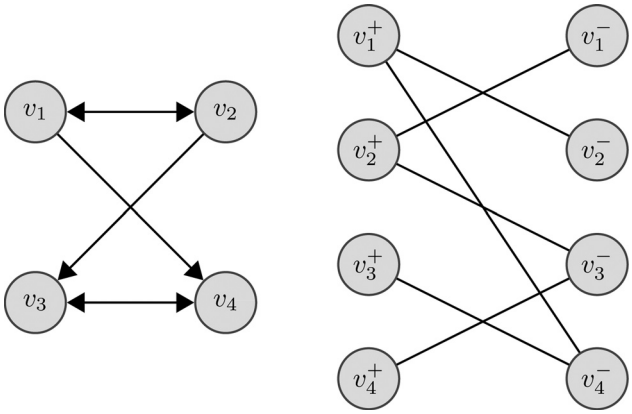


FIG. 1. Left: Directed graph $\overrightarrow{G} = (N, E)$ with node set $N = \{v_1, v_2, v_3, v_4\}$. Right: Bipartite representation $\overrightarrow{G}^* = (N^*, E^*)$, where $N^* = N^+ \cup N^-$, $N^+ = \{v_1^+, \ldots, v_n^+\}$, $N^- = \{v_1^-, \ldots, v_n^-\}$, and $e^* = (v_i^+, v_j^-) \in E^*$ if and only if $e = (v_i, v_j) \in E$.

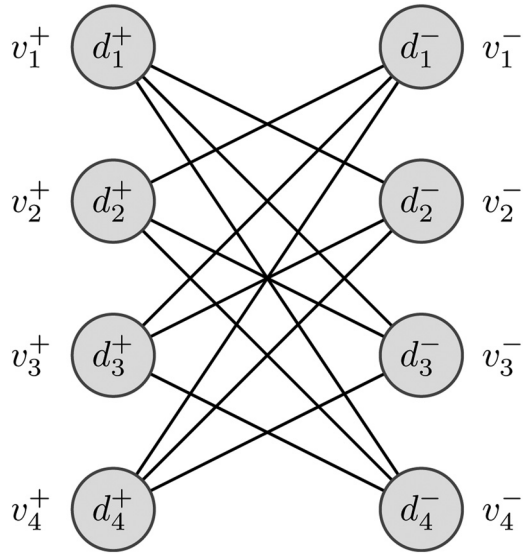


FIG. 2. Bipartite setup of the generalized Randić index degree sequence problem for an example directed graph $\overrightarrow{G} = (N, E)$ with node set $N = (v_1, \ldots, v_4)$, out-degree sequence $(d_1^+, \ldots, d_4^+)$ and in-degree sequence $(d_1^-, \ldots, d_4^-)$. The network shown is the bipartite form $\overrightarrow{K}_4^*$ of the complete (loop-free) directed graph $\overrightarrow{K}_4$, with corresponding node set $N^* = (v_1^+, \ldots, v_4^+, v_1^-, \ldots, v_4^-)$ and positive vector $b = (d_1^+, \ldots, d_4^+, d_1^-, \ldots, d_4^-)$. For the minimum weight perfect b-matching algorithm, we would set $w_{ij} = d_i^p d_j^q$, where $p, q \in \{+, -\}$.

we denote the out-degree and in-degree sequences by d^+ and d^-, respectively. There are four different Randić index-type measures [35] given by

$$R^{pq}(G) = \sum_{(v_i, v_j) \in E} d_i^p d_j^q, \tag{3}$$

where $p, q \in \{-, +\}$.

We define a perfect b-matching in a directed graph as follows.

Definition 2.4. (Perfect b-Matching for Directed Graphs)**.** *For a loop-free, directed graph $\overrightarrow{G} = (N, E)$ and positive integer-pair sequence $b = (b^+, b^-) = \{(b_i^+, b_i^-) \in \mathbb{Z}^n \times Z^n \mid i = 1, \ldots, n\}$, a perfect b-matching is a subgraph (N, M) with $M \subseteq E$ such that for node $v_i \in N$, the out and in-degree of v_i in the subgraph (N, M) are b_i^+ and b_i^-, respectively.*

To use the existing algorithm for undirected graphs, we consider the equivalent bipartite form of $\overrightarrow{G}$ given by $\overrightarrow{G}^* = (N^*, E^*)$, where $N^* = N^+ \cup N^-$, $N^+ = \{v_1^+, \ldots, v_n^+\}$, $N^- = \{v_1^-, \ldots, v_n^-\}$, and $e^* = (v_i^+, v_j^-) \in E^*$ if and only if $e = (v_i, v_j) \in E$ (see Fig. 1 for an example). We can define a perfect b-matching on $\overrightarrow{G}^*$ as a subset of edges $M^* \subseteq E^*$ such that for node $v_i \in N^+$ or N^-, the degree of node v_i in (N^*, M^*) is b_i^+ or b_i^-, respectively. This modified definition is a special case of the undirected version with node set $N = (v_1^+, \ldots, v_n^+, v_1^-, \ldots, v_n^-)$ and the positive vector $b = (b_1^+, \ldots, b_n^+, b_1^-, \ldots, b_n^-)$. Thus, to find a minimum weight perfect b-matching for a directed graph $\overrightarrow{G}$, we simply find a minimum weight perfect b-matching for its bipartite form $\overrightarrow{G}^*$.

The generalized Randić index degree sequence problem can also be formulated for directed graphs $\overrightarrow{G}$ as follows (see Fig. 2):

- Consider the bipartite form $\overrightarrow{K}_n^*$ of the complete directed graph $\overrightarrow{K}_n$.
- Let $b^+ = d^+$ and $b^- = d^-$.
- For $p, q \in \{+, -\}$, let edge weights $w_{ij} = d_i^p d_j^q$ [see Equation (3)].

3. COMPLEXITY

In this section, we show that even approximating the connected generalized Randić index problem is NP-hard. We first define approximation algorithms (see [32] for further details about approximation algorithms). Let $S \subset \mathbb{R}^n$ and $f : S \to \mathbb{R}$ be a given feasibility set and objective function, respectively. Define an η-approximation algorithm for the minimization problem $v^* = \min_{x \in S} f(x)$ as a polynomial time algorithm that finds a solution $y \in S$ with $f(y) \leq \eta v^*$. We say that we can *approximate* a minimization problem if there exists an η such that an η-approximation algorithm exists. Note that $\eta \geq 1$ is implicit with $\eta = 1$ only if an exact algorithm exists.

Theorem 3.1. *Approximating the connected generalized Randić index problem is NP-hard.*

Proof. Recall that a Hamiltonian cycle on G is a simple cycle that includes all nodes of G. We claim the existence of a Hamiltonian cycle on a given graph is equivalent to the feasibility of the connected minimum Randić index problem on a related instance. Recall that an instance of Hamiltonian

cycle consists of a graph, so let such an instance be given with $G = (N, E)$. Now define the vector $b \in \mathbb{R}^{|N|}$ by setting $b_i = 2$ for $i \in \{1, \ldots, |N|\}$ and consider the resulting connected generalized Randić index instance using the graph G and vector b.

We first show that if the connected generalized Randić index instance (G, b) is feasible, then there is a Hamiltonian cycle on G. Suppose there is a feasible solution $H = (N, F)$, which means $F \subseteq E$, each node $u \in N$ has degree 2, and H is connected. As each node has even degree and H is connected, there is an Eulerian cycle T on H. We claim that T is a Hamiltonian cycle on G, which means each node is visited exactly once by T except the start node which is visited exactly twice. Choose $u \in N$ and note that two edges of T are incident with u. Then, because T traverses every edge, the node u is visited. Denote the start node of T by $s \in N$ and consider traversing T beginning at s. If the traversal visits a node $u \in N \setminus \{s\}$ more than once then an edge was traversed into u, a second distinct edge was traversed out of u, and a third distinct edge was traversed into u, a contradiction as there are exactly two distinct edges adjacent to u in H. The same argument applies if s is visited more than once before the traversal is complete. So each node is visited exactly once by T except the start node, which is returned to when the traversal is complete, that is, T is a Hamiltonian cycle. Thus, if the instance (G, b) is feasible, then G possess a Hamiltonian cycle.

We now show that if there is a Hamiltonian cycle on G, then the connected generalized Randić index problem on (G, b) is feasible. Consider a Hamiltonian cycle C on G and the subgraph induced by C. Such a subgraph is connected as each node is visited. Also, each node has degree 2 as each node $u \in N$ has one arc used to enter u and exactly one distinct arc used to exit u. Thus, if G possesses a Hamiltonian cycle, then (G, b) must be feasible to the given connected minimum Randić index instance.

Now suppose there were an η-approximation algorithm to the connected generalized Randić index problem for some $\eta \geq 1$. If the algorithm returns a solution to the instance (G, b) then G possesses a Hamiltonian cycle. If it does not, then the instance (G, b) was not feasible and G does not possess a Hamiltonian cycle. Note that the argument does not rely on what value η is. ∎

Note that we have not shown what the complexity is when the input graph G is the complete graph, that is, we do not know the complexity of the connected generalized Randić index degree sequence problem.

4. COMPUTATIONAL RESULTS

In this section, we seek to gain insight into the complexity of the connected generalized Randić index degree sequence problem. To investigate this problem, we use a simplified version of our heuristic to try and find connected realizations of degree sequences with small Randić index. We generate these sequences based on three graph families and investigate how successful the heuristic is on each family.

To solve the connected generalized Randić index degree sequence problem we use an implementation of an algorithm for solving the minimum weight perfect b-matching problem that was written by Andrews, Huang, Jebara, and Schogolev (http://www.cs.columbia.edu/∼jebara/code/bmatch/) and uses the GOBLIN graph library (http://goblin2.sourceforge.net/). Our experiments were all conducted in MATLAB using an Apple iMac machine with 2.2 GHz i7 quad core processors and 16 GB of RAM. We were unable to find other implementations of b-matching algorithms on MATLAB although more efficient algorithms of b-matching algorithms have been studied [25]. Based on the limitations of our code and environment, we found that running instances with more than 100 nodes was not tractable. Therefore, we focused on instances of that size or smaller.

We used three families of graphs in our experiments: Erdős-Rényi, geometric and scale-free which we describe below. We limited our computational experiments to degree sequences for which connected realizations were known to exist. Theorem 1.2 was used to discard graphs with a degree sequence that had no connected realizations. The MATLAB functions used to generate the geometric and scale-free graphs are from CONTEST: A Controllable Test Matrix Toolbox for MATLAB [30].

We now describe the three families of graphs:

4.1. *Erdős-Rényi Graphs*

A number of nodes n and a probability of edge connection p are chosen. A uniform random number on the interval [0, 1] is generated for each possible edge. If the number generated for an edge is less than p then the edge is added to the graph.

4.2. *Geometric Graphs*

A number of nodes n and a radius r are chosen. Each node v_i is placed uniformly at random in the unit square, giving coordinates (x_i, y_i). We connect nodes v_i and v_j if $(x_i - x_j)^2 + (y_i - y_j)^2 \leq r^2$ [30].

4.3. *Scale-Free Graphs*

A preferential attachment algorithm is used to create graphs whose degree sequences follow a power-law distribution. Following the convention in the literature, we will refer to these graphs as "scale-free." A number of nodes n is chosen. New nodes are added and connected to existing nodes, based on a probability proportional to the current degree of the nodes, until you reach n nodes, making it more likely that a new node will be connected to a higher degree node [30]. The algorithm allows a minimum node degree to be specified.

For each family of graphs, we generated graphs with 25, 50, and 100 nodes, and generated enough graphs until we found 100 instances of each graph type and size that corresponded to a connected degree sequence. For the Erdős-Rényi graphs, we used an average degree per node of 4.25. The corresponding p values used were calculated using $p = \frac{4.25}{n}$

TABLE 1. Maximum percent increase in Randić index

Number of nodes	Erdős-Rényi (%)	Geometric (%)	Scale-free (%)
25	0.35	0.57	3.2
50	0.54	0.77	4.5
100	1.1	0.37	4.6

TABLE 2. Number of degree sequences with minimum Randić index graph realizations that were not connected

Number of nodes	Erdős-Rényi	Geometric	Scale-free
25	1	3	3
50	9	9	9
100	8	6	14

where n is the number of nodes in the graph. Thus $p = 0.17$ for $n = 25$, $p = 0.085$ for $n = 50$, and $p = 0.043$ for $n = 100$. For the geometric graphs we used an average degree per node of 6. The radii were calculated using $r = \sqrt{\frac{6}{\pi n}}$. Our corresponding radii were $r = 0.276$ for $n = 25$, $r = 0.195$ for $n = 50$, and $r = 0.138$ for $n = 100$. We used scale-free graphs with a minimum node degree of 2.

For each graph, we solved the associated minimum Randić index degree sequence problem, that is, we used the degree sequence of the graph as input. Then, if the resulting graph was not connected, we used a simplified version of our heuristic to connect the graph by choosing one edge at random in each component to connect them. We report the maximum percent that the heuristic would increase the Randić index in Table 1 and the number of degree sequences where the graph realization with the minimum Randić index was not connected in Table 2.

As shown in Table 1, when the heuristic was used to reconnect the graphs, the maximum percent increase in the Randić index was less than 5%. Also, the graph realization associated with the minimum Randić index was often connected. These results provide evidence that the connected generalized Randić index degree sequence problem is easier than the connected generalized Randić index problem. We, therefore, conjecture that the connected generalized Randić index degree sequence problem is NP-hard, but that there does exist a constant-factor approximation algorithm.

5. APPLICATIONS

We will now show results for two applications where the effects of network connectivity measures have shown or are hypothesized to play an important role: neuronal synchronization and dementia.

5.1. *Neuronal Synchronization*

We first show how finding graphs with optimized assortativity can be useful in predicting neuronal network dynamics. We do this by considering the synchronous firing of neurons in the preBötzinger complex. This collection of neurons is responsible for the control of respiration in mammals [8].

Using NeuronetExperimenter [17], we simulated 150 rhythmogenic neurons in the preBötzinger complex using the Rubin–Hayes neuron model [28]. It is unknown what degree distribution and network connectivity arise in neuronal networks. However, it seems reasonable to expect that neurons closer to each other are more likely to be connected. Hence, we have chosen to model these networks with 3D geometric (directed) graphs.

A plot of the simulation results are displayed in Figure 3, which shows a spike raster plot of 150 neurons firing over time for three different realizations of a fixed degree sequence: the first, middle, and right plots correspond to network realizations with minimum, original, and maximum Randić indices, respectively [$R^{--}(G)$, see Equation (3)]. The length of a breath is given by the time between when all neurons fire simultaneously, given approximately by the solid vertical lines in the plots. For this specific example, we see a tendency for faster breathing rhythms at the minimum Randić index, as well as more synchronous behavior with network-wide quiescence and activity, as evidenced by the larger gaps of no activity and less noise in the network-wide firing phase in the minimum Randić index case. A more extensive study would be required to extend this proof-of-concept work.

5.2. *Normalizing the Randić Index in Cortical Thickness Networks*

Recently, researchers have begun using graph measures of connectivity to investigate the difference between structural magnetic resonance images taken from healthy individuals and individuals diagnosed with dementia (for example, [2, 18, 19, 34]). Networks in these studies are formed by calculating correlations between the cortical thicknesses of different brain regions, a technique based on the correlation between cortical thickness loss and dementia [10, 12]. This method of network creation has been speculated to provide more insight on the functional relationships between brain regions [11]. Details of how the networks are formed can be found in [18, 26] and we present one such method below. After networks for each subject population (e.g., normal, subjects with dementia) are formed, network measures (e.g., Randić index) are calculated and evaluated for significance.

A particular challenge to using comparisons between the network measures is complicated by the varying number of edges in each of the networks. In particular, some kind of normalization is required for some of the measures used. To normalize, we propose the following scheme:

1. For each network N, determine the underlying degree distribution b.
2. Calculate U_b, the maximum Randić index for b, respectively.
3. Use U_b to normalize the Randić index for N.

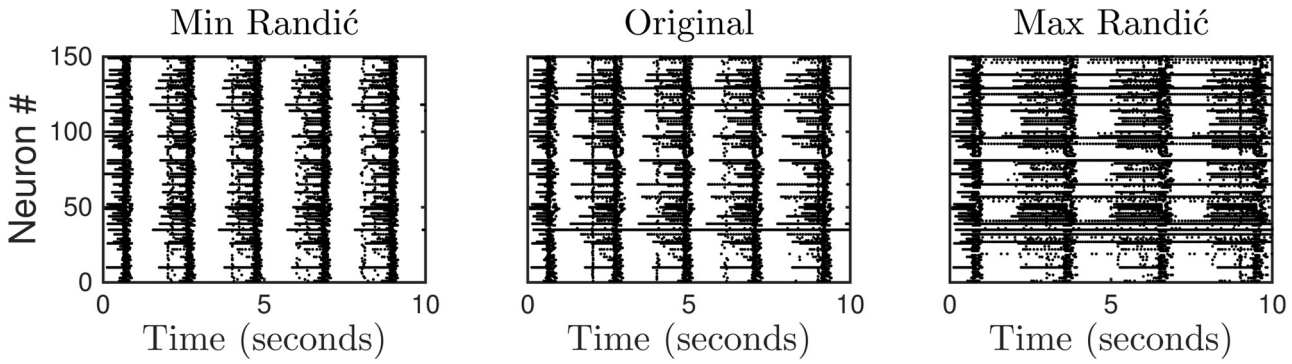


FIG. 3. Neuron spike raster plots for realizations of a specified geometric directed network with minimum Randić index ($R^{--}(G)$) on left, Randić index from the original network in the center, and maximum Randić index on the right. The x-axis is time and the y-axis represents neuron number. The dots represent when each neuron is firing an action potential. The directed Randić index $R(G)$ is as defined in Equation (3).

We use data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and further analyzed by FreeSurfer, a technique developed by Fischl and Dale [9] to measure cortical thicknesses. When applied to the ADNI data, Fischl and Dale discretized the cortical layer into 68 different regions. Phillips et al. [26] used the following steps to generate networks:

1. Subset the population into categories based on whether they were diagnosed as normal (Normal), diagnosed with mild cognitive impairment for three or more years without disease progression (MCI), diagnosed with mild cognitive impairment and then progressed to Alzheimer's Disease within three years (MCI-AD), or diagnosed with Alzheimer's Disease (AD).
2. Within each population, use regression to control for subject age, gender, education level, and interaction effects between age and gender.
3. Use either partial or Pearson's correlations to calculate coefficients and p-values between each of the 68 regions.
4. Use False Discover Rate [4] calculations to determine significant correlation coefficients with an error rate of 5%.
5. Use one of the following schemes to determine edge weights:
   (a) Use no weights: simply include edges or not.
   (b) Use the absolute value of the correlation coefficients as the edge weight.
   (c) Use the product of the normalized cortical thicknesses (i.e., the cortical thicknesses divided by the maximum) between the two regions connected by the significant edge.
   (d) Use both 5b and 5c.

Given the four diagnostic categories, the two possibilities in Step 3, and the four possibilities in Step 5, there are a total of 8 different networks. After calculating the associated degree distributions, we can calculate the maximium Randić index. As an example, we display the results for one network in Tables 3 and 4.

By normalizing the Randić index, comparisons between the groups can more accurately determine whether the assortativity was due to the actual network topology versus other features, such as the total number of edges. For example, as shown in Table 3, the normalized Randić index differentiates normal subjects from subjects with dementia symptoms. Also, the normalized Randić index shows that the disease progression does increase the Randić index monotonically with increasing symptoms whereas the original Randić index is not monotonic. Many studies simply delete edges in networks, to compare non-normalized graph measures [18, 19, 34]. Such methods effectively discard significant data about the relationships between different brain regions. Normalizing by dividing out by the optimized metric allows for comparisons without ignoring network features. We note that normalization can also demonstrate when some graph creation methods are not useful in finding discernable differences in the Randić index. Based on our results, we would conclude that partial correlations are necessary for finding significant Randić index differences. This can be seen by comparing the percentages in Table 3 versus those in Table 4. However, to actually use the optimization in such a statistical study, significance testing (e.g., permutation testing to determine whether the comparisons are valid) would be required, for example, see [26].

TABLE 3. Original, maximized, and normalized Randić index. CT weighting used with partial Pearson's correlations

Group	Original Randić index	Maximum Randić index	Percentage
AD	1293.97	1651.69	78.34%
MCI-AD	1491.38	1908.27	78.15%
MCI	615.10	809.41	75.99%
Normal	129.30	193.47	66.83%

TABLE 4. Original, maximum, and normalized Randić index. CT weighting used with ordinary Pearson's correlations

Group	Original Randić index	Maximum Randić index	Percentage
AD	172910.5	182773.6	94.6%
MCI-AD	182098.8	192005.4	94.84%
MCI	146719	155715.5	94.22%
Normal	231957.8	244387.6	94.91%

6. CONCLUSIONS AND FUTURE WORK

To the best of our knowledge, we have described the first known polynomial algorithm for the generalized Randić index and the generalized Randić index degree sequence problems. A key feature of our algorithm is that it works for all non-zero values of the Randić index parameter α. In addition, we showed that the connected generalized Randić index problem is NP-hard to approximate. We provided computational evidence to support our conjecture that the connected generalized Randić index degree sequence problem is at least approximable. Finally, we showed two applications of our algorithms to two different computational biology problems.

In addition to resolving our conjecture about the connected generalized Randić index degree sequence, future work includes implementing faster versions of our algorithm and heuristics. In addition, we are interested in finding further applications of our algorithms and heuristics. We are also interested in developing an effective heuristic for the connected generalized Randić index problem.

ACKNOWLEDGMENTS

The authors would like to thank the referees and editor for their useful comments, suggestions, and corrections.

REFERENCES

[1] M. Aouchiche, P. Hansen, and G. Caporossi, Variable neighborhood search for extremal graphs, 20: Automated comparison of graph invariants, MATCH Commun Math Comput Chem 58 (2007), 365–384.

[2] F. Bai, N. Shu, Y. Yuan, Y. Shi, H. Yu, D. Wu, J. Wang, M. Xia, Y. He, and Z. Zhang, Topologically convergent and divergent structural connectivity patterns between patients with remitted geriatric depression and amnestic mild cognitive impairment, J Neuroscience 32 (2012), 4307–4318.

[3] I. Beichl and B. Cloteaux, Measuring the effectiveness of the s-metric to produce better network models, Proc 2008 Winter Simulation Conference, IEEE, Miami, FL, pp. 1020–1028.

[4] Y. Benjamini and Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing, J R Stat Soc Ser B 57 (1995), 289–300.

[5] B. Bollobás, P. Erdös, and A. Sarkar, Extremal graphs for weights, Discr Math 200 (1999), 5–19.

[6] W. Chen, Graph theory and its engineering applications, World Scientific Pub. Co. Inc., Singapore, 1997.

[7] J. Edmonds, Maximum matching and a polyhedron with 0, 1 vertices, J Res Natl Bur Stand 69B (1965), 125–130.

[8] J.L. Feldman and C.A. del Negro, Looking for inspiration: New perspectives on respiratory rhythm, Nat Rev Neurosci 7 (2006), 232–242.

[9] B. Fischl and A.M. Dale, Measuring the thickness of the human cerebral cortex from magnetic resonance images, Proc Natl Acad Sci USA 97 (2000), 11050–11055.

[10] A.M. Fjell, K.B. Walhovd, I. Reinvang, A. Lundervold, D. Salat, B.T. Quinn, B. Fischl, and A.M. Dale, Selective increase of cortical thickness in high-performing elderly—Structural indices of optimal cognitive aging, NeuroImage 29 (2006), 984–994.

[11] G. Gong, Y. He, Z.J. Chen, and A.C. Evans, Convergence and divergence of thickness correlations with diffusion connections across the human cerebral cortex, NeuroImage 59 (2012), 1239–1248.

[12] A.L. Gross, J.J. Manly, J. Pa, J.K. Johnson, L.Q. Park, M.B. Mitchell, R.J. Melrose, S.K. Inouye, and D.G. McLaren, Cortical signatures of cognition and their relationship to Alzheimer's disease, Brain Imaging Behav 6 (2012), 584–598.

[13] S. Hakimi, On realizability of a set of integers as degrees of the vertices of a linear graph. I, J Soc Ind Appl Math 10 (1962), 496–506.

[14] S. Hakimi, On realizability of a set of integers as degrees of the vertices of a linear graph II: Uniqueness, J Soc Ind Appl Math 11 (1963), 135–147.

[15] P. Hansen and H. Mélot, Variable neighborhood search for extremal graphs, 6: Analyzing bounds for the connectivity index, J Chem Inf Comput Sci 43 (2003), 1–14.

[16] V. Havel, A remark on the existence of finite graphs, Casopis Pest Mat 80 (1955), 477–480.

[17] J.A. Hayes and J.L. Mendenhall, NeuronetExperimenter, webpage. Available at: https://sourceforge.net/projects/neuronetexp/, Access on 10 February 2014.

[18] Y. He, Z. Chen, and A. Evans, Structural insights into aberrant topological patterns of large-scale cortical networks in Alzheimer's disease, J Neuroscience 28 (2008), 4756–4766.

[19] Y. He, Z. Chen, G. Gong, and A. Evans, Neuronal networks in Alzheimer's disease, Neuroscientist 15 (2009), 333–350.

[20] R.K. Kincaid, N. Alexandrov, and M. Holroyd, An investigation of synchrony in transport networks, Complexity 14 (2009), 34–43.

[21] R.K. Kincaid and D. Phillips, Network topology measures, Wiley Interdiscip Rev 3 (2011), 557–565.

[22] L. Li, D. Alderson, J. Doyle, and W. Willinger, Towards a theory of scale-free graphs: Definition, properties, and implications, Internet Math 2 (2005), 431–523.

[23] X. Li, Y. Shi, and L. Wang, "An updated survey on the Randić index," Recent results in the theory of Randić index, Vol. 6 of Mathematical Chemistry Monographs, I. Gutman and B. Furtula (Editors), University of Kragujevac and Faculty of Science Kragujevac, Kragujevac, Serbia, 2008, pp. 9–47.

[24] Y.Y. Lui, J.J. Slotine, and A.L. Barabasi, Controllability of complex networks, Nature 473 (2011), 167–173.

[25] M. Müller-Hannemann and A. Schwartz, "Implementing weighted *b*-matching algorithms: Insights from a computational study," Algorithm engineering and experimentation, Vol. 1619 of Lecture Notes in Computer Science, M.T. Goodrich and C.C. McGeoch (Editors), Springer, Baltimore, MD, 1999, pp. 18–36.

[26] D.J. Phillips, A. McGlaughlin, D. Ruth, L.R. Jager, A. Soldan, and the Alzheimers Disease Neuroimaging Initiative, Graph theoretic analysis of structural connectivity across the spectrum of Alzheimers disease: The importance of graph creation methods, NeuroImage 7 (2015), 377–390.

[27] M. Randić, Characterization of molecular branching, J Am Chem Soc 97 (1975), 6609–6615.

[28] J.E. Rubin, J.A. Hayes, J.L. Mendenhall, and C.A. Del Negro, Calcium-activated nonspecific cation current and synaptic

depression promote network-dependent burst oscillations, Proc Natl Acad Sci USA 106 (2009), 2939–2944.

[29] A. Schrijver, Combinatorial optimization, Springer, Berlin, Germany, 2003.

[30] A. Taylor and D. Higham, CONTEST: A controllable test matrix toolbox for MATLAB, ACM Trans Math Software (TOMS) 35, Article 26 (2009).

[31] W.T. Tutte, A short proof of the factor theorem for finite graphs, Canad J Math 6 (1954), 347–352.

[32] V.V. Vazirani, Approximation algorithms, Springer-Verlag, Berlin, Germany, 2001.

[33] D. West, Introduction to graph theory, 2nd edition, Prentice Hall, Upper Saddle River, NJ, 2001.

[34] Z. Yao, Y. Zhang, L. Lin, Y. Zhou, C. Xu, T. Jiang, and the Alzheimer's Disease Neuroimaging Initiative, Abnormal cortical networks in mild cognitive impairment and Alzheimer's disease, PLoS Comput Biol 6 (2010), e1001006.

[35] G. Zamora-López, C. Zhou, V. Zlatić, and J. Kurths, The generation of random directed networks with prescribed 1-node and 2-node degree correlations, J Phys A 41 (2008), 224006.