Maximizing Barber’s bipartite modularity is also hard

Atsushi Miyauchi · Noriyoshi Sukegawa

Abstract Modularity introduced by Newman and Girvan (Phys Rev E 69:026113, 2004) is a quality function for community detection in networks. Numerous methods for modularity maximization have been developed so far. In 2007, Barber (Phys Rev E 76:066102, 2007) introduced a variant of modularity called bipartite modularity which is appropriate for bipartite networks. Although maximizing the standard modularity is known to be NP-hard, the computational complexity of maximizing bipartite modularity has yet to be revealed. In this study, we prove that maximizing bipartite modularity is also NP-hard. More specifically, we show the NP-completeness of its decision version by constructing a reduction from a classical partitioning problem.

Keywords Network analysis · Community detection · Modularity maximization · Bipartite networks · Computational complexity

1 Introduction

Networks have attracted much attention from diverse fields such as physics, informatics, chemistry, biology, sociology, and so forth. Many complex systems arising in such fields can be represented as networks, and analyzing the structures and dynamics of these networks provides meaningful information about the underlying systems [22, 24].

In network analysis, finding community structures is one of the most fundamental tasks. Roughly speaking, a community is a set of vertices densely connected inside, but sparsely connected with the rest of the network. Community detection analysis

A. Miyauchi (✉) · N. Sukegawa
Graduate School of Decision Science and Technology, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan
e-mail: miyauchi.a.aa@m.titech.ac.jp
is increasingly applied in various fields. For more details, see the useful survey by Fortunato [13] with over 450 references.

Community detection is now often conducted through maximizing a quality function called modularity introduced by Newman and Girvan [25]. This function was solely a quality measure at first, but nowadays it is widely used as an objective function of optimization problems for finding community structures. Modularity represents the sum, over all communities, of the fraction of the number of edges connecting vertices in a community minus the expected fraction of the number of such edges assuming that they are put at random with the same distribution of vertex degree. Let us consider an undirected network $G = (V, E)$ consisting of $n = |V|$ vertices and $m = |E|$ edges, and take a division $C$ of $V$, then modularity $Q$ can be written as:

$$Q(C) = \sum_{C \in \mathcal{C}} \left( \frac{m_C}{m} - \left( \frac{D_C}{2m} \right)^2 \right),$$

where $m_C$ is the number of edges connecting vertices in community $C$, and $D_C$ is the sum of the degrees of the vertices in community $C$. In 2008, Brandes et al. [6] provided the first computational complexity result for modularity maximization. More precisely, they showed that modularity maximization is NP-hard. Numerous heuristics based on greedy techniques [4,9,25], simulated annealing [17,19,20], spectral optimization [23], extremal optimization [12], dynamical clustering [5], mathematical programming [1,7,8,21], and other techniques have been developed. In addition, a few exact algorithms [2,26] have also been proposed.

In recent years, some authors have reported that modularity is not perfect because it has two drawbacks: the resolution limit [14] and degeneracies [16]. The former means that, when the number of edges is large, small communities tend to be put together even if they are cliques connected by only one edge. The latter means that there exist a large number of nearly optimal divisions in terms of modularity maximization, which makes finding communities with maximal modularity extremely difficult. Nevertheless, modularity maximization is regarded as the most popular approach for community detection.

In 2007, Barber [3] proposed a variant of modularity called bipartite modularity for community detection in bipartite networks. Needless to say, the standard modularity is applicable to bipartite networks. However, it does not reflect a structure and restrictions of bipartite networks, that is, the vertices in a bipartite network can be divided into two disjoint sets of red and blue vertices such that every edge connects a red vertex and a blue vertex. Barber’s bipartite modularity $Q_b$ does reflect them, and it can be represented as:

$$Q_b(C) = \sum_{C \in \mathcal{C}} \left( \frac{m_C}{m} - \frac{R_C B_C}{m^2} \right),$$

where $R_C$ is the sum of the degrees of the red vertices in community $C$, and $B_C$ is the same for the blue vertices. It can be seen that the terms $2m$ and $D^2_C$ in the standard modularity are replaced by $m$ and $R_C B_C$, respectively. These modifications
are due to the structure and restrictions of bipartite networks. As the standard modularity, many approaches for maximizing bipartite modularity have been proposed so far [3,11,27] because bipartite networks arise in various real-world systems. We note that Guimerà, Sales-Pardo, and Amaral [18] proposed another variant of modularity for community detection in bipartite networks. Their variant focuses on dividing either red or blue vertices only, while Barber’s bipartite modularity simultaneously divides the whole network. This is the most striking difference between these two variants.

As mentioned above, maximizing the standard modularity is known to be NP-hard. On the other hand, the computational complexity of maximizing bipartite modularity has yet to be revealed. In 2011, Zhan et al. [27] stated that maximizing the standard modularity can be reduced to maximizing bipartite modularity. If this is correct, then we can conclude that maximizing bipartite modularity is NP-hard. However, as pointed out by Costa and Hansen [10], their analysis includes a crucial error. In 2013, Costa and Hansen [11] stated that the computational complexity of maximizing bipartite modularity still remains open.

In this study, we prove that maximizing bipartite modularity is also NP-hard. To this end, we show the NP-completeness of its decision version by constructing a reduction from a classical partitioning problem. We note that our analysis is based on that of Brandes et al. [6] who succeeded in showing the NP-hardness of maximizing the standard modularity. However, due to the change of the objective function, we need a more detailed analysis.

2 Preliminaries

We study the following problem which is the decision version of maximizing bipartite modularity.

**Problem 1 (BIMODULARITY)** Given a bipartite network \( G = (V, E) \) and a real number \( K \), does there exist a division \( C \) of \( V \) such that \( Q_b(C) \geq K \)?

Our analysis employs the following partitioning problem as Brandes et al. [6] did for the standard modularity.

**Problem 2 (3-PARTITION)** Given a set of \( 3k \) positive integers \( A = \{a_1, a_2, \ldots, a_{3k}\} \) such that \( a = \sum_{i=1}^{3k} a_i = kb \) and \( b/4 < a_i < b/2 \) for \( i = 1, 2, \ldots, 3k \), for some integer \( b \), does there exist a partition of \( A \) into \( k \) sets such that the sum of the numbers in each set is equal to \( b \)?

3-PARTITION is NP-complete in the strong sense [15], which means that the problem cannot be solved even in pseudo-polynomial time, unless \( P = NP \). Therefore, to show the NP-completeness of BIMODULARITY, it is enough to construct a pseudo-polynomial time reduction from 3-PARTITION. In other words, we need to show that a given instance \( A \) of 3-PARTITION can be transformed in pseudo-polynomial time into a certain instance \( (G(A), K(A)) \) of BIMODULARITY such that \( G(A) \) has a division \( C \) of \( V \) which satisfies \( Q_b(C) \geq K(A) \) if and only if \( A \) can be partitioned into \( k \) sets with sum equal to \( b \) each.
We now introduce a variant of 3-PARTITION where instances are restricted to sets $A$ such that each $a_i \in A$ is a multiple of some positive integer $c$. This variant is denoted by 3-PARTITION($c$). We present the following proposition.

**Proposition 1** For any positive integer $c$, 3-PARTITION($c$) is NP-complete in the strong sense.

**Proof** 3-PARTITION($c$) belongs to the class NP. We now construct a polynomial time reduction from 3-PARTITION to 3-PARTITION($c$). Let us take an arbitrary instance $A = \{a_1, a_2, \ldots, a_{3k}\}$ of 3-PARTITION. Note that $A$ satisfies the conditions $a = \sum_{i=1}^{3k} a_i = kb$ and $b/4 < a_i < b/2$ for $i = 1, 2, \ldots, 3k$, for some integer $b$. We generate a new set $A' = \{a_1', a_2', \ldots, a_{3k}'\}$ by multiplying each element of $A$ by $c$, that is, $a_i' = ca_i$ for $i = 1, 2, \ldots, 3k$. Clearly, it can be done in polynomial time in the input size of $A$. We see that $A'$ is an instance of 3-PARTITION($c$) because (1) each $a_i' \in A'$ is a positive integer and a multiple of $c$ and (2) $\sum_{i=1}^{3k} a_i' = ca = k \cdot cb$ and $cb/4 < a_i' < cb/2$ for $i = 1, 2, \ldots, 3k$, for integer $cb$. It is clear that $A$ can be partitioned into $k$ sets with sum equal to $b$ each if and only if $A'$ can be partitioned into $k$ sets with sum equal to $cb$ each. □

Strictly speaking, we construct a pseudo-polynomial time reduction to BIMODULARITY not from 3-PARTITION but from 3-PARTITION(21). This is the reason why we introduced the variant here.

### 3 NP-completeness

We construct a pseudo-polynomial time reduction from 3-PARTITION(21) to BIMODULARITY. Let us take an arbitrary instance $A = \{a_1, a_2, \ldots, a_{3k}\}$ of 3-PARTITION(21).

We first consider the case when $k \leq 14$. In this particular case, we can confirm in polynomial time in the input size of $A$ whether $A$ can be partitioned into $k$ sets with sum equal to $b$ each by checking all possible divisions of $A$. Thus, it suffices to generate a yes-instance (for example, $(G(A), K(A))$ such that $G(A)$ is a singleton and $K(A) \leq 0$) if it turns out that $A$ can be partitioned, and to generate a no-instance (for example, $(G(A), K(A))$ such that $G(A)$ is a singleton and $K(A) > 0$) otherwise. This is the desired reduction for the case when $k \leq 14$.

In what follows, we consider the case when $k > 14$. We initially propose a procedure for generating appropriate bipartite network $G(A)$ from $A = \{a_1, a_2, \ldots, a_{3k}\}$ as follows:

**Step 1:** Construct $k$ complete bipartite networks (bicliques for short) $K_1, K_2, \ldots, K_k$ consisting of $a$ red vertices in one subset of vertices and $a$ blue vertices in the other subset of vertices.

**Step 2:** For each $a_i \in A$, put a red vertex $x_i$ and a blue vertex $y_i$. These are termed *element vertices*.

**Step 3:** For $i = 1, 2, \ldots, 3k$, connect $x_i$ to $a_i$ blue vertices in each of $k$ bicliques constructed in Step 1 such that each blue vertex in bicliques is connected...
Maximizing Barber’s bipartite modularity is also hard

Step 4: For \( i = 1, 2, \ldots, 3k \), connect the pair of element vertices \( x_i \) and \( y_i \).

Step 5: For \( i = 1, 2, \ldots, 3k \), construct a star \( X_i \) consisting of one blue internal vertex and \( a^2/7 \) red leaves, and construct a star \( Y_i \) consisting of one red internal vertex and \( a^2/7 \) blue leaves. (Note that \( a^2 \) is a multiple of 7 because \( A \) is an instance of 3-PARTITION(21)).

Step 6: For \( i = 1, 2, \ldots, 3k \), connect \( x_i \) to the internal vertex of \( X_i \), and connect \( y_i \) to the internal vertex of \( Y_i \).

This procedure generates a bipartite network \( G(A) \) consisting of

\[
  n = \frac{6}{7}ka^2 + 2ka + 12k
\]

vertices and

\[
  m = \frac{13}{7}ka^2 + 2ka + 9k
\]

edges. Clearly, it can be done in pseudo-polynomial time, that is, polynomial time in the sum of the input values of \( A \). We note that each vertex of the bicliques \( K_1, K_2, \ldots, K_k \) has degree \( a + 1 \), and for each \( a_i \in A \), the element vertices \( x_i \) and \( y_i \) have degrees \( ka_i + 2 \). In Fig. 1, \( G(A) \) constructed from \( A = \{2, 2, 2, 2, 3, 3\} \) is shown as an example.

Before determining appropriate parameter \( K(A) \) for the instance of BIMODULARITY, we observe several conditions satisfied by divisions of \( G(A) \) with maximal bipartite modularity.

**Lemma 1** In any division of \( G(A) \) with maximal bipartite modularity, none of the bicliques \( K_1, K_2, \ldots, K_k \) is divided.

**Proof** Let us consider an arbitrary division \( C \). Suppose that a biclique \( K_t \) is divided into \( l \) communities with \( l > 1 \) in division \( C \). We denote the communities containing vertices of \( K_t \) by \( C_1, C_2, \ldots, C_l \). The contribution of \( C_1, C_2, \ldots, C_l \) to \( Q_b \) can be written as:

\[
  \frac{1}{m} \sum_{i=1}^{l} m_i - \frac{1}{m^2} \sum_{i=1}^{l} R_i B_i,
\]

where \( m_i \) is the number of edges connecting vertices in \( C_i \), \( R_i \) is the sum of the degrees of the red vertices in \( C_i \), and \( B_i \) is the same for the blue vertices.

Transform \( C_1, C_2, \ldots, C_l \) into \( C'_1, C'_2, \ldots, C'_l \) by removing all the vertices of \( K_t \) from each community. We construct a new division \( C' \) by replacing \( C_1, C_2, \ldots, C_l \) in \( C \) with \( K_t, C'_1, C'_2, \ldots, C'_l \). For \( i = 1, 2, \ldots, l \), we denote the number of red vertices removed from \( C_i \) by \( r_i \), the same for the blue vertices by \( b_i \), and the number of edges between vertices of \( K_t \) in \( C_i \) and the element vertices in \( C'_j \) by \( f_{ij} \). Then, the decrement, due to the transformation from \( C \) into \( C' \), of the number of edges within communities is given by \( \sum_{i=1}^{l} f_{ii} \). On the other hand, the increment of the number of such edges can be represented as \( \sum_{i=1}^{l} \sum_{j \neq i} r_i b_j \) because biclique \( K_t \) is added to \( C' \) as a new
Fig. 1 Bipartite network $G(A)$ constructed from $A = \{2, 2, 2, 2, 3, 3\}$; labels of vertices represent corresponding elements in $A$

community. Additionally, as for the sum of the degrees of $C_i$, the red one decreases $(a + 1)r_i$ and the blue one decreases $(a + 1)b_i$ because each vertex of $K_t$ has degree $a + 1$. From the above, the contribution of $K_t, C'_1, C'_2, \ldots, C'_l$ to $Q_b$ is calculated by

$$
\frac{1}{m} \left( \sum_{i=1}^{l} m_i - \sum_{i=1}^{l} f_i + \sum_{i=1}^{l} \sum_{j \neq i} r_i b_j \right) - \frac{1}{m^2} \left( (a + 1)^2 a^2 + \sum_{i=1}^{l} (R_i - (a + 1)r_i)(B_i - (a + 1)b_i) \right).
$$

Thus, we see that

© Springer
Maximizing Barber’s bipartite modularity is also hard

\[
\Delta := Q_b(C') - Q_b(C) = \frac{1}{m} \left( \sum_{i=1}^{l} \sum_{j \neq i}^{l} r_i b_j - \sum_{i=1}^{l} f_i \right) \\
+ \frac{1}{m^2} \left( (a + 1) \left( \sum_{i=1}^{l} R_i b_i + \sum_{i=1}^{l} B_i r_i - (a + 1) \sum_{i=1}^{l} r_i b_i \right) \\
- (a + 1)^2 a^2 \right).
\]

Now, for \( i = 1, 2, \ldots, l \), decompose the set of the edges enumerated by \( f_i \) into the set of \( f_{i1} \) edges and the set of \( f_{i2} \) edges which are incident to red vertices and blue vertices of \( K_t \), respectively. Then, we have that \( R_i \geq (a + 1)r_i + kf_{i2} \) for \( i = 1, 2, \ldots, l \). This inequality holds because (1) \( C_i \) has at least \( r_i \) red vertices of \( K_t \) with degree \( a + 1 \) and (2) it also contains some red element vertices which have at least \( f_{i2} \) edges connected to biclique \( K_t \), and such element vertices again have at least \( f_{i2} \) edges to each of the other \( k - 1 \) bicliques. Arguing similarly for blue vertices, we also have that \( B_i \geq (a + 1)b_i + kf_{i1} \) for \( i = 1, 2, \ldots, l \). Therefore, it holds that

\[
\sum_{i=1}^{l} R_i b_i \geq \sum_{i=1}^{l} ((a + 1)r_i + kf_{i2}) b_i = (a + 1) \sum_{i=1}^{l} r_i b_i + k \sum_{i=1}^{l} f_{i2} b_i,
\]

and

\[
\sum_{i=1}^{l} B_i r_i \geq \sum_{i=1}^{l} ((a + 1)b_i + kf_{i1}) r_i = (a + 1) \sum_{i=1}^{l} r_i b_i + k \sum_{i=1}^{l} f_{i1} r_i.
\]

Using these inequalities and the following equalities

\[
\sum_{i=1}^{l} \sum_{j \neq i}^{l} r_i b_j = \sum_{i=1}^{l} \sum_{j=1}^{l} r_i b_j - \sum_{i=1}^{l} r_i b_i = a^2 - \sum_{i=1}^{l} r_i b_i,
\]

we see that

\[
\Delta \geq \frac{1}{m} \left( a^2 - \sum_{i=1}^{l} r_i b_i - \sum_{i=1}^{l} f_i \right) \\
+ \frac{1}{m^2} \left( (a + 1)^2 \sum_{i=1}^{l} r_i b_i + k(a + 1) \sum_{i=1}^{l} (f_{i1} r_i + f_{i2} b_i) - (a + 1)^2 a^2 \right) \\
= \frac{1}{m^2} \left( ma^2 - (a + 1)^2 a^2 - (m - (a + 1)^2) \sum_{i=1}^{l} r_i b_i \right).
\]

Springer
\[-m \sum_{i=1}^{l} f_i + k(a + 1) \sum_{i=1}^{l} (f_{i1} r_i + f_{i2} b_i)\].

Focusing on the last two terms in the above parenthesis, we see that

\[-m \sum_{i=1}^{l} f_i + k(a + 1) \sum_{i=1}^{l} (f_{i1} r_i + f_{i2} b_i)\]

\[= \sum_{i=1}^{l} f_{i1} (k(a + 1) r_i - m) + \sum_{i=1}^{l} f_{i2} (k(a + 1) b_i - m)\]

\[\geq \sum_{i=1}^{l} r_i (k(a + 1) r_i - m) + \sum_{i=1}^{l} b_i (k(a + 1) b_i - m)\]

\[= k(a + 1) \sum_{i=1}^{l} \left( r_i^2 + b_i^2 \right) - 2ma \geq 2k(a + 1) \sum_{i=1}^{l} r_i b_i - 2ma.\]

The first equality follows because \(f_i = f_{i1} + f_{i2}\) for \(i = 1, 2, \ldots, l\). The next inequality follows because for \(i = 1, 2, \ldots, l\), we have (1) \(r_i \geq f_{i1}\) and \(k(a + 1) r_i - m < 0\) and (2) \(b_i \geq f_{i2}\) and \(k(a + 1) b_i - m < 0\). Therefore, it holds that

\[\Delta \geq \frac{1}{m^2} \left( ma^2 - (a + 1)^2 a^2 - 2ma - \left(m - (a + 1)^2 - 2k(a + 1) \right) l r_i b_i \right).\]

We now have \(m - (a + 1)^2 - 2k(a + 1) > 0\) as \(k > 14\). Moreover, we can show that \(\sum_{i=1}^{l} r_i b_i \leq a(a - 1)\). To obtain this, it suffices to show that there exist at least \(a\) edges in \(K_t\) which connect different communities in \(C\). If there exists a community \(C_i\) such that \(r_i = 0\), then we immediately obtain the claim. Thus, we consider only the case when \(r_i \geq 1\) for \(i = 1, 2, \ldots, l\). Let us take an arbitrary community \(C_i\) from \(C_1, C_2, \ldots, C_l\). Since there exists another community \(C_j\) with \(r_j \geq 1\), we have \(r_i \leq a - 1\). Hence, the number of edges in \(K_t\) which connect different communities in \(C\) is at least

\[r_i (a - b_i) + (a - r_i) b_i \geq (a - b_i) + b_i = a.\]

Using the above inequalities, we see that

\[\Delta \geq \frac{a}{m^2} \left( ma - (a + 1)^2 a - 2m - \left(m - (a + 1)^2 - 2k(a + 1) \right) (a - 1) \right)\]

\[= \frac{a}{m^2} \left( 2k(a + 1) a - m - (a + 1)^2 - 2k(a + 1) \right)\]

\[= \frac{a}{m^2} \left( \left(\frac{1}{7} a^2 - 2a - 11 \right) k - a^2 - 2a - 1 \right).\]
Since \( k > 14 \) and \( a \geq 3k > 42 \), we have

\[
\Delta > \frac{a}{m^2} \left( a^2 - 30a - 155 \right) > 0.
\]

Taking an optimal division as \( C \) at first, we obtain a contradiction. \( \square \)

**Lemma 2** In any division of \( G(A) \) with maximal bipartite modularity, every community contains at most one of the bicliques \( K_1, K_2, \ldots, K_k \).

**Proof** Let us consider an arbitrary optimal division \( C^* \). From Lemma 1, none of the bicliques \( K_1, K_2, \ldots, K_k \) is divided. Suppose that \( C^* \) has a community \( C \) which contains \( I \) of the bicliques with \( I \geq 2 \). The set of indices of the red element vertices in \( C \) is denoted by \( I_R \), and the same for the blue vertices is denoted by \( I_B \). The number of edges in \( C \) whose at least one endpoint is a vertex of \( X_i \) or \( Y_i \) for some \( i \in \{1, 2, \ldots, 3k\} \) is denoted by \( m_{\text{star}} \). Additionally, the sum of the degrees of the red vertices in \( C \) other than the ones in \( I \) bicliques is denoted by \( R \), and the same for the blue vertices is denoted by \( B \). Then, the contribution of \( C \) to \( Q_b \) is calculated by

\[
\frac{1}{m} \left( a^2 I + l \left( \sum_{i \in I_R} a_i + \sum_{i \in I_B} a_i \right) + |I_R \cap I_B| + m_{\text{star}} \right)
\]

\[
- \frac{1}{m^2} \left( (a + 1)l + (a + 1)(l - 1) \right). \]

Note that \( |I_R \cap I_B| \) enumerates the number of edges between the element vertices \( x_i \) and \( y_i \) in \( C \).

Let us take an arbitrary biclique \( K_t \) from the bicliques contained in \( C \). Construct a new division \( C' \) by dividing \( C \) into \( K_t \) and the rest \( C' \). Clearly, the increment, due to this transformation, of the number of edges within communities is 0. On the other hand, the decrement is given by \( \sum_{i \in I_R} a_i + \sum_{i \in I_B} a_i \) because only the edges between \( K_t \) and the element vertices in \( C \) are cut. Thus, the contribution of \( K_t \) and \( C' \) to \( Q_b \) is calculated by

\[
\frac{1}{m} \left( a^2 I + (l - 1) \left( \sum_{i \in I_R} a_i + \sum_{i \in I_B} a_i \right) + |I_R \cap I_B| + m_{\text{star}} \right)
\]

\[
- \frac{1}{m^2} \left( (a + 1)^2 l^2 + (a + 1)l(l - 1) + R \right). \]

Hence, we see that

\[
\Delta := Q_b(C') - Q_b(C^*)
\]

\[
= -\frac{1}{m} \left( \sum_{i \in I_R} a_i + \sum_{i \in I_B} a_i \right) + \frac{1}{m^2} \left( 2(a + 1)^2 l^2 + (a + 1)l(R + B) \right)
\]

\[
= \frac{1}{m^2} \left( 2(a + 1)^2 l^2 + (a + 1)l(R + B) - m \left( \sum_{i \in I_R} a_i + \sum_{i \in I_B} a_i \right) \right)
\]
\[ \geq \frac{2a}{m^2} ((a+1)^2 a - m) \geq \frac{2a}{m^2} (3k(a+1)^2 - m) = \frac{2ka}{m^2} \left( \frac{8}{7} a^2 + 4a - 6 \right) > 0. \]

The first inequality follows because we have \( l \geq 2, R + B \geq 0 \), and \( \sum_{i \in I_B} a_i + \sum_{i \in I_R} a_i \leq 2a \). The second inequality follows because we have \( 2(a+1)^2/m^2 > 0 \) and \( a \geq 3k \). This contradicts the optimality of division \( C^* \). \( \square \)

**Lemma 3** In any division of \( G(A) \) with maximal bipartite modularity, none of the stars \( X_1, X_2, \ldots, X_{3k} \) and \( Y_1, Y_2, \ldots, Y_{3k} \) is divided.

**Proof** Let us consider an arbitrary optimal division \( C^* \), and an arbitrary leaf \( l \). It suffices to show that \( l \) is contained in the community to which its adjacent vertex belongs. Suppose otherwise, that is, \( l \) belongs to a community \( C_1 \) and its adjacent vertex belongs to another community \( C_2 \). Assume now that \( l \) is a red leaf. (Note that the following discussion is also applicable to every blue leaf.)

Construct a new division \( C' \) by transferring \( l \) from \( C_1 \) to \( C_2 \). The increment of \( Q_b \) is \( \frac{1}{m} \) due to the edge between \( l \) and its adjacent vertex. On the other hand, the decrement of \( Q_b \) is strictly less than \( \frac{1}{m} \) because (1) the degree of \( l \) is 1 and (2) the sum of the degrees of the blue vertices in \( C_2 \) is less than \( m \) as the set of the blue vertices in the whole network is divided into at least \( k > 14 \) communities by Lemma 2. Thus, we obtain \( Q_b(C') > Q_b(C^*) \), which contradicts the optimality of division \( C^* \). \( \square \)

**Lemma 4** In any division of \( G(A) \) with maximal bipartite modularity, every red element vertex \( x_i \) and the adjacent star \( X_i \) are not contained in the same community. The same statement holds for every blue element vertex \( y_i \) and the adjacent star \( Y_i \).

**Proof** Let us consider an arbitrary optimal division \( C^* \), and an arbitrary red element vertex \( x_i \). Suppose that \( C^* \) has a community \( C \) which contains both \( x_i \) and the adjacent star \( X_i \). Now, the sum of the degrees of the red vertices in \( C \) other than \( X_i \) is denoted by \( R \), and the same for the blue vertices is denoted by \( B \). Then, the contribution of \( C \) to \( Q_b \) is given by

\[ \frac{m_C}{m} - \frac{1}{m^2} \left( R + \frac{a^2}{7} \right) \left( B + \left( \frac{a^2}{7} + 1 \right) \right), \]

because \( X_i \) consists of one blue vertex with degree \( a^2/7 + 1 \) and \( a^2/7 \) red leaves.

Construct a new division \( C' \) by dividing \( C \) into \( X_i \) and the rest \( C' \). Clearly, the contribution of \( X_i \) and \( C' \) to \( Q_b \) is given by

\[ \frac{m_C - 1}{m} - \frac{1}{m^2} \left( RB + \frac{a^2}{7} \left( \frac{a^2}{7} + 1 \right) \right). \]

Hence, we see that

\[ \Delta := Q_b(C') - Q_b(C^*) = -\frac{1}{m} + \frac{1}{m^2} \left( \left( \frac{a^2}{7} + 1 \right) R + \frac{a^2}{7} B \right). \]
Since $C$ at least contained element vertex $x_i$ other than $X_i$, we have $R \geq k a_i + 2$. Using this inequality and $B \geq 0$, we see that

$$\Delta \geq -\frac{1}{m} + \frac{1}{m^2} \left( \frac{a^2}{7} + 1 \right) (k a_i + 2).$$

Since $A$ is an instance of 3-PARTITION(21), we have $a_i \geq 21$. Thus, we immediately obtain $\Delta > 0$. This contradicts the optimality of division $C^*$. It is easy to see that the above discussion is applicable to every blue element vertex $y_i$ and the adjacent star $Y_i$. □

**Lemma 5** In any division of $G(A)$ with maximal bipartite modularity, every star $X_1, X_2, \ldots, X_{3k}$ and $Y_1, Y_2, \ldots, Y_{3k}$ itself forms a community.

**Proof** Let us consider an arbitrary optimal division $C^*$, and an arbitrary star $X_i$. From Lemma 3, $X_i$ is entirely contained in a community $C$. Therefore, it suffices to show that $C$ contains no vertices other than $X_i$. Suppose otherwise, that is, some vertex other than $X_i$ belongs to $C$. Since the only adjacent vertex $x_i$ of $X_i$ is not contained in $C$ from Lemma 4, $X_i$ is not connected with the other vertices in $C$. Constructing a new division $C'$ by dividing $C$ into $X_i$ and the rest, we obtain $Q_b(C') > Q_b(C^*)$. This contradicts the optimality of $C^*$. The above discussion also holds for an arbitrary star $Y_i$. □

**Lemma 6** In any division of $G(A)$ with maximal bipartite modularity, every element vertex belongs to one of the communities corresponding to the bicliques $K_1, K_2, \ldots, K_k$.

**Proof** Let us consider an arbitrary optimal division $C^*$. From Lemma 5, it suffices to show that there exists no community consisting of element vertices only. Suppose otherwise, that is, $C^*$ has a community $C$ which consists of element vertices only. In what follows, we consider the following two cases: when $C$ contains both red and blue element vertices, and when $C$ contains either red or blue element vertices only.

First, we analyze the former case. If there exists a vertex which has no neighbors in $C$, then the objective value can be strictly improved by removing the vertex as a new community. Thus, $C$ consists of some pairs $x_i$ and $y_j$. Note that if $C$ contains two or more such pairs, then we similarly obtain a contradiction. Hence, we see that $C$ consists of only one pair $x_i$ and $y_j$.

From Lemmas 1 and 2, there exist communities $C_1, C_2, \ldots, C_k$ corresponding to the bicliques $K_1, K_2, \ldots, K_k$. In the following, these communities are termed *biclique communities*. Assume that $C_{\text{min}}$ is one of those communities whose sum of the degrees is minimal. Now, the set of indices of the red element vertices in $C_{\text{min}}$ is denoted by $I_R$, and the same for the blue vertices is denoted by $I_B$. Then, the contribution of $C$ and $C_{\text{min}}$ to $Q_b$ is calculated by

$$\frac{1}{m} \left( a^2 + \sum_{i \in I_R} a_i + \sum_{i \in I_B} a_i + |I_R \cap I_B| + 1 \right)$$

$$- \frac{1}{m^2} \left( (a+1)a + \sum_{i \in I_R} (k a_i + 2) \right) \left( (a+1)a + \sum_{i \in I_B} (k a_i + 2) \right) + (k a_i + 2)^2.$$
Construct a new division $C'$ by merging $C$ and $C_{\min}$ into one community $C'$. The contribution of $C'$ to $Q_b$ is calculated by

$$\frac{1}{m} \left( a^2 + \sum_{i \in I_R} a_i + \sum_{i \in I_B} a_i + |I_R \cap I_B| + 1 + 2a_t \right)$$

$$- \frac{1}{m^2} \left( (a+1)a + \sum_{i \in I_R} (ka_i + 2) + (ka_i + 2) \right) \left( (a+1)a + \sum_{i \in I_B} (ka_i + 2) + (ka_i + 2) \right).$$

Thus, we see that

$$\Delta := Q_b(C') - Q_b(C^*)$$

$$= \frac{2}{m} a_t - \frac{1}{m^2} \left( 2(a+1)a(ka_t + 2) + (ka_t + 2) \left( \sum_{i \in I_R} (ka_i + 2) + \sum_{i \in I_B} (ka_i + 2) \right) \right).$$

Recall now that $C_{\min}$ is the biclique community whose sum of the degrees is minimal. Thus, the sum of the degrees of $C_{\min}$ is less than or equal to the average of that of all biclique communities. Moreover, no biclique community contained element vertices $x_t$ and $y_t$. Therefore, it holds that

$$\sum_{i \in I_R} (ka_i + 2) + \sum_{i \in I_B} (ka_i + 2) \leq \frac{1}{k} (2ka + 12k - 2(ka_t + 2)) < 2(a + 6).$$

Using these inequalities, we see that

$$\Delta > \frac{2}{m^2} \left( ma_t - (ka_t + 2)(a^2 + 2a + 6) \right)$$

$$\geq \frac{2}{m^2} \left( 18ka^2 + 63k - 2a^2 - 4a - 12 \right)$$

$$> \frac{2}{m^2} \left( 250a^2 - 4a + 870 \right) > 0.$$ 

The second inequality follows from $a_t \geq 21$. The third inequality follows from $k > 14$. This contradicts the optimality of division $C^*$.  

Next, we analyze the latter case. Assume that $C$ consists of red element vertices only. (Note that the following discussion is also applicable to every community which consists of blue element vertices only). In this case, we assume that $C_{\min}$ is one of the biclique communities whose sum of the degrees of the blue vertices is minimal. The set of indices of the red element vertices in $C_{\min}$ is denoted by $I_R$, and the same for the blue vertices is denoted by $I_B$. Then, the contribution of $C$ and $C_{\min}$ to $Q_b$ is calculated by

$$\frac{1}{m} \left( a^2 + \sum_{i \in I_R} a_i + \sum_{i \in I_B} a_i + |I_R \cap I_B| \right).$$

\(\text{Springer}\)
Maximizing Barber’s bipartite modularity is also hard

\[-\frac{1}{m^2} \left( (a + 1)a + \sum_{i \in I_R} (ka_i + 2) \right) \left( (a + 1)a + \sum_{i \in I_B} (ka_i + 2) \right).\]

Construct a new division \(C'\) by transferring an arbitrary vertex \(x_t\) from \(C\) to \(C_{\text{min}}\). If the corresponding element vertex \(y_t\) is not contained in \(C_{\text{min}}\), then the contribution of updated \(C\) and \(C_{\text{min}}\) to \(Q_b\) is calculated by

\[\frac{1}{m} \left( a^2 + \sum_{i \in I_R} a_i + \sum_{i \in I_B} a_i + |I_R \cap I_B| + a_t \right)\]

\[\quad - \frac{1}{m^2} \left( (a + 1)a + \sum_{i \in I_R} (ka_i + 2) + (ka_t + 2) \right) \left( (a + 1)a + \sum_{i \in I_B} (ka_i + 2) \right).\]

Note that if \(y_t\) is contained in \(C_{\text{min}}\), \(1/m\) is added to the contribution. Thus, we see that

\[\Delta := Q_b(C') - Q_b(C^*)\]

\[\geq \frac{1}{m} a_t - \frac{1}{m^2} \left( (a + 1)a(ka_t + 2) + (ka_t + 2) \sum_{i \in I_B} (ka_i + 2) \right).\]

Recall now that \(C_{\text{min}}\) is the biclique community whose sum of the degrees of the blue vertices is minimal. Thus, the sum of the degrees of the blue vertices in \(C_{\text{min}}\) is less than or equal to the average of that of all biclique communities. Moreover, no biclique community contained element vertex \(x_t\). Therefore, it holds that

\[\sum_{i \in I_B} (ka_i + 2) \leq \frac{1}{k} (ka + 6k - (ka_t + 2)) < a + 6.\]

Using these inequalities, we have

\[\Delta > \frac{1}{m^2} \left( ma_t - (ka_t + 2)(a^2 + 2a + 6) \right).\]

In the analysis of the former case, we have already shown \(2\Delta > 0\), which contradicts the optimality of \(C^*\).

\[\Box\]

**Lemma 7** In any division of \(G(A)\) with maximal bipartite modularity, every pair of element vertices \(x_i\) and \(y_i\) belongs to the same community.

**Proof** Let us consider an arbitrary optimal division \(C^*\), and an arbitrary pair of element vertices \(x_i\) and \(y_i\). Suppose that \(x_i\) belongs to a community \(C_1\) and \(y_i\) belongs to another community \(C_2\). From Lemma 6, we see that \(C_1\) and \(C_2\) are both biclique communities. The set of indices of the blue element vertices in \(C_1\) and \(C_2\) are denoted by \(I_{B_1}\) and \(I_{B_2}\), respectively.
Construct a new division $C'$ by transferring $x_t$ from $C_1$ to $C_2$. The increment of $Q_b$ is $1/m$ due to the edge between $x_t$ and $y_t$. On the other hand, the decrement of $Q_b$ is calculated by

$$\frac{1}{m^2}(ka_t + 2)\left(\sum_{i \in I_{B_2}} (ka_i + 2) - \sum_{i \in I_{B_1}} (ka_i + 2)\right),$$

because (1) the degree of $x_t$ is $ka_t + 2$ and (2) the sum of the degrees of the blue vertices in $C_2$ minus that of $C_1$ is $\sum_{i \in I_{B_2}} (ka_i + 2) - \sum_{i \in I_{B_1}} (ka_i + 2)$. Thus, we see that $\Delta := Q_b(C') - Q_b(C^*)$

$$= \frac{1}{m} - \frac{1}{m^2}(ka_t + 2)\left(\sum_{i \in I_{B_2}} (ka_i + 2) - \sum_{i \in I_{B_1}} (ka_i + 2)\right)$$

$$\geq \frac{1}{m} - \frac{1}{m^2}(ka_t + 2)(ka + 6k)$$

$$> \frac{1}{m} - \frac{1}{m^2}\left(\frac{a}{2} + 2\right)(ka + 6k) = \frac{k}{m^2}\left(\frac{19}{14}a^2 - 3a - 3\right).$$

The first inequality follows because the sum of the degrees of the blue element vertices in the whole network is $ka + 6k$. The second inequality follows because we have $a_t < b/2 = a/2k$. Since $a \geq 3k > 42$, we immediately obtain $\Delta > 0$. This contradicts the optimality of $C^*$. □

So far, we have observed the conditions satisfied by divisions of $G(A)$ with maximal bipartite modularity. Finally, combining these findings, we present our result.

**Theorem 1** BIMODULARITY is NP-complete in the strong sense.

**Proof** Since bipartite modularity for a given division $C$ can be computed in polynomial time, BIMODULARITY belongs to the class NP. We now complete the reduction to show the NP-completeness. Recall that it suffices to provide appropriate parameter $K(A)$ such that $G(A)$ has a division $C$ of $V$ which satisfies $Q_b(C) \geq K(A)$ if and only if $A$ can be partitioned into $k$ sets with sum equal to $b$ each.

From the above observations, an arbitrary optimal division $C^*$ of $G(A)$ in terms of maximizing bipartite modularity can be represented as:

$$\{C_1, C_2, \ldots, C_k, X_1, X_2, \ldots, X_{3k}, Y_1, Y_2, \ldots, Y_{3k}\},$$

where $C_1, C_2, \ldots, C_k$ are the biclique communities. Note that every pair of element vertices $x_j$ and $y_j$ belongs to one of the biclique communities. In this situation, the number of edges within communities is unvarying. More specifically, denoting the number of such edges by $m_{\text{intra}}$, we have

$$m_{\text{intra}} = m - (2a(k - 1) + 6k),$$

because the number of edges between different communities is always exactly $2a(k - 1) + 6k$. Thus, we see that division $C^*$ minimizes $\sum_{i=1}^{k} R_C; B_{C_i}$. Since
Maximizing Barber’s bipartite modularity is also hard

\( R_{C_i} = B_{C_i} \) for \( i = 1, 2, \ldots, k \), it can be replaced by \( \sum_{i=1}^{k} R_{C_i}^2 \). By the arithmetic mean-geometric mean inequality, this sum of squares has a lower bound:

\[
k \left( \prod_{i=1}^{k} R_{C_i}^2 \right)^{1/k}.
\]

Now, the sum of the degrees of the red vertices in all the biclique communities is given by

\[
\sum_{i=1}^{k} R_{C_i} = ka(a + 1) + ka + 6k = k(a^2 + 2a + 6).
\]

The above lower bound is attained if and only if

\[
R_{C_1}^2 = R_{C_2}^2 = \cdots = R_{C_k}^2,
\]

that is, for \( i = 1, 2, \ldots, k \),

\[
R_{C_i} = \frac{1}{k} \sum_{i=1}^{k} R_{C_i} = a^2 + 2a + 6,
\]

which means that all the sum of the degrees of the red vertices in each biclique community are the same. The lower bound is

\[
k(a^2 + 2a + 6)^2.
\]

Assume now that the lower bound is attained by division \( \mathcal{C}^* \). Then, we see that the sum of the degrees of the red element vertices in each biclique community is equal to

\[
a^2 + 2a + 6 - (a + 1)a = a + 6.
\]

This implies that the number of red element vertices in each biclique community is exactly three because we have \( b/4 < a_i < b/2 \) for \( i = 1, 2, \ldots, 3k \). Thus, for each biclique community, three red element vertices, say \( x_s, x_t, \) and \( x_u \), satisfy

\[
(ka_s + 2) + (ka_t + 2) + (ka_u + 2) = a + 6.
\]

This leads that \( a_s + a_t + a_u = a/k = b \). Therefore, \( A \) of 3-PARTITION can be partitioned into \( k \) sets with sum equal to \( b \) each.

Conversely, assume that \( A \) can be partitioned into \( k \) sets with sum equal to \( b \) each. Then, we can assign three red element vertices, say \( x_s, x_t, \) and \( x_u \), such that \( a_s + a_t + a_u = b \) to each biclique community. It is easy to see that the lower bound can be attained by \( G(A) \) constructed from such instance \( A \).
From the above, we should take $K(A)$ which is realized when the lower bound of the sum of squares is attained. Thus, we determine $K(A)$ as follows:

$$K(A) = \frac{m_{\text{intra}}}{m} - \frac{k(a^2 + 2a + 6)^2 + \frac{6}{7}ka^2\left(\frac{1}{7}a^2 + 1\right)}{m^2} = 1 - \frac{2a(k - 1) + 6k}{m} - \frac{k(a^2 + 2a + 6)^2 + \frac{6}{7}ka^2\left(\frac{1}{7}a^2 + 1\right)}{m^2}.$$ 

This completes the desired reduction. □

Acknowledgments

The authors would like to thank the anonymous referee for helpful comments. The authors are supported by the Grant-in-Aid for JSPS Fellows.

References

1. Agarwal, G., Kempe, D.: Modularity-maximizing graph communities via mathematical programming. Eur. Phys. J. B 66, 409–418 (2008)
2. Aloise, D., Cafieri, S., Caporossi, G., Hansen, P., Perron, S., Liberti, L.: Column generation algorithms for exact modularity maximization in networks. Phys. Rev. E 82, 046112 (2010)
3. Barber, M.J.: Modularity and community detection in networks. Phys. Rev. E 76, 066102 (2007)
4. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. P10008 (2008)
5. Boccaletti, S., Ivanchenko, M., Latora, V., Pluchino, A., Rapisarda, A.: Detecting complex network modularity by dynamical clustering. Phys. Rev. E 75, 045102 (2007)
6. Brandes, U., Delling, D., Gaertler, M., Görgke, R., Hoefer, M., Nikoloski, Z., Wagner, D.: On modularity clustering. IEEE Trans. Knowl. Data Eng. 20, 172–188 (2008)
7. Cafieri, S., Costa, A., Hansen, P.: Reformulation of a model for hierarchical divisive graph modularity maximization. Ann. Oper. Res. (in press). doi: 10.1007/s10479-012-1286-z
8. Cafieri, S., Hansen, P., Liberti, L.: Locally optimal heuristic for modularity maximization of networks. Phys. Rev. E 83, 056105 (2011)
9. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Phys. Rev. E 70, 066111 (2004)
10. Costa, A., Hansen, P.: Comment on “Evolutionary method for finding communities in bipartite networks”. Phys. Rev. E 84, 058101 (2011)
11. Costa, A., Hansen, P.: A locally optimal hierarchical divisive heuristic for bipartite modularity maximization. Optim. Lett. 8, 903–917 (2014)
12. Duch, J., Arenas, A.: Community detection in complex networks using extremal optimization. Phys. Rev. E 72, 027104 (2005)
13. Fortunato, S.: Community detection in graphs. Phys. Rep. 486, 75–174 (2010)
14. Fortunato, S., Barthélemy, M.: Resolution limit in community detection. Proc. Natl. Acad. Sci. USA 104, 36–41 (2007)
15. Garey, M.R., Johnson, D.S.: Computers and intractability: a guide to the theory of NP-completeness. WH Freeman, New York (1979)
16. Good, B.H., de Montjoye, Y.-A., Clauset, A.: Performance of modularity maximization in practical contexts. Phys. Rev. E 81, 046106 (2010)
17. Guimerà, R., Amaral, L.A.N.: Functional cartography of complex metabolic networks. Nature 433, 895–900 (2005)
18. Guimerà, R., Sales-Pardo, M., Amaral, L.A.N.: Module identification in bipartite and directed networks. Phys. Rev. E 76, 036102 (2007)
19. Massen, C.P., Doye, J.P.K.: Identifying communities within energy landscapes. Phys. Rev. E 71, 046101 (2005)
20. Medus, A.D., Acuña, G., Dorso, C.O.: Detection of community structures in networks via global optimization. Physica A 358, 593–604 (2005)
21. Miyauchi, A., Miyamoto, Y.: Computing an upper bound of modularity. Eur. Phys. J. B 86, 302 (2013)
22. Newman, M.E.J.: The structure and function of complex networks. SIAM Rev. 45, 167–256 (2003)
23. Newman, M.E.J.: Modularity and community structure in networks. Proc. Natl. Acad. Sci. USA 103, 8577–8582 (2006)
24. Newman, M.E.J.: Networks: an introduction. Oxford University Press, Oxford (2009)
25. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E 69, 026113 (2004)
26. Xu, G., Tsoka, S., Papageorgiou, L.G.: Finding community structures in complex networks using mixed integer optimisation. Eur. Phys. J. B 60, 231–239 (2007)
27. Zhan, W., Zhang, Z., Guan, J., Zhou, S.: Evolutionary method for finding communities in bipartite networks. Phys. Rev. E 83, 066120 (2011)