Maximizing edge-ratio is NP-complete

Steven D. Noble*, Pierre Hansen† and Nenad Mladenović‡

February 27, 2012

Abstract

Given a graph G and a bipartition of its vertices, the edge-ratio is the minimum for both classes so defined of their number of internal edges divided by their number of cut edges. We prove that maximizing edge-ratio is NP-complete.

Keywords: graph clustering, community detection, edge-ratio, NP-complete

1 Introduction

Clustering in graphs, also known as community detection, has been much studied recently (see [4] for a review with over 400 references). The algorithms proposed are hierarchical divisive and agglomerative as well as partitioning ones. Roughly speaking, a cluster should have more internal edges, joining two vertices within the cluster, than cut edges, joining a vertex in a cluster to a vertex outside of it. Many criteria have been proposed to evaluate the quality of a partition into clusters. The best known among these are *modularity* [7] for partitioning and *conductance* [6] for (recursive) bi-partitioning. Modularity maximization, and conductance minimization have been shown to be NP-complete [1, 6]. Therefore many heuristic optimization methods have been proposed for solving large problem instances [4]. Very recently a new criterion has been proposed for hierarchical divisive clustering [2]. There the *edge-ratio* is defined as the minimum for both classes so defined of their number of internal edges divided by their number of cut edges. We prove that maximizing edge-ratio is NP-complete. Based on this result, future work may consist of designing heuristic solution methods for solving clustering problems using edge-ratio as the criterion. It would also be interesting to determine bounds on the performance ratios of polynomial time approximation algorithms.

*Corresponding author: Brunel University, UB8 3PH, United Kingdom, steven.noble@brunel.ac.uk

†GERAD, HEC Montréal, Montréal (Québec) Canada, and LIX, Ecole Polytechnique Palaeseau, France, pierre.hansen@gerad.ca

‡Brunel University, UB8 3PH, United Kingdom, nenad.mladenovic@brunel.ac.uk

Most of our notation is standard, see for example [3]. All graphs considered are simple, that is, they do not have multiple edges or loops. Given a graph G with vertex set $V(G)$ and edge set $E(G)$ and a bipartition of G, comprising $X, Y \subseteq V(G)$, with $X \neq \emptyset$, $Y \neq \emptyset$, $X \cap Y = \emptyset$, we use $e_G(X)$ to denote the number of edges of G with both endpoints in X and $e_G(X, Y)$ to denote the number of edges of G which have one endpoint in X and one endpoint in Y. Most of the time, we will drop the subscript G. We use $e(G)$ to denote $|E(G)|$.

Given a graph G and a partition (X, Y) of $V(G)$, we define the *edge-ratio* of (X, Y), denoted by $r(X, Y)$, to be

$$r(X, Y) = \min\left\{\frac{e(X)}{e(X, Y)}, \frac{e(Y)}{e(X, Y)}\right\}.$$

The *edge-ratio* of a graph G is the maximum edge-ratio of any partition (X, Y) of its vertices. The decision problem Maximum Edge-Ratio (MaxER) is defined as follows.

Problem 1 Maximum Edge-Ratio (MaxER)
Input: A connected graph G, non-negative rational number K.
Question: Is the edge-ratio of G at least K?

The restriction that G is connected ensures that $e(X, Y) > 0$. In the next section we show that MaxER is NP-complete, by reducing from the well-known NP-complete problem Minimum Cut Into Equal Parts (MinEP). We say that a partition (X, Y) of a set is *equal* if $|X| = |Y|$. Now MinEP is defined as follows.

Problem 2 Minimum Cut Into Equal Parts (MinEP)
Input: A graph G with an even number of vertices, integer K satisfying $0 < K < e(G)$.
Question: Does there exist an equal partition (X, Y) of $V(G)$ with $e(X, Y) \leq K$?

MinEP is shown to be NP-complete in [5]. Our reduction works in two main steps. In the first we show that MinEP remains NP-complete when the input is restricted to regular graphs. (A graph is *d-regular* if every vertex has d neighbours.) We believe that this result is interesting in its own right. It is more convenient to describe the reduction between the corresponding maximization problems. A simple reduction from these to the minimization problems and vice versa consists of taking the complement and adjusting the value of K appropriately. Our reduction between the maximization problems takes an input graph and transforms it into a regular graph by replacing each edge by a pair of parallel edges, inserting a copy of a large complete bipartite graph with one edge deleted into each edge, and attaching many copies of the same graph to each vertex of the original input graph.

In the second step the key observation required is that if G is a d-regular graph with $2n$ vertices, then finding an equal partition (X, Y) of its vertex set minimizing $e(X, Y)$ is equivalent to finding an equal partition (X, Y) maximizing

$r(X, Y)$. By blowing up each vertex of an input graph to a large clique, we can ensure a partition maximizing the edge ratio must be an equal partition and consequently reduce MINEP to MAXER.

2 Results

As we described briefly above, to show that MAXER is NP-complete, we proceed through three auxiliary problems. The most important of these is REGULAR MINIMUM CUT INTO EQUAL PARTS (RMINEP), which is the restriction of MINEP to regular graphs. So formally the definition of RMINEP is as follows.

Problem 3 REGULAR MINIMUM CUT INTO EQUAL PARTS (RMINEP)
Input: A regular graph G with an even number of vertices, integer K satisfying $0 < K < e(G)$.
Question: Does there exist an equal partition (X, Y) of $V(G)$ with $e(X, Y) \leq K$?

It turns out to be convenient to introduce two more auxiliary problems which are formed by replacing minimization by maximization in two of the previous problems. We make the following definitions.

Problem 4 MAXIMUM CUT INTO EQUAL PARTS (MAXEP)
Input: A graph G with an even number of vertices, integer K satisfying $0 < K < e(G)$.
Question: Does there exist an equal partition (X, Y) of $V(G)$ with $e(X, Y) \geq K$?

Problem 5 REGULAR MAXIMUM CUT INTO EQUAL PARTS (RMAXEP)
Input: A regular graph G with an even number of vertices, integer K satisfying $0 < K < e(G)$.
Question: Does there exist an equal partition (X, Y) of $V(G)$ with $e(X, Y) \geq K$?

Note that it is easy to see that each of MAXER, RMINEP, MAXEP and RMAXEP belongs to NP because a suitable certificate is just a bipartition of the vertices of the input graph G, satisfying the required properties, and this can be verified in time $O(n^2)$, where n is the number of vertices of G.

Our first lemma is obvious well-known folklore, but we cannot find a precise reference, so we give the simple proof.

Lemma 1. MAXEP *is NP-complete.*

Proof. We noted above that MAXEP belongs to NP. To prove hardness, it is simple to see that MINEP $\propto$ MAXEP. Suppose that (G, K) is an instance of MINEP and let $n = |V(G)|/2$. Construct the instance $(\overline{G}, n^2 - K)$, where $\overline{G}$ is the complement of G. This can clearly be done in polynomial time. The result follows by observing that if (X, Y) is a partition of $V(G)$, such that $|X| = |Y|$, then $e_G(X, Y) + e_{\overline{G}}(X, Y) = n^2$. $\square$

Lemma 2. RMAXEP *is NP-complete.*

Proof. We noted above that RMAXEP belongs to NP. To establish the hardness of RMAXEP, we show that MAXEP $\propto$ RMAXEP. Let (G, K) be an instance of MAXEP, let $n = |V(G)|$, $N = 2n^2$.

We now describe how to form an N-regular graph G' from G, but first we need a little extra notation. A graph obtained from $K_{N,N}$ by deleting one edge is called a *near-$K_{N,N}$*. We say that the two endpoints of the deleted edge are *depleted*. Now, for each edge $e = uv$ of G, delete uv and take two copies of a near-$K_{N,N}$, join one depleted vertex in each near-$K_{N,N}$ to u and the other to v. This situation is depicted by the right-hand graph in Figure 1. The subgraph of G' spanned by u, v and just one near-$K_{N,N}$, to which both u are v are connected, is called an *edge gadget*. Consequently both the top half and the bottom half of the right-hand graph in Figure 1 are edge gadgets. Once this stage of the construction is complete, each vertex that was originally present in G has twice as many neighbours as before and each new vertex has degree N.

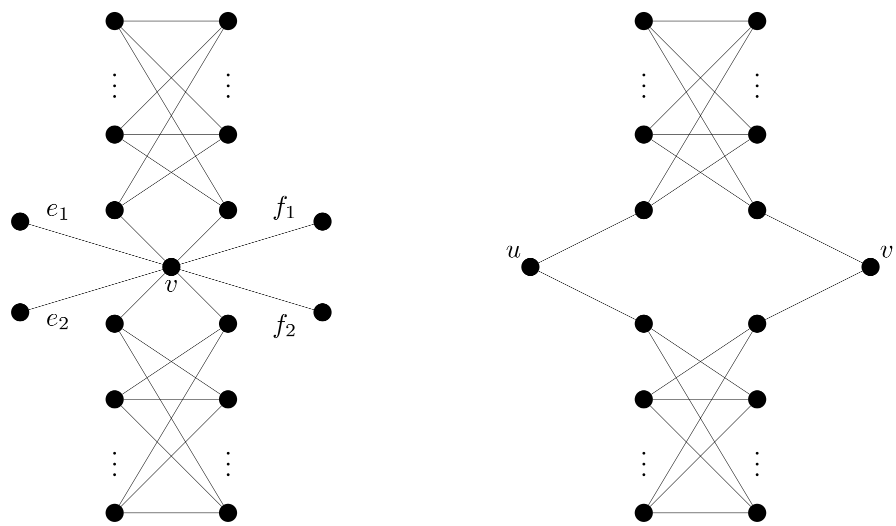

Figure 1: Graphs replacing a vertex and an edge

For each vertex v of the original graph G, let $d(v)$ denote its degree in G. Now for each vertex v of G, add $(N - 2d(v))/2$ copies of a near-$K_{N,N}$ and join all the depleted vertices to v. A subgraph spanned by v and the vertices of just one near-$K_{N,N}$, to which it is joined by two edges, is called a *vertex gadget*. For example suppose that $N = 8$ and v is a vertex of degree 2. (Note that these values of the parameters are in practice incompatible, because we are working with simple graphs.) Then two vertex gadgets would be attached at v and both of the edges leaving v present in the original graphs would be replaced by a pair of edge gadgets. Consequently v would end up having degree $N = 8$. This situation is depicted in the left-hand graph of Figure 1. The edges e_1, e_2, f_1 and f_2 are all part of edge gadgets corresponding to the original edges leaving v. This completes the construction of G', which is a regular graph of degree N. Both an edge gadget and a vertex gadget contain $O(N)$ vertices. G' contains $O(n^2)$ edge gadgets and each vertex is adjacent to $O(N)$ vertex gadgets. Therefore there are in total $O(n^5)$ vertices and consequently $O(n^7)$ edges, so G' can be constructed in polynomial time.

We claim that G has an equal partition (X, Y) with $e(X, Y) \geq K$, if and only if, G' has an equal partition (X', Y') with $e(X', Y') \geq K'$ where

$$K' = \Big(2e(G) + \sum_{v \in V(G)} \Big(\frac{N}{2} - d(v)\Big)\Big)N^2 + 2K.$$

Establishing this claim is enough to prove the lemma.

Let H be an edge gadget and let u, v denote the endpoints of the edge in G to which it corresponds. Let (X, Y) be a partition of $V(H)$. If there is some edge of the near-$K_{N,N}$ forming part of H having both endpoints on the same side of the partition then $e(X, Y) \leq N^2 - N + 2$.

On the other hand, if we ensure that each edge of the near-$K_{N,N}$ forming part of H has both endpoints on opposite sides of the partition, then we can arrange X and Y to achieve $e(X, Y) = N^2 + 1$ if u and v lie in different sets of the partition and $e(X, Y) = N^2$ if u and v lie in the same set of the partition. Both of these are the maximum values possible.

Now let H denote a vertex gadget and (X, Y) denote a partition of $V(H)$. If there is some edge of the near-$K_{N,N}$ forming part of H having both endpoints on the same side of the partition then $e(X, Y) \leq N^2 - N + 2$. On the other hand, if we ensure that each edge of the near-$K_{N,N}$ forming part of H has both endpoints on opposite sides of the partition, then we can arrange X and Y to achieve $e(X, Y) = N^2$, which is best possible.

The remarks above imply that if (X', Y') is a partition of $V(G')$ such that there is a near-$K_{N,N}$ with an edge having both endpoints on the same side of the partition, then

$$\begin{aligned} e(X', Y') &\leq \overbrace{2e(G)(N^2+1)}^{\text{edge gadgets}} + \overbrace{\sum_{v \in V(G)} \Big(\frac{N}{2} - d(v)\Big)N^2}^{\text{vertex gadgets}} - \overbrace{(N-2)}^{\text{missing edges}} \qquad (1) \\ &= \Big(2e(G) + \sum_{v \in V(G)} \Big(\frac{N}{2} - d(v)\Big)\Big)N^2 + 2e(G) - (N-2) < K'. \end{aligned}$$

The first term in (1) is an upper bound (over all possible partitions (X', Y') ignoring the requirement for at least one near-$K_{N,N}$ with an edge having both endpoints on the same side of the partition) for the number of edges in edge gadgets that may contribute to $e(X', Y')$. The second term is an upper bound for the number of edges in vertex gadgets that may contribute to $e(X', Y')$. The final term counts the minimum number of "missing edges", i.e. the smallest possible overcount in the first term due to neglecting the requirement for at least one near-$K_{N,N}$ with an edge having both endpoints on the same side of the partition. So if (X', Y') is an equal partition of $V(G')$ satisfying $e(X', Y') \geq K'$ then $|X' \cap (V(G') \setminus V(G))| = |Y' \cap (V(G') \setminus V(G))|$ and so $|X' \cap V(G)| = |Y' \cap V(G)|$.

If there is an equal partition (X, Y) of $V(G)$ with $e(X, Y) = C$ then we can add each vertex of $V(G') \setminus V(G)$ to either X or Y to get an equal partition (X', Y') of $V(G')$ with

$$e(X', Y') = \overbrace{2e(G)N^2 + 2C}^{\text{edge gadgets}} + \overbrace{\sum_{v \in V(G)} \Big(\frac{N}{2} - d(v)\Big)N^2}^{\text{vertex gadgets}}.$$

(All edges lie in either a vertex gadget or an edge gadget. The first terms count edges belonging to edge gadgets and includes edges of the original graph. The final term counts edges belonging to vertex gadgets.) Moreover this is the largest possible value of $e(X', Y')$ such that (X', Y') is an equal partition of $V(G')$ with $X' \cap V(G) = X$ and $Y' \cap V(G) = Y$. Consequently G has an equal partition (X, Y) with $e(X, Y) \geq K$, if and only if, G' has an equal partition (X', Y') with $e(X', Y') \geq K'$. □

Lemma 3. RMinEP *is NP-complete.*

Proof. Since the complement of a regular graph is again regular, the proof is almost identical to that of Lemma 1. □

We can now establish our main theorem, which is that MaxER is NP-complete. Recall the key observation required to prove the theorem. Suppose that G is a d-regular graph with $2n$ vertices. If (X, Y) is an equal partition of $V(G)$ with $e(X, Y) = K$, then $e(X) = e(Y) = (dn - K)/2$. Hence $r(X, Y) = \dfrac{dn - K}{2K}$. So minimizing $e(X, Y)$ is equivalent to maximizing $r(X, Y)$.

Theorem 4. MaxER *is NP-complete.*

Proof. We have already noted that MaxER is in NP, so it suffices to prove hardness. We show that RMinEP $\propto$ MaxER.

We first deal with an easy degenerate case when the value K in an instance of RMinEP is so large that the answer is guaranteed to be yes. Let P be the graph consisting of just two vertices joined by an edge and let $n = |V(G)|/2$. Every graph has an equal partition (X, Y) of its vertices with $e(X, Y) \leq n^2$. Consequently if $K > n^2$ then (G, K) is a 'yes' instance of RMinEP, so let $(G', K') = (P, 0)$.

From now on we assume that $1 \leq K \leq \min\{e(G), n^2\}$ and continue with the main part of the proof.

Suppose that G is d-regular and that $V(G) = \{1, \ldots, 2n\}$. Let $M = 4(n + 2)^2 n$ and $N = n + 1$. The graph G' has vertex set

$$V(G') = \{v_{i,j} : 1 \leq i \leq 2n \text{ and } 0 \leq j \leq M\},$$

and edge set

$$\begin{aligned}E(G') = \{v_{i,j}v_{i,k} : 1 \le i \le 2n \text{ and } 0 \le j < k \le M\} \\ \cup \{v_{i,k}v_{j,k} : 1 \le i < j \le 2n \text{ and } 1 \le k \le N\} \cup \{v_{i,0}v_{j,0} : ij \in E(G)\}.\end{aligned}$$

So G' is formed by replacing each vertex of G by a clique of size $M+1$, adding N edges between each pair of cliques and finally adding a further edge between each pair of cliques corresponding to vertices which were joined in G. Therefore $|V(G')| = 2n(M+1)$ and

$$e(G') = \overbrace{2n\binom{M+1}{2}}^{\text{within cliques}} + \overbrace{N\binom{2n}{2}}^{\text{between cliques}} + \overbrace{e(G)}^{\text{original edges}}.$$

(The first term in $e(G')$ comes from the edges within cliques, the second term from the N edges added between every pair of cliques and the final term from the edges corresponding to edges of the original graph.) Let

$$K' = \frac{n\binom{M+1}{2} + N\binom{n}{2} + \frac{e(G)-K}{2}}{Nn^2 + K} \ge \frac{n\binom{M+1}{2} + N\binom{n}{2}}{(N+1)n^2}.$$

Clearly (G', K') can be constructed in polynomial time.

Suppose there is an equal partition (X, Y) of $V(G)$ with $e(X, Y) = C$. Let $X' = \{v_{i,j} : i \in X \text{ and } 0 \le j \le M\}$. So X' consists of all the vertices in the cliques of G' corresponding to vertices in X. Now

$$\begin{aligned}e(Y') = e(X') &= n\binom{M+1}{2} + N\binom{n}{2} + \frac{dn - C}{2} \\ &= n\binom{M+1}{2} + N\binom{n}{2} + \frac{e(G) - C}{2}.\end{aligned}$$

Furthermore $e(X', Y') = Nn^2 + C$. Hence

$$r(X', Y') = \frac{n\binom{M+1}{2} + N\binom{n}{2} + \frac{e(G)-C}{2}}{Nn^2 + C}. \tag{2}$$

So, in particular, if there is an equal partition (X, Y) of $V(G)$ with $e(X, Y) = C \ge K$ then $r(X', Y') \ge K'$ and G' has edge-ratio at least K'.

Now suppose that G' has edge-ratio at least K'. Then there exists a partition (X', Y') of $V(G')$ such that $r(X', Y') \ge K'$. First, suppose that for some i the set $U = \{v_{i,j} : 0 \le j \le M\}$ has vertices on both sides of the partition (X', Y'). We have

$$\min\{e(X'), e(Y')\} \le \frac{e(G')}{2} = \frac{1}{2}\Big(2n\binom{M+1}{2} + N\binom{2n}{2} + e(G)\Big).$$

Furthermore because U has vertices on both sides of the partition, we have $e(X', Y') \ge M$. Hence

$$r(X', Y') \le \frac{2n\binom{M+1}{2} + N\binom{2n}{2} + e(G)}{2M} < \frac{n\binom{M+1}{2} + N\binom{n}{2}}{(N+1)n^2} \le K'.$$

So for each i, the vertices in $\{v_{i,j} : 0 \leq j \leq M\}$ lie on the same side of the partition (X', Y').

Next suppose that $X' = \{v_{i,j} : i \in X, \text{ and } 0 \leq j \leq M\}$ where $X \subseteq V(G)$ and $|X| \neq n$. We can assume without loss of generality that $|X| = m < n$. We have

$$e(X') \leq m\binom{M+1}{2} + N\binom{m}{2} + e(G) \leq e(Y')$$

and

$$e(X', Y') \geq Nm(2n - m).$$

So

$$r(X', Y') \leq \frac{m\binom{M+1}{2} + N\binom{m}{2} + e(G)}{Nm(2n-m)} < \frac{n\binom{M+1}{2} + N\binom{n}{2}}{(N+1)n^2} \leq K'.$$

So the only partitions (X', Y') of $V(G')$ with $r(X', Y') \geq K'$ are such that $X' = \{v_{i,j} : i \in X \text{ and } 0 \leq j \leq M\}$ for some $X \subseteq V(G)$ and moreover $|X| = |V(G) \setminus X|$. Therefore if there is a partition (X', Y') of $V(G')$ with $r(X', Y') \geq K'$ then there is an equal partition (X, Y) of $V(G)$ with $e(X, Y) \geq K$. $\square$

Acknowledgement

We thank the anonymous referee for a careful reading and for several helpful comments which have improved the presentation of the paper.

References

[1] U. Brandes, D. Delling, M. Gaertler, R. Görke, M. Hoefer, Z. Nikoloski, D. Wagner, On modularity clustering, IEEE Transactions on Knowledge and Data Engineering 20 (2008) 172–188.

[2] S. Cafieri, P. Hansen, L. Liberti, Edge ratio and community structure in networks, Physical Review E 81 (2010) 026105.

[3] R. Diestel, Graph Theory, 3rd ed., Springer, Berlin, 2005.

[4] S. Fortunato, Community detection in graphs, Physics Reports 486 (2010) 75–174.

[5] M. R. Garey, D. S. Johnson, L. Stockmeyer, Some simplified NP-complete graph problems, Theoret. Comput. Sci. 1 (1976) 237–267.

[6] R. Kannan, S. Vempala, A. Vetta, On clustering: good, bad and spectral, J. ACM 51 (2004) 497–515.

[7] M. E. J. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2004) 0216113.