Journal of Combinatorial Optimization manuscript No.
(will be inserted by the editor)

Polarization reduction by minimum-cardinality balanced edge additions:

Formulations, complexity, and integer programming approaches

Ruben Interian · Jorge R. Moreno · Celso C. Ribeiro

Received: date / Accepted: date

Abstract Real-world networks are often extremely polarized, because the communication between different groups of vertices can be weak and, most of the time, only vertices within the same groups or sharing the same beliefs communicate to each other. In this work, we introduce the Minimum-Cardinality Balanced Edge Addition Problem (MinCBEAP) as a strategy for reducing polarization in real-world networks based on a principle of minimum external interventions. We present the problem formulation and discuss its complexity, showing that its decision version is NP-complete. We also propose three integer linear programming formulations for the problem and discuss computational results on artificially generated and real-life instances. Randomly generated instances with up to 1000 vertices are solved to optimality. On the real-life instances, we show that polarization can be reduced to the desired threshold with the addition of a few edges. The minimum intervention principle and the methods developed in this work are shown to constitute an effective strategy for tackling polarization issues in practice in social, interaction, and commu-

Work of Ruben Interian and Jorge R. Moreno was supported by CNPq scholarships. Work of Celso C. Ribeiro was partially supported by CNPq research grant 303958/2015-4 and by FAPERJ research grant E-26/201.198/2014, and concluded during a visit to the State Key Laboratory of Scientific and Engineering Computing of the Chinese Academy of Sciences, China. This work was also partially sponsored by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Finance Code 001.

R. Interian
Universidade Federal Fluminense, Institute of Computing, Niterói, RJ 24271-270, Brasil.
E-mail: rinterian@ic.uff.br

J. R. Moreno
Universidade Federal Fluminense, Institute of Computing, Niterói, RJ 24271-270, Brasil.
E-mail: jmoreno@ic.uff.br

C. C. Ribeiro
Universidade Federal Fluminense, Institute of Computing, Niterói, RJ 24271-270, Brasil.
E-mail: celso@ic.uff.br

nication networks, which is a relevant problem in a world characterized by extreme political and ideological polarization.

Keywords Polarization · Minimum-cardinality balanced edge addition problem · Polarized networks · Complexity · Integer programming

1 Motivation

The issue of polarization has been discussed by politicians, media, and researchers [8, 25]. This subject has also attracted the attention of thinkers throughout history. John Stuart Mill, an important philosopher and political theorist, claimed that dialogue across lines of political difference is a key prerequisite for sustaining a democratic citizenry [16]. Hannah Arendt also asseverated that debate is irreplaceable for forming enlightened opinions that reach beyond the limits of one's own subjectivity to incorporate the standpoints of others [4]. From sociologists to economists, many are interested in studying the behavior and interactions in social networks that rule the opinion formation process.

According to the Oxford Dictionaries, polarization is the division into sharply contrasting groups or sets of opinions or beliefs [7]. Academic articles, newspapers, and the media in general constantly report the growth of fake news, misinformation spreading, and polarization of increasingly isolated groups of individuals. These phenomena are closely interrelated with each other. Fake news spread faster in polarized networks or groups [24]. At the same time, fake and tendentious news can accentuate polarization within already existing echo chambers in the social networks.

Recently, the causes of the proliferation of flat-earth believers, i.e., people who believe that the Earth is actually flat, were investigated by Landrum [14], revealing the role of the video-sharing platform YouTube on this proliferation. This work showed that the algorithms the platform uses to guide people to topics that might interest them makes it easier for a user to end up in a misinformation echo chamber. The study concludes that the most effective instrument to combat disinformation – i.e., false information spread deliberately to deceive – is to provide (or even "to flood") users of the platform with quality information, to ensure that the public also receives accurate, scientific or simply plural information when watching videos on some subject.

Interian and Ribeiro [13] have shown that many case-study real-world networks are extremely polarized. A polarized network is divided into two or more strongly connected groups, with few edges between vertices belonging to different groups. Communication between different groups is weak: there are many vertices for which all or most of its neighbors belong to the same group. In practice, this corresponds to a situation where, most of the time, only same-group vertices communicate to each other and most of the information that a vertex can receive comes from inside the same group to which it belongs. These groups may correspond to large cliques or quasi-cliques [1, 2, 22, 23, 26, 27].

Interian and Ribeiro also showed in [13] that in order to reduce polarization, networks can be treated by external interventions. An intervention can be seen as any externally-induced process that modifies the structure of the network, such as a fact-checking campaign, a marketing campaign, a regulatory action or some direct manipulation that adds or removes vertices or edges of the network. The process of adding new vertices is often difficult to be performed in real networks. On the other hand, removing vertices or edges may be controversial, because it can be interpreted as the permanent exclusion or deletion of elements such as users, sites, or posts from a social network. This kind of intervention has been widely used in moderation systems for inspecting or removing objectionable content at the discretion of the moderator and such exclusions are often seen as aggressions against freedom of expression in the digital environment.

Suppose that we have a network formed by a set of vertices V partitioned into disjoint subsets $V_1, V_2, \ldots, V_k$. Two vertices that belongs to the same subset V_i are called same-type vertices, while two vertices that belong to different subsets V_i and V_j, $i \neq j$, are called different-type vertices. We consider the addition of edges between different-type vertices of the network as a less invasive treatment method. A typical example of the use of this kind of treatment in real networks is the suggestion of new friendship relations in social networks. By adding edges between vertices of different groups, a super-graph containing the original graph is built. There are more connections between different-type vertices inside this super-graph and, consequently, inter-group communication is improved.

A new optimization problem addressing the issue of polarization reduction by edge additions is presented in this work. Other optimization problems have already used the idea of adding edges to a graph with the goal of improving specific performance measures. Constant-factor approximation algorithms were developed in [6] for the problem of adding k shortcut edges to the graph in order to minimize its diameter. A game that models the creation of a network by selfish agents that benefit from shortest paths to all destinations is analyzed in [9], considering that the agents pay for the links they establish. Two variants of the diameter minimization problem are studied in [15]: the minimum-cardinality-bounded-diameter and the Bounded-cardinality-minimum-diameter edge addition problems, where it is shown that both problems are NP-hard even if the value of the diameter is fixed to 2. Improved approximation ratios of $O(\log n)$ and 2 were proposed in [5] for both problems, respectively. Some results were also extended to the edge-weighted versions of the problems.

Other works in the area of analysis of social networks explored the idea of adding edges to a graph in order to improve its ability to disseminate information. A problem addressing the minimization of the average shortest path distance between all pairs of vertices was studied in [20], adding a limited number of additional "ghost edges" with the objective of improving the network efficiency of information propagation. This approach prioritizes the shortest path distance between each pair of vertices, while in the present work the connectivity between groups of vertices that represent different opinions, ideas, or beliefs is analyzed.

A measure called characteristic path length was minimized in [19]. The characteristic path length is another name for the average shortest path distance between all pairs of vertices. Some properties of the problem are proved and methods for computing the utility of all candidate edges in large graphs are described and evaluated.

Another edge recommendation problem was introduced in [11]. In this case, the goal of the recommendation is to reduce the "controversy score" of the graph, using a metric based on random walks. The controversy score relies on how controversial a topic is or, in other words, on how much polarization it generates. The probability of acceptance of the recommended edge is also evaluated.

In this article, we propose the minimal intervention principle, which consists in assuming that the lowest number of changes should be made in the original network in order to attend any proposed condition for polarization reduction. We formulate the minimum-cardinality balanced edge addition problem and discuss integer programming formulations for its solution. This work is organized as follows. In the next section, we present the problem formulation and its complexity. Integer programming models are presented in Section 3. Computational results on randomly generated and real-life instances are discussed in Section 4. Concluding remarks are drawn in the last Section 5.

2 Problem formulation and complexity

Let $G = (V, E)$ be a graph defined by a set $V = \{v_1, \dots, v_n\}$ of vertices and a set $E \subseteq V \times V$ of edges. We use the term *group* to refer to any subset of the vertex set V.

We assume that graph G is polarized to some extent and it is necessary to solve the issue of insufficient communication between the groups. The reduction of the polarization in a polarized graph can be treated and formulated as a mathematical optimization problem is discussed next.

2.1 Minimum-cardinality balanced edge addition problem

Interian and Ribeiro [13] observed that, in many graphs, there may be an important number of vertices that are not connected to other groups, i.e., there may be only intra-group edges adjacent to these vertices. Consider, for example, a network of books about U.S. politics sold by Amazon.com [18]. Edges between books represent frequent copurchasing of those books by the same buyers. Most of the books are classified as conservative or liberal, and a small number of them as neutral. There are 105 vertices in this instance and 56 of them are adjacent only to neighbors of the same group, as shown in Figure 1. Another example is that of a network of political blogs that emerged during the 2004 U.S. presidential election [3]. Blogs are divided into two groups:

republican and democratic. Among the 1065 non-isolated vertices in this instance, there are 572 blogs with links exclusively to blogs of the same political orientation, as shown in Figure 2.

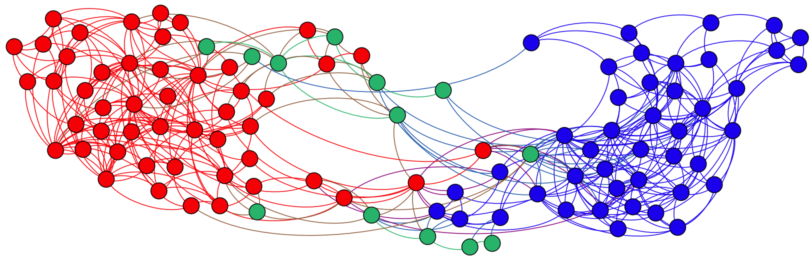


Fig. 1: Network of books about U.S. politics sold by Amazon.com. Red, green, and blue vertices represent, respectively, conservative, neutral and liberal books.

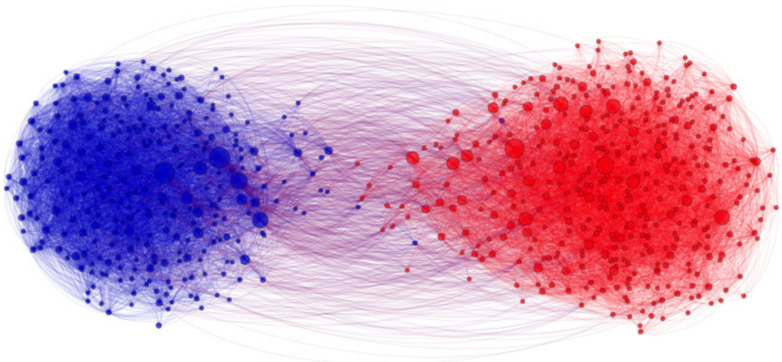


Fig. 2: Network of political blogs during the 2004 U.S. presidential election. Red and blue vertices represent republican and democratic blogs, respectively.

In practice, it can be unrealistic to add an expressive number of edges for each vertex, since this kind of intervention should be minimal. We refer to this assumption as the minimal intervention principle. These statement led us to consider the following optimization problem, in which we seek to minimize the number of edges to be added to a polarized graph in order that any vertex in a proper vertex subset $A \subset V$ can reach some vertex of $V \setminus A$ in the resulting graph by a path with a limited number of edges. If we denote by $d_G(v, V')$ the minimum number of edges in a path from a vertex v of graph G to any vertex in $V' \subseteq V$, then this problem can be formulated as:

Minimum-cardinality Balanced Edge Addition Problem (MinCBEAP)
Instance: Graph $G = (V, E)$, subset $A \subset V$, integer D.
Goal: Find a minimum-cardinality set $E' \subseteq (V \times V) \setminus E$ such that $d_{G'=(V,E \cup E')}(v, V \setminus A) \leq D,\ \forall v \in A$.

Given an integer L as an additional parameter, the decision version of MinCBEAP amounts to the question: "Is there a set $E' \subseteq (V \times V) \setminus E$ with at most L edges such that $d_{G'=(V,E \cup E')}(v, V \setminus A) \leq D,\ \forall v \in A$?"

To prove that MinCBEAP is NP-complete, we first define the eccentricity $\epsilon(v)$ of a vertex $v \in V$ as the longest of the shortest paths in G from v to all other vertices in V [12].

Bearing this definition in mind, we introduce the Minimum-cardinality-bounded-eccentricity edge addition problem [6] (MCBE), which consists in reducing the eccentricity of some vertex v by adding edges to the graph the vertex belongs. More formally, its decision version can be stated as:

Minimum-cardinality-bounded-eccentricity edge addition problem (MCBE)
Instance: Graph $G = (V, E)$, source vertex $s \in V$, integer p, integer B.
Question: Is there a supergraph $G' = (V, E \cup E')$ of G with $E' \subseteq (V \times V) \setminus E$ such that $|E'| \leq p$ and $\epsilon_{G'}(s) \leq B$?

Lemma 1 *There is a concise certificate for MCBE with all edges incident to vertex s.*

Proof Let E' be any concise certificate for MCBE. Consider the shortest-path tree T in graph $G' = (V, E \cup E')$ rooted at s. Each edge in the tree is traversed in the direction of the shortest path to s. Any edge (u, v) in $E' \cap T$ used in this direction can be replaced by edge (u, s), since all vertices that use edge (u, v) in their shortest paths to s will not have their their distance to s increased, therefore creating a new concise certificate with all edges incident to the source vertex s.

Although the NP-completeness of MCBE has been suggested by some authors [6][21], to the best of our knowledge a formal proof does not exist. We give a proof using a polynomial reduction from the set covering problem [10]:

Set covering problem (SC)
Instance: Collection $C = \{S_1, \ldots, S_m\}$ of subsets of a finite set $S = \{x_1, \ldots, x_n\}$, integer k.

Question: Is there a cover $C' \subseteq C$ such that each element of S belongs to at least one member of C' and $|C'| \leq k$?

Theorem 1 *MCBE is NP-complete.*

Proof MCBE is in NP, since for any of its instances defined by a graph $G = (V, E)$, a source vertex $s \in V$, and integers p and B, the eccentricity of the source vertex s in a supergraph $G' = (V, E \cup E')$ of G can be calculated in polynomial time, where $E' \subseteq (V \times V) \setminus E$.

We show that any instance of set covering problem can be transformed into an instance of MCBE with $B = 2$. Consider an instance of the set covering problem defined by subsets $S_1, \ldots, S_m$, with $|S_1 \cup \ldots \cup S_m| = n$, and by an integer k that indicates the size of the target cover C'. Build an instance of MCBE as follows. Let G be a graph with vertex set $V = \{u_1, \ldots, u_m, v_1, \ldots, v_n, s, s'\}$. There is an edge between vertices u_j and v_i if element x_i belongs to S_j. Vertices s and s' are connected and vertex s' is connected with vertices $u_1, \ldots, u_m$. In addition, set $B = 2$ and $p = k$.

Figure 3 illustrates an example of the construction of an instance of MCBE with $B = 2$ and $p = 3$. Note that $\epsilon(s) = 3$ and let $E' \subseteq (V \times V) \setminus E$ be a set with at most p edges such that $\epsilon(s) \leq 2$ in $G' = (V, E \cup E')$, i.e., E' is a concise certificate for MCBE for this instance.

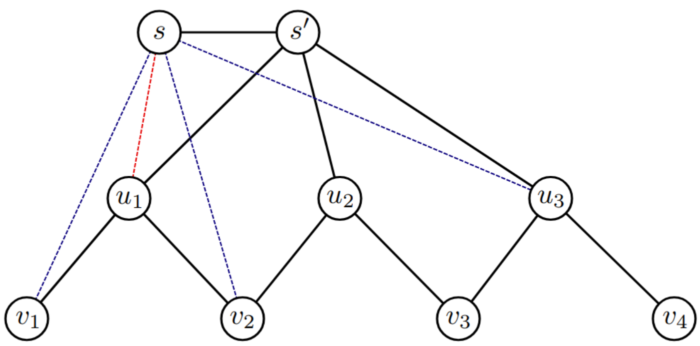


Fig. 3: Example of instance used in the proof of the NP-completeness of MCBE. The concise certificate $E' = \{(s, v_1), (s, v_2), (s, u_3)\}$, highlighted in blue, is replaced by the certificate $\overline{E} = \{(s, u_1), (s, u_3)\}$, with edges (s, v_1) and (s, v_2) replaced by edge (s, u_1).

The distance from vertex s to any vertex $v_1, \ldots, v_n$ in G is greater than 2. From Lemma 1, without loss of generality, we may pick the certificate E' in such a way that all its edges are incident to s. The other extremities of the edges in E' necessarily belong to either $\{v_1, \ldots, v_n\}$ or $\{u_1, \ldots, u_m\}$.

To build another set $\overline{E}$ with at most p edges such that all of them are incident to $\{u_1, \ldots, u_m\}$, we replace every edge $(s, v_i), i = 1, \ldots, n$, in E' by an edge (s, u_j) in $\overline{E}$, with $j : x_i \in S_j$. $\overline{E}$ remains a concise certificate for MCBE, because the distance from s to vertex v_i in $\overline{G} = (V, E \cup \overline{E})$ is still less than 3 for any $i = 1, \ldots, n$ for which there is an edge $(s, v_i) \in E'$. Therefore, $\epsilon(s)$ in $\overline{G}$ is also at most 2.

To conclude, we note that for each vertex v_i there is a vertex u_j such that there is an edge in $\overline{E}$ from s to u_j, because v_i is at most at distance 2 from s in $\overline{G}$. In consequence, the edges in $\overline{E}$ are incident to at most k vertices, each one associated with a set S_j. These k sets represent a concise certificate for the set covering instance.

In order to prove the NP-completeness of MinCBEAP, a polynomial transformation from MCBE is used:

Theorem 2 *MinCBEAP is NP-complete.*

Proof The problem is in NP, since the distance from any vertex $v \in A$ to any vertex in $V \setminus A$ can be calculated in polynomial time.

Now, consider an instance of MCBE defined by graph G, vertex s and integers p and B, and build an instance of MinCBEAP by setting $A = V \setminus \{s\}$ as the proper vertex subset of V. Then, $V \setminus A = \{s\}$. Set $D = B$ and $L = p$.

Let $E' \subseteq (V \times V) \setminus E$ be a set with at most L edges such that all vertices in A are at a distance of at most D from s in $G' = (V, E \cup E')$, i.e., E' is a concise certificate to MinCBEAP. Then, adding E' to G reduces the eccentricity of s to at most $B = D$ using at most $p = L$ edges, since the graph $G' = (V, E \cup E')$ is undirected. Consequently, E' is also a concise certificate to MCBE.

Exact integer programming approaches for the Minimum-cardinality balanced edge addition problem are developed in the next section.

3 Integer programming formulations

Given a non-oriented graph $G = (V, E)$, a vertex subset $A \subset V$, and a non-negative integer D, the optimization version of MinCBEAP amounts to finding a minimum cardinality set $S_G \subseteq (V \times V) \setminus E$ such that $d_{G'=(V, E \cup S_G)}(v, V \setminus A) \leq D, \ \forall v \in A$.

3.1 Instance transformation

There are no edges in an optimal solution S_G to MinCBEAP with both extremities in $V \setminus A$, because adding edges with both extremities in $V \setminus A$ would not affect the distance from any vertex in A to those in $V \setminus A$.

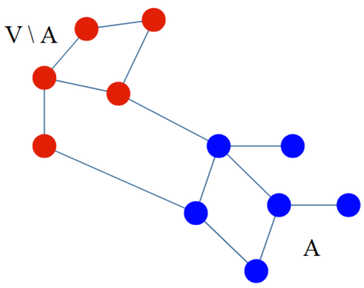
The following proposition holds:

Proposition 1 *Let S_G be a solution to MinCBEAP. Let $(u, v) \in S_G$ be an edge with $u \in A$ and $v \in V \setminus A$. Then, $(S_G \setminus \{(u, v)\}) \cup \{(u, w)\}$, with $w \in V \setminus A$ and $w \neq v$, is also a solution to MinCBEAP.*

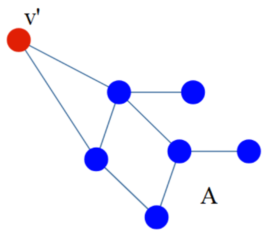
Proof Replacing edge (u, v) by edge (u, w) does not change the distance from any vertex in A to set $V \setminus A$.

Given a non-oriented graph $G = (V, E)$, a source vertex s, and a non-negative integer B, the optimization version of MCBE amounts to finding a minimum cardinality set $S_H \subseteq (V \times V) \setminus E$ such that $\epsilon_{G'=(V,E\cup S_H)}(s) \leq B$.

Then, consider the following transformation from an instance of MinCBEAP defined on graph $G = (V, E)$, as illustrated in Figure 4a, that creates an instance of MCBE on graph $H = (V_H, E_H)$, as illustrated in Figure 4b. In the transformed MCBE instance, $V_H = A \cup \{v'\}$, $s = v'$, and $B = D$, with the dummy vertex v' representing the collapsed set $V \setminus A$. Furthermore, for any vertex $u \in A$ such that there is an edge between u and some vertex $v \in V \setminus A$ in G, then there is an edge between u and v' in H. We also observe that while the number of vertices in $G = (V, E)$ is $|V|$, there are only $|A| + 1$ vertices in the graph $H = (V_H, E_H)$ that defines the MCBE instance.



(a) Instance of MinCBEAP on $G = (V, E)$



(b) Transformed instance of MCBE on $H = (V_H, E_H)$

Fig. 4: Instance transformation.

We make use of this transformation to find a solution for the transformed instance of MCBE, which is then used to obtain a solution for the original instance of MinCBEAP. Let S_H be an optimal solution for the transformed MCBE instance. A solution S_G for the original instance of MinCBEAP can be obtained as follows. Let $e = (u, v) \in S_H$. If both $u, v \in A$, then edge $e = (u, v)$ also belongs to S_G. In case one of the extremities – say, extremity v – of edge e coincides with $v' \notin A$, then we chose at random a vertex $w \in V \setminus A$, and substitute edge $e = (u, v')$ in S_H by edge $e' = (u, w)$ in S_G. Therefore, by construction, the solution S_G obtained for MinCBEAP has $|S_G| = |S_H|$.

3.2 First ILP formulation

Any optimal solution S_H to problem MCBE can be seen as an oriented spanning tree of the graph $H' = (V_H, E_H \cup S_H)$ rooted at vertex v'. The distance from any vertex in the tree to vertex v' should be at most D. The arcs of the oriented spanning tree indicate the paths from each vertex to the root v'.

This formulation makes use of a variant of the Miller-Tucker-Zemlin constraints to avoid cycles [17]. They create an arborescence in which each vertex v is labeled with an integer d_v. The root is labeled with $d_{v'} = 0$ and the vertices in any tree arc (v_1, v_2) are labeled with $d_{v_1} > d_{v_2}$.

The edges in the optimal solution are those associated with arcs that belong to the oriented spanning tree and not to E_H.

For each vertex $u \neq v$, we define the following decision variable:

$$x_{uv} = \begin{cases} 1, & \text{if arc } (u,v) \in A \times (A \cup \{v'\}), \text{ belongs to the oriented spanning tree,} \\ 0, & \text{otherwise.} \end{cases}$$

The integer variable d_v indicates the label of vertex $v \in V_H$. The formulation makes use of weights defined as $w_{uv} = 0$ if the associated edge $(u, v) \in E_H$, $w_{uv} = 1$ otherwise:

$$\min \sum_{u \in A} \sum_{v \in A \cup \{v'\}} w_{uv} x_{uv} \tag{1}$$

subject to:

$$\sum_{v \in A \cup \{v'\}, v \neq u} x_{uv} = 1, \ \forall u \in A \tag{2}$$

$$x_{uv} + x_{vu} \leq 1, \ \forall u, v \in A, \tag{3}$$

$$d_u \geq x_{uv} + d_v - (1 - x_{uv})D, \ \forall u \in A, \forall v \in A \cup \{v'\}, u \neq v \tag{4}$$

$$d_u \leq D, \ \forall u \in A \tag{5}$$

$$d_u \geq 1, \ \forall u \in A \tag{6}$$

$$d_{v'} = 0, \tag{7}$$

$$d_u = 1, \ \forall u \in A, \ v' \in N_H(u) \tag{8}$$

$$x_{uv'} = 1, \ \forall u \in A, \ v' \in N_H(u) \tag{9}$$

$$x_{uv} \in \{0, 1\}, \ \forall (u, v) \in A \times (A \cup \{v'\}) \tag{10}$$

$$d_v \in \{0, ..., D\}, \ \forall u \in A \cup \{v'\}, \tag{11}$$

with $N_H(u) = \{v \in A \cup \{v'\} : (u, v) \in E_H\}$.

The objective function (1) minimizes the number of edges, since the weights of edges in E_H are zero. Constraints (2) indicate that an arc must come out from every vertex of A, tracing the path (i.e., the last vertex before) to vertex v'. Constraints (3) enforce that there is at most one arc between any pair of vertices. Constraints (4) ensure that if $x_{uv} = 1$, i.e., arc (u, v) belongs to the oriented spanning tree, then $d_u > d_v$. On the other hand, if $x_{uv} = 0$, i.e., arc (u, v) does not belong to the oriented spanning tree, then the constraint becomes $d_u \geq d_v - D$ and is satisfied for any $d_u, d_v \in \{0, \dots, D\}$. Constraints (5) and (6) indicate, respectively, upper and lower bounds to the vertex labels. Constraint (7) sets the label of vertex v' to zero. Constraints (8) set to one the labels of the vertices of A that are adjacent to v', while constraints (9) set to one the variables associated with the vertices of A that are adjacent to v'. Constraints (10) and (11) are the integrality requirements.

We observe that although the model can be solved without constraints (8) and (9), they are added to accelerate the solution process.

3.3 Second ILP formulation

We recall that our formulation addresses the transformed instance of MCBE on graph $H = (V_H, E_H)$, with v' being the dummy vertex. From Lemma 1, we know that there is always a solution S_H with all edges having v' as one of the extremities. Therefore, the problem can be solved by considering only this particular subset of solutions and deciding, for each vertex u, if edge (u, v') should be added to the graph.

Demaine and Zadimoghaddam [6] proposed a model solving the linear feasibility problem associated to the MCBE. The adaptation of this model to an optimization problem is described next. The following decision variabes are defined:

$$y_u = \begin{cases} 1, & \text{if there is an edge between vertex } u \in V_H \text{ and } v', \\ 0, & \text{otherwise;} \end{cases}$$

$$t_{uv} = \begin{cases} 1, & \text{if the shortest path from vertex } v \in V_H \text{ to } v' \text{ makes use of edge } (u, v'), \\ 0, & \text{otherwise.} \end{cases}$$

If the distance between v and v' is greater than D, then a path from v will reach v' using any of the vertices that are at a distance from v' that is smaller than D. Moreover, as noted in [6], vertex v can not use edge (u, v') if the distance between v and u is greater than D:

$$\min \sum_{u \in V_H : (u, v') \notin E_H} y_u \tag{12}$$

subject to:

$$t_{uv} \leq y_u, \ \ \forall u, v \in A \tag{13}$$

$$\sum_{u : dist(u,v) < D} t_{uv} = 1, \ \ v \in A, dist(v, v') > D \tag{14}$$

$$y_u \in \{0, 1\}, \ \ \forall u \in A \cup \{v'\} \tag{15}$$

$$t_{uv} \in \{0, 1\}, \ \ \forall u, v \in A \cup \{v'\} \tag{16}$$

The objective function (12) minimizes the number of edges adjacent to vertex v' to be added. Constraints (13) indicate that if vertex v_j reaches v' using edge (v_i, v'), then vertex v_i must be counted in the objective function.

Moreover, constraint (14) expresses that if the distance between v_j and v' is greater than D, then vertex j reaches v' using exactly one of the vertices that are at a distance to v' that is smaller than D. Constraints (15) and (16) are the integrality requirements.

3.4 Third ILP formulation

In this formulation based only on 0-1 variables, we also make use of Lemma 1 that establishes that there is always a solution to MCBE with all edges incident to the source vertex v'. In addition to the variables

$$y_u = \begin{cases} 1, & \text{if there is an edge between vertex } u \in V_H \text{ and } v', \\ 0, & \text{otherwise;} \end{cases}$$

already used in the previous formulation, we also define

$$d_{vk} = \begin{cases} 1, & \text{if there is an path of size } k \text{ from vertex } v \text{ to vertex } v', \\ 0, & \text{otherwise.} \end{cases}$$

The problem may then be formulated as:

$$\min \sum_{(u,v') \notin E_H} y_u \tag{17}$$

subject to:

$$\sum_{k=1}^{D} d_{uk} = 1, \ \forall u \in A \tag{18}$$

$$d_{u0} = 0, \ \forall u \in A \tag{19}$$

$$\sum_{k=1}^{D} d_{v'k} = 0, \tag{20}$$

$$d_{v'0} = 1, \tag{21}$$

$$d_{uk} \leq \sum_{v \in N_H(u)} d_{vk-1}, \ \forall u \in A, k \in \{2, ..., D\} \tag{22}$$

$$d_{u1} = y_u, \ \forall u \in A, v' \notin N_H(u) \tag{23}$$

$$d_{u1} = 1, \ \forall u \in A, v' \in N_H(u) \tag{24}$$

$$y_u = 0, \ \forall u \in A, v' \in N_H(u) \tag{25}$$

$$y_u \in \{0, 1\}, \ \forall u \in A \cup \{v'\} \tag{26}$$

$$d_{vk} \in \{0, 1\}, \ \forall u \in A \cup \{v'\}, \forall k \in \{0, ..., D\} \tag{27}$$

The objective function (17) minimizes the number of edges adjacent to vertex v' to be added. Constraints (18) and (19) indicate that $1 \leq dist(u, v') \leq D, \forall u \in A$. Moreover, constraints (20) and (21) express that the distance from vertex v' to itself is zero. Constraints (22) indicate that if the distance from vertex $u \in A$ to vertex v' is $k \geq 2$, then the distance from one of its adjacent vertices to v' must be $k-1$. Constraints (23) ensure that for each vertex u that is not adjacent to v' in H, then its distance to v' will become equal to 1 in $H' = (V_H, E_H \cup S_H)$ if there is an edge between u and v' in the optimal solution. Constraints (24) and (25) fix the variables of the vertices adjacent to v' in H. Constraints (26) and (27) are the integrality requirements.

We now observe that the following property holds:

Proposition 2 *Let S_H be an optimal solution of the MCBE problem defined by a graph $H = (V_H, E_H)$, a source vertex v' and a constant D, and let $u \in V_H$, $u \neq v'$ be a vertex. If $d_H(u, v') = d \leq D$, then $d_{H'=(V_H, E_H \cup S_H)}(u, v') \leq d$.*

Proof Since $H' = (V_H, E_H \cup S_H)$ is a supergraph of $H = (V_H, E_H)$, then it contains all paths from u to v' that already exists in H. Consequently, the distance from vertex u to v' can not increase in H'.

In other words, all vertices $u \neq v'$ with $d(u, v') \leq D$ can not be, in the optimal solution, at a distance greater than their current distance to v'.

Therefore, constraints (18) can be replaced by the constraints below in an improved formulation:

$$\sum_{k=1}^{D} d_{uk} = 1, \ \forall u \in A, \, d(u, v') > D \tag{28}$$

$$\sum_{k=1}^{d(u,v')} d_{uk} = 1, \ \forall u \in A, \, d(u, v') \leq D \tag{29}$$

$$\sum_{k=d(u,v')+1}^{D} d_{uk} = 0, \ \forall u \in A, \, d(u, v') \leq D, \tag{30}$$

where constraints (28), (29) and (30) make use of the additional information about the distances from vertex v' to all other vertices in graph H.

Table 1 compares the three formulations in terms of their number of variables and constraints, where $n = |V_H|$.

Table 1: ILP formulations: number of variables and constraints.

	Variables	Constraints	All variables binary?
First formulation	$O(n^2)$	$O(n^2)$	No
Second formulation	$O(n^2)$	$O(n^2)$	Yes
Third formulation	$O(nD)$	$O(nD)$	Yes

4 Numerical results

The models were implemented and tested using version 12.7.1 of the CPLEX solver on an Intel Core i7 machine with a 3.2 GHz processor and 8 GB of RAM, running under the Windows 10 operating system.

4.1 Ramdomly generated test problems

Several experiments were performed to assess the performance of the integer programming models presented in the previous section. We created two set of instances: small and medium sized instances. The instances were generated as random graphs with two parameters: the number of vertices n and the number of randomly generated edges m inside the set A. The parameter D of the problem was set to 2, which is a reasonable target in practice. The instances are named indicating the values of n and m. For example, the instance named "inst_200v_4x" has $n = 200$ vertices and $m = n \times 4 = 800$ edges in set A.

Tables 2 and 3 contain the experimental results for the small and medium instances, respectively. For each instance and formulation, the tables display the number of added edges in the best solution found by the solver, the running time in seconds, and an indication whether the instance was solved to optimality or not within a time limit of 3600 seconds.

Table 2 shows that the third formulation outperforms the others, solving to optimality all small instances instances with up to 1000 vertices in much smaller running times.

Table 3 reports the same experimental results for the second and third formulations for the medium-sized instances, but exclusively for those where the number of edges is four or eight times the number of vertices, because for them the optimal solution is not quickly reached. The third formulation obtains better results when the number of vertices increases. We also observe that the memory space requirements of the second formulation increase very quickly with the number of vertices, making it impractical on a machine with a limited amount of memory space: not even feasible solutions were found for the instances with 5000 or more vertices due to memory limitations.

Table 4 illustrates the variation of the linear relaxation gap for the instances with 1000 and 2000 vertices with the increase in the number of edges. For the same instances, Figure 5 displays the evolution of the absolute gap when the number of edges increases. We observe that the largest absolute gap values are reached when the number of edges is 5 or 6 times greater than the number of vertices. For the same instances, the third – and best – formulation takes the longest times to reach the optimum. Therefore, instances with these densities seem to be the hardest to be solved by integer programming techniques.

Another observation that can be drawn from Table 4 is that the higher is the density of each of the polarized groups of vertices in a network, the smaller is the number of edges that should be added in the optimal solution, which makes these problems easier to be solved in practice.

Table 2: Results for small instances.

Instance	First formulation			Second formulation			Third formulation		
	Edges	Time (s)	Solved?	Edges	Time (s)	Solved?	Edges	Time (s)	Solved?
inst_100v_1x	29	0.091	yes	29	0.047	yes	29	0.013	yes
inst_100v_2x	19	0.285	yes	19	0.174	yes	19	0.049	yes
inst_100v_4x	9	1.167	yes	9	0.053	yes	9	0.028	yes
inst_100v_8x	3	1.029	yes	3	0.037	yes	3	0.024	yes
inst_100v_16x	1	1.303	yes	1	0.034	yes	1	0.034	yes
inst_200v_1x	75	0.310	yes	75	0.145	yes	75	0.005	yes
inst_200v_2x	42	0.469	yes	42	0.149	yes	42	0.007	yes
inst_200v_4x	17	6.754	yes	17	0.391	yes	17	0.110	yes
inst_200v_8x	7	9.108	yes	7	0.278	yes	7	0.069	yes
inst_200v_16x	1	27.874	yes	1	0.103	yes	1	0.021	yes
inst_500v_1x	177	2.146	yes	177	0.960	yes	177	0.043	yes
inst_500v_2x	100	8.139	yes	100	1.024	yes	100	0.067	yes
inst_500v_4x	45	171.620	yes	45	21.182	yes	45	3.177	yes
inst_500v_8x	20	101.572	yes	20	6.104	yes	20	0.324	yes
inst_500v_16x	7	250.697	yes	7	1.053	yes	7	0.038	yes
inst_1000v_1x	362	1690.160	yes	362	4.335	yes	362	0.057	yes
inst_1000v_2x	253	3610.160	(no)	197	5.064	yes	197	0.083	yes
inst_1000v_4x	187	3607.410	(no)	97	166.762	yes	97	9.414	yes
inst_1000v_8x	53	3647.120	(no)	38	1963.990	yes	38	50.929	yes
inst_1000v_16x	17	3618.540	(no)	12	4.127	yes	12	0.129	yes
inst_1500v_1x	571	3642.440	(no)	544	11.897	yes	544	0.058	yes
inst_1500v_2x	425	3651.650	(no)	294	12.627	yes	294	0.277	yes
inst_1500v_4x	292	3688.460	(no)	141	3607.990	(no)	139	3602.090	(no)
inst_1500v_8x	201	3647.860	(no)	62	3611.450	(no)	62	3602.940	(no)
inst_1500v_16x	100	3654.160	(no)	15	9.635	yes	15	0.132	yes
inst_2000v_1x	875	3736.490	(no)	702	22.718	yes	702	0.110	yes
inst_2000v_2x	1800	3784.000	(no)	390	25.002	yes	390	0.329	yes
inst_2000v_4x	475	3677.580	(no)	186	3617.370	(no)	186	3601.790	(no)
inst_2000v_8x	1800	4301.730	(no)	83	3623.620	(no)	78	3601.030	(no)
inst_2000v_16x	492	3862.770	(no)	19	22.158	yes	19	0.137	yes

Table 3: Results for medium-sized instances.

Instance	Second formulation			Third formulation		
	Edges	Time (s)	Solved?	Edges	Time (s)	Solved?
inst_3000v_4x	286	3776.23	(no)	280	3600.78	(no)
inst_3000v_8x	118	3723.63	(no)	115	3600.71	(no)
inst_4000v_4x	384	3963.45	(no)	373	3600.53	(no)
inst_4000v_8x	157	4176.90	(no)	153	3600.59	(no)
inst_5000v_4x	-	-	(no)	472	3600.57	(no)
inst_5000v_8x	-	-	(no)	200	3600.57	(no)
inst_6000v_4x	-	-	(no)	573	3600.42	(no)
inst_6000v_8x	-	-	(no)	240	3600.32	(no)
inst_7000v_4x	-	-	(no)	674	3600.56	(no)
inst_7000v_8x	-	-	(no)	278	3600.46	(no)
inst_8000v_4x	-	-	(no)	766	3600.62	(no)
inst_8000v_8x	-	-	(no)	306	3600.47	(no)
inst_9000v_4x	-	-	(no)	863	3600.66	(no)
inst_9000v_8x	-	-	(no)	368	3600.49	(no)
inst_10000v_4x	-	-	(no)	976	3600.86	(no)
inst_10000v_8x	-	-	(no)	398	3600.72	(no)

4.2 Real networks and interpretation

We also applied the solution approach proposed in Section 2 to the two instances recovered from Interian and Ribeiro [13] that appear in Figures 1 and 2: books and blogs, respectively. The third formulation of problem MinCBEAP described in Section 3 was solved for both instances.

Table 5 shows the results. We note that the number of edges in the solution that solves optimally each instance is very small in each case. The intervention associated with the addition of these edges to the graph represents, indeed, a small increase of less than 1% in the number of edges. This fact reflects the minimum intervention principle proposed in the problem formulation, showing that polarization can be reduced by small modifications in the structure of the graph.

These results also illustrate that edge additions make it possible to break the isolation of polarized groups by providing them with more plural information coming from other groups, as noticed in [14].

5 Concluding remarks

In this work, we introduced the Minimum-Cardinality Balanced Edge Addition Problem as a strategy for reducing polarization in real-world networks. We proved the NP-completeness of its decision version. We also proposed three new integer linear formulations for the optimization version, discussing computational results on both randomly generated and real-life instances. On the real-life instances, we showed that polarization can be reduced to the desired

Table 4: Linear relaxation gap on the instances with 1000 and 2000 vertices.

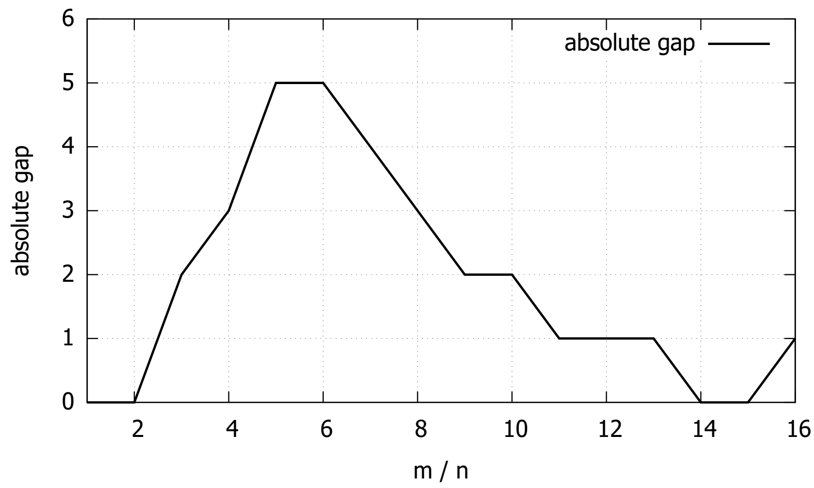
Instance name	Third formulation			Third formulation, LP relaxation			Gap = $\frac{x-\lceil y \rceil}{x}$
	Edges	Time (s)	Solved?	Edges	Time (s)	Solved?	
inst_1000v_1x	362	0.057	yes	362	0.027	yes	0
inst_1000v_2x	197	0.148	yes	196.29	0.034	yes	0
inst_1000v_3x	130	15.463	yes	127.44	0.043	yes	0.015
inst_1000v_4x	97	9.488	yes	93.84	0.041	yes	0.031
inst_1000v_5x	72	779.852	yes	66.79	0.058	yes	0.069
inst_1000v_6x	59	3600.98	(no)	53.41	0.053	yes	0.085
inst_1000v_7x	47	574.383	yes	42.36	0.064	yes	0.085
inst_1000v_8x	38	51.265	yes	34.42	0.041	yes	0.079
inst_1000v_9x	33	144.737	yes	30.14	0.029	yes	0.061
inst_1000v_10x	27	3.942	yes	24.05	0.029	yes	0.074
inst_1000v_11x	23	0.408	yes	21.07	0.044	yes	0.043
inst_1000v_12x	17	0.102	yes	15.89	0.026	yes	0.059
inst_1000v_13x	15	0.120	yes	13.80	0.023	yes	0.067
inst_1000v_14x	13	0.125	yes	12.57	0.024	yes	0
inst_1000v_15x	10	0.078	yes	9.53	0.023	yes	0
inst_1000v_16x	12	0.129	yes	10.62	0.024	yes	0.083
inst_2000v_1x	702	0.110	yes	702	0.056	yes	0
inst_2000v_2x	390	0.329	yes	388.74	0.099	yes	0.003
inst_2000v_3x	253	3603.000	(no)	247.14	0.195	yes	0.020
inst_2000v_4x	186	3601.790	(no)	175.42	0.224	yes	0.054
inst_2000v_5x	140	3601.060	(no)	127.42	0.273	yes	0.086
inst_2000v_6x	117	3600.940	(no)	104.79	0.205	yes	0.103
inst_2000v_7x	94	3601.130	(no)	84.51	0.146	yes	0.096
inst_2000v_8x	78	3601.030	(no)	69.77	0.132	yes	0.103
inst_2000v_9x	58	3601.710	(no)	52.59	0.099	yes	0.086
inst_2000v_10x	50	459.184	yes	45.72	0.078	yes	0.080
inst_2000v_11x	46	3433.460	yes	40.89	0.084	yes	0.109
inst_2000v_12x	38	6.289	yes	35.41	0.069	yes	0.053
inst_2000v_13x	32	4.755	yes	29.50	0.071	yes	0.063
inst_2000v_14x	25	0.507	yes	23.12	0.069	yes	0.040
inst_2000v_15x	22	0.469	yes	20.73	0.072	yes	0.045
inst_2000v_16x	19	0.137	yes	17.84	0.067	yes	0.053

Table 5: Results for real-life instances.

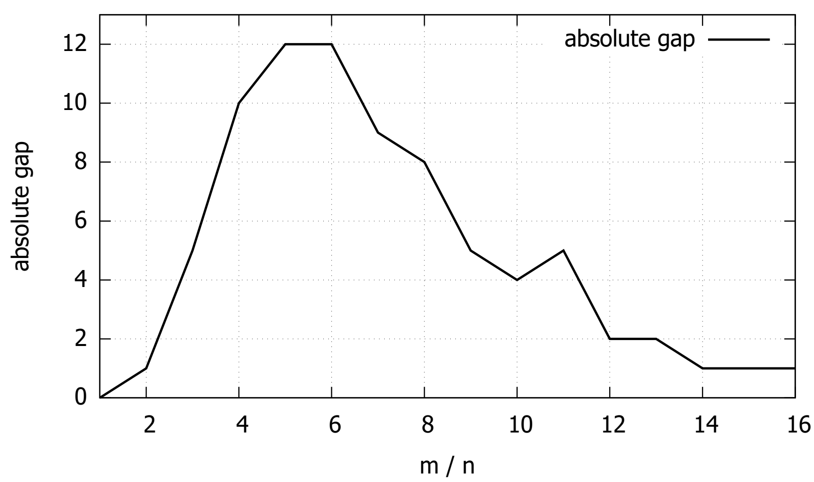
Instance name	Group	Vertices	Edges	ILP Model 3		
				Solution	Time	Solved
books	Conservative	50	420	1	0.006	yes
	Liberal	44	376	2	0.018	yes
	Neutral	14	44	0	0.014	yes
blogs	Republican	637	9352	8	0.048	yes
	Democratic	589	8805	17	0.014	yes

threshold with the addition of a few edges, as established by the minimum intervention principle that guided the problem formulation.

Another interesting conclusion is that in strongly polarized groups, there is often some easy way of spreading polarization-breaking information. This is a consequence of the fact that the higher is the density of a polarized group of vertices in a network, the smaller is the number of edges in the optimal solution, as previously observed in Section 4.1 from the results in Table 4.



(a) Instances with 1000 vertices.



(b) Instances with 2000 vertices.

Fig. 5: Variation of the absolute linear relaxation gaps with the increase in the number of edges: on horizontal axis, the ratio m/n between the number of edges and vertices in set A.

This study also shows that by using edge additions, completely isolated groups mentioned by Landrum [14] can start receiving more plural information, i.e., information coming from more that one group. Therefore, as suggested, disinformation can be broken by providing users a way to encounter diverse views of those practiced by members of the same groups they are trapped in.

Future work involves the study of graph properties that might lead to improvements in the efficiency of exact approaches, as well as the development of heuristic methods for handling hard instances that can not be solved by exact methods.

The minimum intervention principle that guided the approach proposed in this work and the exact methods developed here constitute an effective strategy for tackling polarization problems in real social, interaction, and communication networks. They make it possible to build concrete tools and strategies to address polarization issues in practice, since this is a relevant problem in a world characterized by extreme political and ideological polarization.

References

1. Abello, J., Pardalos, P.M., Resende, M.G.C.: On maximum clique problems in very large graphs. In: J.M. Abello, J.S. Vitter (eds.) External Memory Algorithms, pp. 119–130. American Mathematical Society (1999)
2. Abello, J., Resende, M., Sudarsky, S.: Massive quasi-clique detection. In: J. Abello, J. Vitter (eds.) Proceedings of the 5th Latin American Symposium on the Theory of Informatics, *Lecture Notes in Computer Science*, vol. 2286, pp. 598–612. Springer, Berlin (2002)
3. Adamic, L.A., Glance, N.: The political blogosphere and the 2004 U.S. election: Divided they blog. Proceedings of the 3rd International Workshop on Link Discovery pp. 36–43 (2005)
4. Arendt, H.: Between Past and Future. Viking Press, New York (1968)
5. Bilò, D., Gualà, L., Proietti, G.: Improved approximability and non-approximability results for graph diameter decreasing problems. Theoretical Computer Science **417**, 12–22 (2012)
6. Demaine, E.D., Zadimoghaddam, M.: Minimizing the diameter of a network using shortcut edges. Proceedings of the 12th Scandinavian Workshop on Algorithm Theory pp. 420–431 (2010)
7. Dictionaries, O.: Definition of polarization. Last access on 2017-09-06. https://en.oxforddictionaries.com/definition/polarization
8. Economist, T.: Not persuaded – political polarization. Last access on 2017-11-13. https://www.economist.com/blogs/democracyinamerica/2015/05/political-polarisation
9. Fabrikant, A., Luthra, A., Maneva, E., Papadimitriou, C.H., Shenker, S.: On a network creation game. Proceedings of the 22nd Annual Symposium on Principles of Distributed Computing pp. 347–351 (2003)
10. Garey, M.R., Johnson, D.S.: Computers and Intractability; A Guide to the Theory of NP-Completeness. W. H. Freeman & Co., New York (1990)
11. Garimella, K., De Francisci Morales, G., Gionis, A., Mathioudakis, M.: Reducing controversy by connecting opposing views. In: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, pp. 81–90. ACM, New York (2017)
12. Harary, F.: Graph Theory. Addison Wesley (1969)
13. Interian, R., Ribeiro, C.C.: An empirical investigation of network polarization. Applied Mathematics and Computation **339**, 651 – 662 (2018)
14. Landrum, A.: YouTube as a primary propagator of flat Earth philosophy. In: American Association for the Advancement of Science Annual Meeting. Washington, DC (2019). https://aaas.confex.com/aaas/2019/meetingapp.cgi/Session/21893
15. Li, C.L., McCormick, S.T., Simchi-Levi, D.: On the minimum-cardinality-bounded-diameter and the bounded-cardinality-minimum-diameter edge addition problems. Operations Research Letters **11**, 303–308 (1992)
16. Mill, J.S.: On liberty. J. W. Parker and Son, London (1859)
17. Miller, C.E., Tucker, A.W., Zemlin, R.A.: Integer programming formulation of traveling salesman problems. Journal of the ACM (JACM) **7**, 326–329 (1960)
18. Newman, M.E.J.: Network data – Books about US politics. Last access on 2017-12-11. http://www-personal.umich.edu/ mejn/netdata/

19. Papagelis, M.: Refining social graph connectivity via shortcut edge addition. ACM Transactions on Knowledge Discovery Data **10**, 12:1–12:35 (2015)
20. Papagelis, M., Bonchi, F., Gionis, A.: Suggesting ghost edges for a smaller world. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, pp. 2305–2308. ACM, New York (2011)
21. Perumal, S., Basu, P., Guan, Z.: Minimizing eccentricity in composite networks via constrained edge additions. In: 2013 IEEE Military Communications Conference, pp. 1894–1899. San Diego (2013)
22. Pinto, B.Q., Ribeiro, C.C., Rosseti, I., Plastino, A.: A biased random-key genetic algorithm for the maximum quasi-clique problem. European Journal of Operational Research **271**, 849–865 (2018)
23. Ribeiro, C.C., Riveaux, J.A.: An exact algorithm for the maximum quasi-clique problem. International Transactions in Operational Research **26**, 2199–2229 (2018)
24. Ribeiro, M.H., Calais, P.H., Almeida, V.A.F., Jr., W.M.: Everything I disagree with is #FakeNews: Correlating Political Polarization and Spread of Misinformation. Workshop on Data Science + Journalism @KDD 2017. https://sites.google.com/view/dsandj2017/accepted-papers
25. Times, N.Y.: How we became bitter political enemies. Last access on 2017-11-13. https://www.nytimes.com/2017/06/15/upshot/how-we-became-bitter-political-enemies.html
26. Vogiatzis, C., Walteros, J.L.: Integer programming models for detecting graph bipartitions with structural requirements. Networks **71**, 432–450 (2018)
27. Walteros, J.L., Veremyev, A., Pardalos, P.M., Pasiliao, E.L.: Detecting critical node structures on graphs: A mathematical programming approach. Networks **73**, 48–88 (2019)