A centrality notion for graphs based on Tukey depth

J. Orestes Cerdeira[a,*], Pedro C. Silva[b]

[a] Centro de Matemática e Aplicações (CMA) and Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa
[b] Centro de Estudos Florestais (CEF), Instituto Superior de Agronomia, Universidade de Lisboa.

Abstract

Centrality on graphs aims at ranking vertices in terms of their contribution to facilitate the communication flow in the network. Tukey depth is one of most widely used statistical measures to assess the centrality of a point within a cloud of points in the multidimensional space. In this paper we propose and discuss how to adapt Tukey depth to develop a novel centrality index for vertices of a graph. We present some properties of the indices on several classes of graphs, show that computing the indices is NP-hard, extend the indices to assess the centrality of group of vertices and give 0/1 linear formulations to calculate them.

Key words: Centrality measures, Median points, Convexity, Unimodal distribution, Quasi-concave function, Computational complexity, Social networks.

1. Introduction

(Data) depth functions measure the centrality of a point with respect to a set of points generalizing ordering ranks for multivariate data. Tukey depth [1] is one of the most popular depth functions in the literature, and the most widely used multivariate generalization of the notion of median for univariate data [2, 3].

If X is a finite set of points of $\mathbb{R}^d$ and x a point of $\mathbb{R}^d$ (belonging or not to X), the (integer version of the) *Tukey depth* of x with respect to X is the minimum number of points of X lying in one side of a hyperplane through x. Equivalently, the Tukey depth of x is the minimum number of points to be removed from X in order that the convex hull of the remaining points of X does not include x.

A point of maximum Tukey depth is called a *Tukey median point*. Points with zero Tukey depth are the points outside the convex hull of X. Points of X with Tukey depth one are in the border of the convex hull of X. Assuming

*Corresponding author
Email addresses: jo.cerdeira@fct.unl.pt (J. Orestes Cerdeira), pcsilva@isa.ulisboa.pt (Pedro C. Silva)

the points of X are in general position, these are the vertices of the convex hull of X. Points with higher Tukey depths are located in the inner regions of X. Figure 1 (Left) illustrates the concept of Tukey depth of a point in the plane with respect to a set X of 10 points of $\mathbb{R}^2$. Figure 1 (Right) depicts the regions with the same Tukey depth with respect to X, ranging from zero (the set of points outside the convex hull of X) to four (points lying in the innermost convex region).

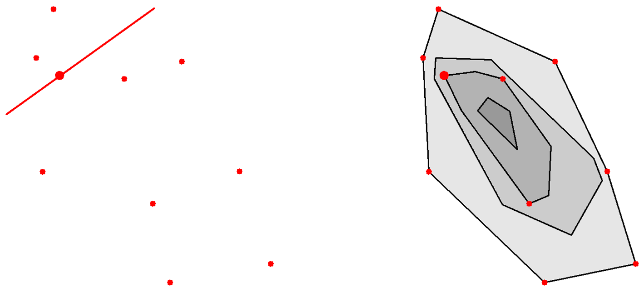

Figure 1: X is the set of the 10 red dots (the set of points is the same in both images) and $x \in X$ is represented by the largest red dot. On the left: the line through x delimits a halfplane containing a subset X' of X of minimum cardinality. Note that X' is also a subset of X of minimum cardinality such that the convex hull of $X \setminus X'$ does not contain x. Hence, the Tukey depth of x equals $|X'| = 3$. On the right: the regions with the same grey tone represent sets of points with the same Tukey depth with respect to X. The Tukey depths range from 0 (points lying outside the outer polygon) to four (points lying in the innermost polygon, i.e., corresponding to Tukey median points).

In the univariate case the Tukey median of a set of points X of $\mathbb{R}$, i.e., a point of $\mathbb{R}$ with maximum depth $\lceil \frac{|X|}{2} \rceil$, corresponds to the usual median of X.

Computing Tukey depth is NP-hard [4].

Tukey depth is a quasi-concave function, which is an important property to settle a center-outward ranking of points. A function $f : \mathbb{R}^d \to \mathbb{R}$ is *quasi-concave* if, for every $\tau \in \mathbb{R}$, the set $\Sigma_\tau(f) = \{x \in \mathbb{R}^d : f(x) \geq \tau\}$ is convex. Equivalently, function f is quasi-concave if $f(\lambda x + (1-\lambda)y) \geq \min\{f(x), f(y)\}$, for $0 < \lambda < 1$. Thus, the Tukey depth of any point of a $u-v$ geodesic is greater than or equal to the minimum value of the Tukey depths of u and v. (For other properties of depth functions we refer to [5, 6].)

Centrality in graphs aims at ranking vertices in terms of their contribution to facilitate the communication flow in a network. A number of centrality measures have been proposed (see, e.g, [7, 8]). Here we propose to adapt the notion of Tukey depth to graphs to assess the centrality of vertices, by employing the graph analogues of geodesic and convexity.

The paper is organized as follows. In Section 2 we introduce the new centrality indices for the vertices, which we call Tukey centrality, illustrate these measures in some classes of graphs including trees, chordless cycles, and grid graphs, and prove that computing Tukey centrality is NP-hard. Tukey centrality naturally extends to groups of vertices. We address this in Section 3. Linear 0/1 formulations to compute Tukey centrality are given in Section 4. Tukey depth is a quasi-concave function in $\mathbb{R}^d$. In Section 5 we draw some implications resulting from Tukey centrality in the light of quasi-concave's analogue in graphs. We finish, in Section 6, summarizing the main aspects of this study and enumerating some topics to be addressed in future work.

2. Tukey centrality: definitions, examples and complexity

Let G denote a simple graph of *order* $n \geq 2$ (the number of vertices) and *size* $m \geq 1$ (the number of edges). The corresponding vertex and edge sets shall be denoted by $V(G)$ and $E(G)$, respectively. The (open) neighborhood of a vertex $v \in V(G)$, denoted by $N_G(v)$, is the set of vertices of G adjacent to v.

If $S \subset V(G)$ is a subset of vertices, $G[S]$ denotes the subgraph of G induced by S, i.e., the graph whose vertex set is S and whose edge set consists of the edges in $E(G)$ that have both endpoints in S. We denote by $G - S$ the graph obtained by removing S and all edges incident with every vertex in S. For simplicity we use $G - v$ to denote $G - \{v\}$.

The distance between a pair of vertices $u, v \in V(G)$, denoted by $d_G(u, v)$, is the length (number of edges) of a shortest $u - v$ path in G. A $u - v$ geodesic is a $u - v$ path of length $d_G(u, v)$. If G is not connected and u, v belong to different connected components, $d_G(u, v) = +\infty$.

Throughout this paper G will be a connected graph.

To define centrality measures for the vertices, that are the discrete counterparts of the (continuous) Tukey depth, we use the graph analogues of convexity.

Convexity in graphs was addressed in [9]. A subset S of vertices of G is a (strong) convex set if for any two vertices $u, v \in S$, $G[S]$, the subgraph of G induced by S, contains every $u - v$ geodesic in G. The (strongly) convex hull of a set S of vertices is the smallest convex set of vertices that contains S.

Definition 1. The *strong Tukey centrality* of $v \in V(G)$, denoted by $STC_G(v)$, is the minimum number of vertices to be removed from $V(G)$ in order that the strongly convex hull of the remaining vertices does not include v.

A weak version of convexity in graphs was introduced in [10, 11]. A subset S of vertices of G is called weakly convex (also called isometric set) if for any two vertices $u, v \in S$, $G[S]$ contains a $u - v$ geodesic in G, i.e., $d_{G[S]}(u, v) = d_G(u, v)$. A weakly convex hull of a set S of vertices is a smallest weakly convex set of vertices containing S.

Definition 2. The *weak Tukey centrality* of vertex $v \in V(G)$, denoted $WTC_G(v)$, is the minimum number of vertices to be removed from $V(G)$ in order that the weakly convex hull of the remaining vertices does not include v.

The above definitions can be equivalently stated as follows.

Remark 1.

- *$STC_G(v) = n - |W|$, where W is a strong convex set of G, not containing v, of maximum cardinality.*
- *$WTC_G(v) = n - |W|$, where W is a weak convex set of G, not containing v, of maximum cardinality.*

Since every strong convex set of G is also weakly convex we have the following.

Remark 2. *$WTC_G(v) \leq STC_G(v)$ for every vertex v of G.*

We call a vertex with maximum strong [weak] Tukey centrality a strong [weak] *Tukey median* vertex of G. We call a vertex v of G a strong [weak] *border* vertex if $STC_G(v) = 1$ [$WTC_G(v) = 1$].

The following examples illustrate strong and weak Tukey centrality $STC_G(v)$ and $WTC_G(v)$, for every vertex v, on several graphs. The vertices in the figures are depicted as colored dots, with a gradient of colors ranging from white for strong and weak border vertices to red for strong Tukey medians.

We shall omit the subscript G from the above notations whenever no ambiguity arises.

Example 1. Figure 2 (Left) and Figure 2 (Right) show the values of the strong and weak Tukey centrality, respectively, for each one of the $n = 6$ vertices of the graph G. Let v denote the vertex located in the center of the graph. Since $V(G) \setminus \{v\}$ is weakly convex in G, $WTC(v) = 1$. However, $V(G) \setminus \{v\}$ is not strongly convex since it does not contain one of the geodesics connecting the two neighbors of v. Hence we need also to remove one of these two neighbors, which we denote by u (it is irrelevant which one we remove). The set of vertices $V(G) \setminus \{u, v\}$ is still not strongly convex, since it does not contain the geodesic connecting the topmost vertex and one of the two bottom vertices. Hence we still need to remove one of these two neighbors of u, say w. The set $G \setminus \{u, v, w\}$ is now strongly convex in G and therefore $STC(v) = 3$. The disparity between the $STC(v)$ and $WTC(v)$ values is related with the quasi-concave property that holds for the strong Tukey centrality but does not hold for the weak case. We shall address this in the next section.

For geodetic graphs, i.e., graphs having a single geodesic connecting every pair of vertices, strong and weak Tukey centrality coincide.

Example 2. For the path graph of order n, P_n with $V = \{v_1, \ldots, v_n\}$ and $E = \{(v_i, v_{i+1}) : i = 1, \ldots, n-1\}$, $STC(v_i) = WTC(v_i) = \min\{i, n - i + 1\}$.

See in Figure 3 the Tukey centrality of the vertices of P_9.

In particular, both strong and weak Tukey median(s) of P_n have centrality $\lceil \frac{n}{2} \rceil$ and consists of one middle vertex if n is odd (Figure 3) and two middle vertices, otherwise.

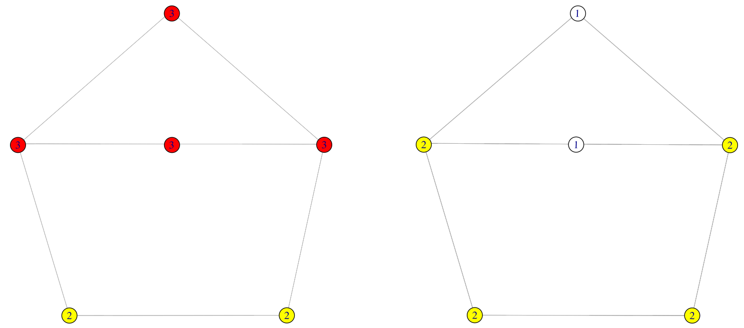

Figure 2: Tukey centrality values of each of the $n = 6$ vertices of a graph G with the values of $STC(v)$ depicted on the left side and the values of $WTC(v)$ on the right side.

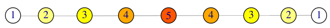

Figure 3: Tukey centrality values $STC(v) = WTC(v)$ of each vertex v of the path graph P_9.

Example 3. Paths are tree graphs with exactly two leaves. Stars are trees with $n-1$ leaves, i.e. complete bipartite graphs, $K_{1,n-1}$, with a singleton bipartite class. In a star, the strong and weak Tukey centrality are both equal to one, for leaves, and $n-1$ for the remaining vertex, which is (if $n \geq 3$) the unique Tukey median vertex. Figure 4 shows the Tukey centrality of the vertices of the star $K_{1,5}$.

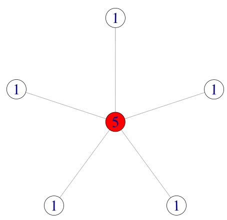

Figure 4: Tukey centrality values $STC(v) = WTC(v)$ of each vertex v of the star $K_{1,5}$.

Example 4. In general, for a tree graph Γ_n with n vertices, strong and weak Tukey centrality of a vertex v equals n minus the order of the largest connected component of the forest $\Gamma_n - v$ (i.e, the subgraph induced by $V \setminus \{v\}$). See Figure 5.

Example 5. For the chordless cycle C_n we have $STC(v) = WTC(v) = \lfloor \frac{n}{2} \rfloor$, if n is odd, and $STC(v) = WTC(v) + 1 = \frac{n}{2}$, if n is even, for all vertices v. In particular, all vertices are strong and weak Tukey medians (see Figure 6).

Example 6. Consider the 2-dimensional $k \times \ell$ regular grid with set of vertices $V = \{(i,j) : 1 \leq i \leq k,\ 1 \leq j \leq \ell\}$ and set of edges $\{\{(i,j),(i',j')\} : |i-i'| + |j-j'| = 1\}$.

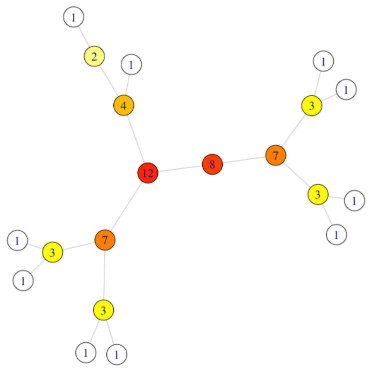

Figure 5: Tukey centrality values $STC(v) = WTC(v)$ of each vertex v of a tree Γ_{20}.

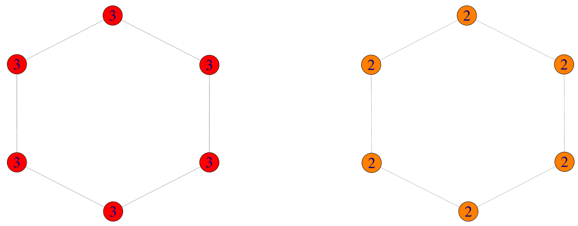

Figure 6: Strong Tukey centrality values $STC(v) = 3$ (Left) and weak Tukey centrality values $WTC(v) = 2$ (Right) of each vertex v of the cycle C_6.

Each vertex (i, j) divides the grid into four rectangular subgrids ("quadrants"), whose common boundaries are contained in row i and column j.

If we remove (i, j), it is also necessary to remove two of the adjacent "quadrants" in order that the set of the remaining vertices is strongly convex. Hence, for the strong Tukey centrality we get

$$STC((i, j)) = \min\{\ell\, i, \ell(k - i + 1), k\, j, k\, (\ell - j + 1)\},$$

for every vertex (i, j) (Figure 7, Left).

For the weak Tukey centrality, when we remove vertex (i, j), to get a weak convex set, we also need to remove one of the "quadrants". Thus,

$$WTC((i, j)) = \min(i, k - i + 1)\ \min(j, \ell - j + 1),$$

for every vertex (i, j) (Figure 7, Right).

Example 7. The Zachary's karate club network is a social network that is widely used to test and compare centrality measures (see, e.g., [12]) and procedures to identify community structures in graphs (see, e.g., [13]). The graph consists of 34 vertices, representing members of a karate club, and edges indicating interactions outside the club between pairs of members (see Figure 8).

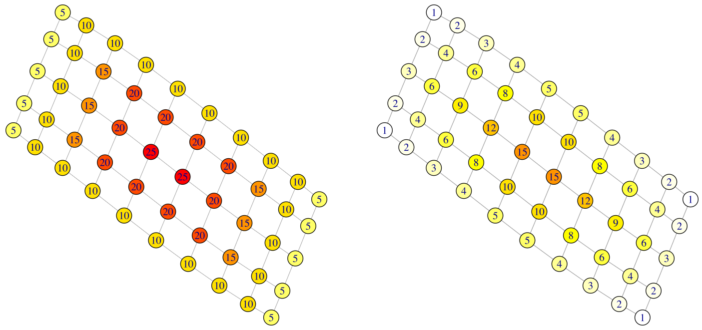

Figure 7: Strong Tukey centrality values $STC(v)$ (Left) and weak Tukey centrality values $WTC(v)$ (Right) of each vertex of v the 2-dimensional regular 5×10 grid.

The network derives from a study that Zachary [14] carried out from 1970 to 1972. During that period a conflict between the instructor (vertex 1) and the administrator (vertex 34) led the members of the club to split with about half of the members following the instructor (vertices represented by squares in Figure 8) which started a new club.

Figure 9 (Left) and Figure 9 (Right) indicate the values of the strong and the weak Tukey centrality, respectively, of each one of the 34 vertices of the Zachary's karate club network. The instructor (vertex 1) is the Tukey median vertex of the graph, with strong and weak Tukey centrality equal to 21 and 12, respectively. The administrator (vertex 34) is the vertex with weak Tukey centrality equal to 8, and has strong Tukey centrality equal to 19. Almost half of the vertices in the Zachary's karate club network have strong Tukey centrality values of 19 or 21, while all the others do not exceed value 3. This behaviour of strong Tukey centrality replicates the performance of Tukey depth on multi-modal data where, typically, the highest depth values are assigned to a vast interior region.

The weak Tukey centrality has a wider range of values and provides a better discrimination of the centrality of the vertices of a graph. All but four vertices in the Zachary's network with highest strong centrality values (greater than or equal to 19) have weak centrality 1, 2 or 3. The other four vertices correspond to the instructor $WTC(1) = 12$, the administrator $WTC(34) = 8$ and vertices 3 and 33, with $WTC(3) = WTC(33) = 6$. The vertex 3 is often misclassified by community detection approaches that test their performance in Zachary's network [15]. The member corresponding to vertex 33 plays in the network a "role" very similar to the administrator (vertex 34). They are adjacent, have a large number of common neighbours among the members of the administrator's faction (represented by circles in Figure 8) and each has exactly two and one neighbour among the instructor members faction (represented by squares in Figure 8).

We saw that for trees Γ_n with n vertices, where every non-pendant vertex

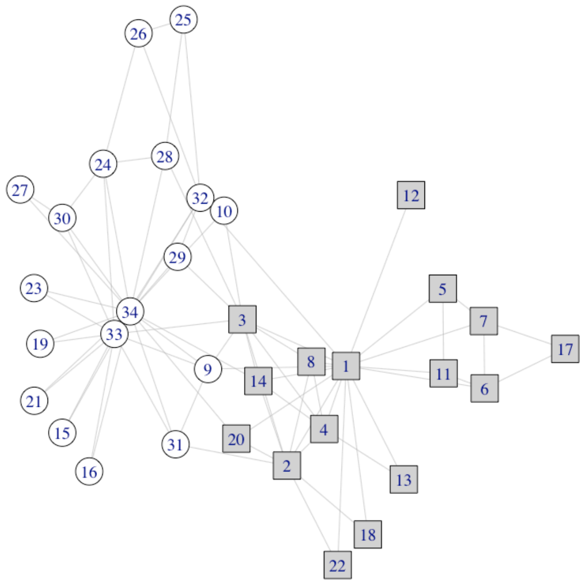

Figure 8: The Zachary's karate club network [14], as it appears in [13]. Squares represent the members of the faction associated with the instructor, vertex 1. Circles represent the members of the faction associated with the administrator, vertex 34.

is a cut-vertex, strong and weak Tukey centrality of a vertex v equals n minus the order of the largest connected component of the forest $\Gamma_n - v$. The same result applies to general graphs.

Lemma 3. *The strong and weak Tukey centrality of a pendant-vertex or a cut-vertex v of graph G with n vertices equals n minus the order of the largest connected component of $G - v$.*

The next result gives (polynomial-time checkable) characterizations of the strong and weak border vertices of a graph.

Lemma 4. *Let v be a vertex of G. Then*

1. $STC(v) = 1$ *iff* $d_{G-v}(u, w) = 1$*, for all* $u, w \in N_G(v)$*.*
2. $WTC(v) = 1$ *iff* $d_{G-v}(u, w) \leq 2$*, for all* $u, w \in N_G(v)$*.*

Proof: Let us prove the first claim. Assume that $STC(v) = 1$. Then $V(G)\setminus\{v\}$ is strongly convex in G, which implies $d_{G-v}(u, w) = 1$, for all $u, w \in N_G(v)$. To prove the converse assume that $STC(v) > 1$. Then there are vertices $u, w \in V(G) \setminus \{v\}$ and a geodesic connecting these vertices that contains v. Hence we can find vertices $u', w' \in N_G(v)$ and a geodesic connecting these vertices also containing v. This implies $d_{G-v}(u', w') > 1$ and the result follows.

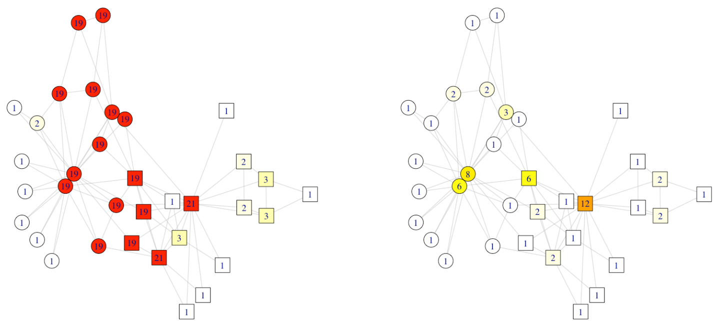

Figure 9: Strong Tukey centrality values $STC(v)$ (Left) and weak Tukey centrality values $WTC(v)$ (Right) of each vertex of v of the Zachary's karate club network [14].

Let us now prove the second claim. If $WTC(v) = 1$, $G - v$ is weakly convex, which implies $d_{G-v}(u, w) = d_G(u, w)$, for all $u, w \in V(G - v)$. If $v, w \in N_G(v)$, we have $d_{G-v}(u, w) = d_G(u, w) \leq 2$. Conversely, if $WTC(v) \geq 2$, $V(G) \setminus \{v\}$ is not weakly convex in G. Hence we can find $u, w \in V(G) \setminus \{v\}$ such that every the geodesic connecting these vertices also contains v. Then we can also find vertices $u', w' \in N_G(v)$ such that every geodesic connecting these vertices must contain v and the result follows. □

For the complete graph of order n, K_n, all vertices are strong and weak border vertices, that is, we have the following result.

Remark 5. $STC(v) = WTC(v) = 1$ *for all* $v \in V(K_n)$.

We have noted previously that Tukey depth is a quasi-concave function in $\mathbb{R}^d$, i.e., for every pair of points $u, w \in \mathbb{R}^d$, the Tukey depth of any point of a (unique) $u{-}w$ geodesic is greater than or equal to the minimum value of the Tukey depths of u and v. The strong Tukey centrality satisfies in graphs a robust version of this property. Specifically, in every graph, the intermediate vertices along all $u - w$ geodesics have strong Tukey centrality values that are greater than or equal to the minimum between $STC(u)$ and $STC(w)$. Equivalently, in every graph G, the set $\Sigma_\tau(STC) = \{v \in V(G) : STC(v) \geq \tau\}$, with $\tau = 1, \ldots, k$, where k is the Tukey centrality of a strong median vertex, is strongly convex. We will say that the strong Tukey centrality is a *(strong) quasi-concave* function in graphs. The proof that the strong Tukey centrality is strongly quasi-concave is given below.

Lemma 6. *Let* G *be an arbitrary graph. If* v *lies in a* $u - w$ *geodesic of* G, $STC(v) \geq \min\{STC(u), STC(w)\}$.

Proof: Let W be a largest strongly convex set not containing v. Then we have $STC(v) = n - |W|$. Since one of these vertices, u or w do not belong to W (otherwise $v \in W$), the strong Tukey centrality of one of these vertices is at most $n - |W|$ and the result follows. □

For the weak Tukey centrality, for vertices v lying in $u-w$ geodesics, it may happen that $WTC(v) < \min\{WTC(u), WTC(w)\}$. In the 2-dimensional 5×10 grid graph represented in Figure 7 there are several $u-w$ geodesics where, for all interior vertices v, $WTC(v) < \min\{WTC(u), WTC(w)\}$ (see Figure 7, Right). We will develop this further in Section 5.

We close this section proving that computing $STC(v)$ and $WTC(v)$ is NP-hard. We do this using the NP-hardness of the maximum clique and of the 2-club problems, respectively. Given a graph H and a positive integer k, the maximum clique and the 2-club problems ask whether there is a subset of vertices $S \subset V(H)$ of size at least k such that $H[S]$, the subgraph of H induced by S, has diameter at most 1, for maximum clique, and at most 2, for the 2-club problem, i.e., $d_{H[S]}(u, v) \leq 1$, resp. 2, for all $u, v \in S$. Maximum clique and the 2-club problem are both NP-complete [16, 17] and [18], respectively.

Lemma 7. *Computing strong and weak Tukey centrality of vertices of a graph are NP-hard.*

Proof: Given graph H and integer k, add a vertex v to $V(H)$ and connect v to every other vertex of H. Let $\tilde{H}$ be the graph thus obtained. The result now follows from the fact that the size of a maximum clique is equal to $|\tilde{H}| - STC_{\tilde{H}}(v)$ and the size of a maximum 2-club in H is equal to $|\tilde{H}| - WTC_{\tilde{H}}(v)$. $\square$

3. Generalizing Tukey centrality to groups of vertices

The centrality of groups of more than one vertex was introduced by Everett and Borgatti [19] to answer questions such as "how central is the engineering department in the informal influence network of this company?" or "among middle managers in a given organization, which are more central, the men or the women?".

Group centrality has been used in different contexts, e.g., to study group features in peer relationship network of school-age children from kindergarten to 2nd grade [20]; to analyse the impact of radiotherapy on cancer patients [21]; to assess conservation strategies for networks of marine protected areas [22]; to locate facilities, in a balanced manner, in spatial networks [23].

In [19], Everett and Borgatti proposed what they called "proper generalization of the corresponding individual measure" for each of three well known centrality measures: degree, closeness and betweenness [24]. "Proper generalization" measure means that when applied to a group consisting of a single vertex, the measure yields the same answer as the individual version.

Tukey centrality provides, in a very natural way, a proper generalization to groups of vertices.

Definition 3. The *strong* [*weak*] *Tukey group centrality* of a proper subset $S \subset V(G)$, denoted $STC(S)$ [$WTC(S)$], is the minimum number of vertices (including S) needed to be removed from $V(G)$ so that the strong [weak] convex hull of the remaining vertices does not intersect S.

From the above definition we have $STC(S) = n - |W|$ $[WTC(S) = n - |W|]$, where W is a strong [weak] convex set of G of maximum cardinality that does not intersect S. In particular, the strong [weak] Tukey group centrality of the singleton $\{v\}$ corresponds to the strong [weak] Tukey centrality of vertex v, i.e., Tukey centrality is a proper generalization measure.

In a tree graph Γ_n with n vertices, the strong [and weak] Tukey centrality of a proper set of vertices S equals n minus the order of the largest connected component of the forest $\Gamma_n - S$ (i.e, the subgraph induced by $V(\Gamma_n) \setminus \{S\}$).

Note that if W is a strong [weak] convex set and $W \cap S = \emptyset$, then for every $v \in S$, $STC(v) \leq |V(G) \setminus W|$ $[WTC(v) \leq |V(G) \setminus W|]$. Thus, the following holds.

Lemma 8. *For every vertex $v \in S$, $STC(v) \leq STC(S)$ $[WTC(v) \leq WTC(S)]$.*

We can therefore write

$$\max_{v\in S} STC(v) \leq STC(S) \quad [\max_{v\in S} WTC(v) \leq WTC(S)] \tag{1}$$

Examples may be given to show that the lower bound of (1) can be achieved, with $S \neq \{v\}$.

For the path graph P_n with $V = \{v_1, \ldots, v_n\}$ and $E = \{(v_i, v_{i+1}) : i = 1, \ldots, n-1\}$ (P_9 is represented in Figure 3), if $S = \{v_1, \ldots, v_k\}$, with $k \leq \frac{n}{2}$, $STC(S) = STC(v_k)[= WTC(v_k) = WTC(S)]$. Any subset of S that includes v_k also has strong [weak] Tukey centrality equal to $STC(v_k)[= WTC(v_k)]$.

Any proper subset of vertices of a star graph (the star graph $K_{1,5}$ is displayed in Figure 4), with $n \geq 3$, that includes the internal vertex (the unique non-leaf), has the same strong [weak] Tukey centrality $(n-1)$ as that of the internal vertex.

Any subset of the set of vertices of a path of length at most $\lfloor\frac{n-1}{2}\rfloor$ included in the chordless cycle C_n (see Figure 6) has the same strong [weak] Tukey centrality as that of any of its individual elements.

For the strong group Tukey the following upper bound on $STC(S)$ can be proved.

Lemma 9. $STC(S) \leq \sum_{v\in S} STC(v)$.

Proof: For each $v \in S$ let W_v be a strong convex set of G of maximum cardinality, such that $v \notin W_v$. In particular, $STC(v) = |V(G) \setminus W_v|$. Set $W_S = \cap_{v\in S} W_v$. Then W_S is a strong convex set and $S \cap W_S = \emptyset$. Moreover,

$$V(G) \setminus W_S = V(G) \setminus (\cap_{v\in S} W_v) = \cup_{v\in S} (V(G) \setminus W_v).$$

Therefore,

$$STC(S) \leq |V(G) \setminus W_S| \leq \sum_{v\in S} |V(G) \setminus W_v| = \sum_{v\in S} STC(v). \quad \square$$

The upper bound on $STC(S)$ given by Lemma 9 can be achieved. This is the case, for instance, when S is any set of leaves of a tree.

Combining lemmas 8 and 9 we can conclude that the value of the strong Tukey centrality of a group of vertices is at least the centrality value of the most central of its members and cannot exceed the sum of the individual Tukey centrality values.

For the weak group Tukey centrality it may happen that

$$WTC(S) > \sum_{v \in S} WTC(v).$$

This is case, for instance, if u and v are 'antipodal' vertices of the cycle C_4, where $WCT(u) = WCT(v) = 1$, while $WCT(\{u,v\}) = 3$. This is a consequence of the fact that the intersection of weak convex sets may not be a weak convex set. However, since every strong convex set of G is also a weak convex set, we have the following remark which extends Remark 2 for the Tukey centrality of groups.

Remark 10. *$WTC(S) \leq STC(S)$ for every proper subset S of $V(G)$.*

We illustrate the group Tukey centrality in the Zacharys karate club network.

Vertices 1 and 34, corresponding to the instructor and the administrator, respectively, are the vertices of maximum weak Tukey centrality. Together they have maximum strong and weak Tukey centrality, $STC(\{1,34\}) = 29$ and $WTC(\{1,34\}) = 19$, among all subsets of two vertices. The set $\{1,33\}$ has the second highest value both for strong and for weak centrality, $STC(\{1,33\}) = 28$ and $WTC(\{1,33\}) = 15$. The "role" similarity in the network between the administrator (vertex 34) and the member corresponding to vertex 33 have been previously noted. This is well captured by the Tukey centrality of the group formed by vertices 33 and 34 together, $STC(\{33,34\}) = STC(33) = STC(34) = 19$, i.e., there is no gain in centrality from having the two members together compared to the centrality of each individual member. The strong Tukey centrality of the vertices corresponding to the instructor members faction (represented by squares in Figure 8) and to the administrator member faction (represented by circles in Figure 8) are equal to 21 and 19, respectively, which are the numbers of members in each of the corresponding factions. This is equivalent to say that each of the two sets is strongly convex.

We finish this section noting that Tukey depth could also be extended to assess the depth of a set of points in $\mathbb{R}^d$. More precisely, if X is a finite set of points of $\mathbb{R}^d$ the Tukey depth of $S \subset X$ is the minimum number of points that should be removed from X so that the convex hull of the remaining points of X does not intersect S. In this context, the analogue of Lemma 8 for group Tukey depth follows directly from the above definition. Moreover, since the proof of Lemma 9 only depends on the fact that the intersection of convex sets is a convex set, the equivalent of Lemma 9 for group Tukey depth for clouds of points also holds. Thus, if X is a finite set of points of $\mathbb{R}^d$ and $TD(S)$ denotes the Tukey depth of a proper subset S of X, we have the following.

$$\max_{x \in S} TD(x) \leq TD(S) \leq \sum_{x \in S} TD(x).$$

4. Linear 0/1 formulations to compute Tukey centrality

It is straightforward to settle models to derive strong and weak Tukey centrality values using 0/1 variables x_u associated to vertices $u \in V(G)$, where $x_u = 1\,(0)$ means that vertex u is (not) removed from $V(G)$.

The strong Tukey centrality of a proper $S \subset V(G)$ is the optimal value of the 0/1 linear problem:

$$STC(S) = \min \sum_{u \in V} x_u \tag{2}$$

$$x_u + x_w \geq x_s, \qquad s \in V(G) : d(u,s) + d(s,w) = d(u,w) \tag{3}$$

$$x_v = 1, \qquad v \in S \tag{4}$$

$$x \in \{0,1\}^n \tag{5}$$

Equation (4) forces all vertices $v \in S$ to be removed. Inequalities (3) ensure that if vertex s, that belongs to an $u - w$ geodesic, is removed, then u or v should also be removed. Constraint (5) establishes the range of variables x, and the objective function (2) defines $STC(v)$ to be the minimum number of vertices removed.

The weak Tukey centrality of $S \subset V(G)$ is given by (4), (5) and

$$WTC(S) = \min \sum_{u \in V} x_u \tag{6}$$

$$x_u + x_w \geq \sum_{s \in S_{uw}} x_s - |S_{uw}| + 1, \tag{7}$$

with $S_{uw} \subset \{s \in V(G) : d(u,s) + d(s,w) = d(u,w)\}$

minimal such that $G - S_{uw}$ has no $u - w$ geodesic.

Since S_{uw} is a set of vertices whose removal eliminates all $u - w$ geodesics, each inequality (7) states that if all vertices of S_{uw} are removed, then u or w should also be removed.

Interestingly, Kratica et al. [25] give 0/1 formulations for the convex dominating and weakly convex dominating set problems, that allow to conclude that i) inequalities (3) may be replaced by

$$x_u + x_w \geq x_s, \qquad s \in N(u) : d(u,s) + d(s,w) = d(u,w) \tag{8}$$

i.e., only the vertices s which are neighbours of u have to be considered, for the strong Tukey centrality model (see also Remark 13); and ii) inequalities (7) may be replaced by

$$x_u + x_w \geq \textstyle\sum_{s \in S^u_{uw}} x_s - |S^u_{uw}| + 1, \tag{9}$$

with $S^u_{uw} = \{s \in N(u) : d(u,s) + d(s,w) = d(u,w)\}$, i.e., only $S^u_{uw} = S_{uw} \cap N(u)$, the set of vertices belonging to $u - w$ geodesics which are neighbours of u, has to be considered for the weak Tukey centrality model (see also Remark 14).

To obtain the strong and weak Tukey for a vertex v it is enough to consider $S = \{v\}$.

As stated above, computing strong and weak Tukey centrality are NP-hard. Moreover, it may well be intend to determine the centrality of all vertices or of several subsets of vertices of the graph. The valid inequalities arising from Lemma 6, Remark 10 and from the Remark 11 below may be added to the formulations in order to accelerate the computation of the Tukey centrality.

Remark 11. *If $v \in V(G)$ and W is a strong convex set of maximum cardinality such that $v \notin W$ then $STC(v) = STC(S)$, where $S := V(G) \setminus W$. In particular, $STC(S) = |S|$ is an upper bound for $STC(u)$, $\forall u \in S$. A similar conclusion holds for the weak version too.*

5. Tukey centrality and quasi-concave functions in graphs

We showed in Section 2 that the strong Tukey centrality satisfies the following property (Lemma 6). If v is an interior vertex of any $u - w$ geodesic in an arbitrary graph G, then $STC(v) \geq \min\{STC(u), STC(w)\}$. We said that strong Tukey centrality is a quasi-concave function in graphs and noted that this property does not hold, in general, for the weak Tukey centrality. We now look at this matter in a more comprehensive way.

Consider a real function f defined on $V(G)$. We say that f is *strongly [weakly] quasi-concave* in G if, for every pair of non-adjacent vertices u, w and for every vertex v that lies in every [at least one] $u - w$ geodesic in G, $f(v) \geq \min\{f(u), f(w)\}$. Note that, if f is quasi-concave, for every pair $u, w \in V(G)$, a $u - w$ geodesic exists such that for all intermediate vertices u', w' and every v' in the geodesic $u' - w'$, $f(v') \geq \min\{f(u'), f(w')\}$. We will say that a $u - w$ geodesic that satisfies this hereditary property is *f-concave.* Thus, if f is *strongly [weakly] quasi-concave* in G, for all pair $u, w \in V(G)$, every [at least one] $u - w$ geodesic is f-concave.

Similarly to the Tukey depth for points in $\mathbb{R}^d$ (see Figure 1, Right), a strong [weak] quasi-concave function f in graph G, defines in $V(G)$ a hierarchical ordering structure of nested strong [weak] convex sets of increasing centrality, $\Sigma_\tau(f) = \{v \in V(G) : f(v) \geq \tau\}$, with τ ranging in the set of values of f, such that $\Sigma_\tau(f) \subset \Sigma_{\tau'}(f)$ if $\tau \geq \tau'$. The graph G can be depicted with the vertices arranged in as many layers as the number of different values of f, with the vertices having values $\min_{v \in V(G)} f(v)$ and $\max_{v \in V(G)} f(v)$ occupying the lowest and highest layers, respectively. Vertices belonging to $\Sigma_{\tau'}(f) \setminus \Sigma_\tau(f)$ are located in the layer immediately above of the layer containing the vertices of $\Sigma_{\tau''}(f) \setminus \Sigma_{\tau'}(f)$, where $\tau > \tau' > \tau''$ are consecutive values of f. In this representation, every [at least one] $u - w$ geodesic does not go under the lowest layer where u or w are located. Moreover, in this setting, the f-concave $u - w$ geodesics have a concave downwards shape, connecting vertices located in layers with indices greater than or equal to $\min\{f(u), f(w)\}$, illustrating the quasi-concave essence of function f in G.

The nested structure of the strong convex sets $\Sigma_\tau(STC)$ can be easily visualized in the graphs of figures 2 - 7 and 9, as sets of vertices depicted with gradient of colors ranging from white to red.

The weak Tukey centrality is strongly quasi-concave in a number of graphs including trees, chordless cycle C_n and the complete graph K_n, but not weakly quasi-concave in general. In the graph of Example 1 (Figure 2, Right), the interior vertices of the geodesics that connect the two topmost vertices with weak Tukey centrality equal to two, have weak centrality one.

Betweenness and closeness are two well known centrality measures based on geodesics [24]. Betweenness counts the number of geodesics through each vertex. Closeness is the inverse of the mean distance from a vertex to all other vertices. Both measures are strongly quasi-concave in certain graphs such as chordless cycles, complete graphs, path graphs and stars. In general, betweenness is not quasi-concave in trees (see Figure 10, Left). For closeness we can prove the

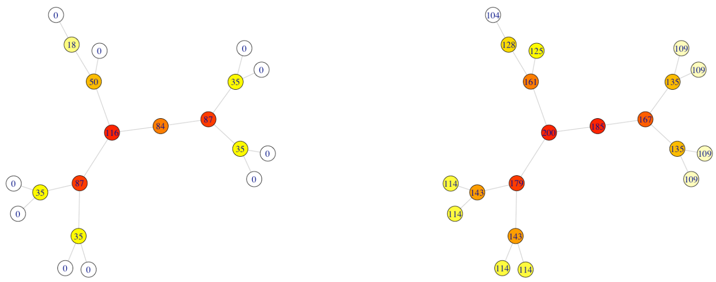

Figure 10: Betweenness centrality values (Left) and closeness centrality values multiplied by 10^4 and rounded to the unity (Right) in a tree with $n = 20$ vertices (the same as Γ_{20} in Figure 5).

following.

Proposition 12. *Closeness centrality is quasi-concave in trees.*

Proof: Let Γ be a tree with n vertices and let $d(u)$ denote the sum of distances from u to the other vertices. A vertex u is a median if $d(u) \leq d(v)$, for every vertex v. Thus, the median vertices are the vertices of maximum closeness centrality. The result now follows directly from the two following facts that are valid for trees [26]. i) A tree has either exactly one median, or exactly two medians joined by an edge. ii) If v is a median of a tree, there are no more than $\lfloor \frac{n}{2} \rfloor$ vertices in any connected component of the forest $\Gamma - v$. Hence, function d increases along any path starting at v toward a leaf vertex, which shows that closeness centrality is quasi-concave in trees. $\square$

The (approximated) closeness centrality values for each one of the 20 vertices of the tree Γ_{20} of Example 4 are shown in Figure 10 (Right).

Another centrality measure, which is related with closeness, is centroid centrality [27]. The centroid of a graph is a vertex that minimizes the maximal distance from other vertices in the graph. Centroid centrality ranks vertices according to their distance to the centroid, and assigns to the centroid the maximum centrality. Since, for trees, a vertex v is a centroid iff v is a median [26], the proof of Property 12 also holds for centroid centrality allowing to conclude that it is a quasi-concave function in trees.

In general, assuming that flow only moves along geodesics, if the centrality function is quasi-concave, it is possible to select intermediates to conduct the flow between any two agents which are at least as central as the less central of the two agents. This may be, in certain contexts and for certain centrality functions, a desirable hierarchical property to achieve. When the property fails, the identification of modules inducing subgraphs where the centrality function is quasi-concave may be considered.

The weak Tukey centrality is not quasi-concave in the Zachary's network. It assigned the highest centrality to the vertices representing the instructor ($WTC(1) = 12$) and the administrator ($WTC(34) = 8$), and values 1, 2 or 3 to all intermediate vertices of all geodesics between these two vertices. This may suggest a clustered structure in the Zachary's network with the ("mode") vertices representing the instructor and the administrator belonging to different modules. When considering the two separate communities, we note that the weak Tukey centrality is strongly quasi-concave in the subgraph induced by the vertices corresponding to the group that followed the instructor (vertices represented by squares in Figure 8), and is weak, but not strongly, quasi-concave in the subgraph induced by the vertices corresponding to members associated with the administrator's faction (represented by circles in Figure 8). If the members corresponding to vertices 28 and 32 would have joined the instructor, the weak Tukey centrality would be strongly quasi-concave in the graphs induced by each of the two resulting communities.

The lemmas below characterize strong and weak quasi-convex functions f defined in graphs G, in terms of the values of f and distances between pairs of vertices in G.

Lemma 13. *f is a strongly quasi-concave function in G iff for every $u, w \in V(G)$, with $w \notin N(u)$, and every $v \in N(u)$ such that $f(v) < \min\{f(u), f(w)\}$, inequality $d(u, w) < d(v, w) + 1$ holds.*

Proof: The condition is clearly necessary. To prove sufficiency suppose that f is not strongly quasi-concave and let P be an $a - b$ geodesic in G that includes an interior vertex where f takes a value less than the minimum between $f(a)$ and $f(b)$. Let v be a vertex in P that has the lower f value, and define u and w to be the two vertices in the $a-v$ geodesic and in the $v-b$ geodesic, respectively, that are at minimum distance from v, and such that $f(u), f(w) > f(v)$. Clearly, these two vertices exist and the geodesic $(u - \cdots - v - \cdots - w)$ has all interior vertices (if there are vertices other than v) with f value equal to $f(v)$. $\square$

Lemma 14. *f is a weakly quasi-concave function in G iff for every $u, w \in V(G)$, with $w \notin N(u)$, there is a vertex $v \in N(u)$ such that $f(v) \geq \min\{f(u), f(w)\}$ and $d(u, w) = d(v, w) + 1$.*

Proof: The condition is clearly necessary and, for $d(u, w) = 3$, it is also clearly sufficient. Suppose $d(u, w) > 3$ and consider a vertex v in the conditions of the lemma. The induction hypothesis ensures the existence of a $v - w$ geodesic, where every vertex v' that lies in the geodesic satisfies $f(v') \geq \min\{f(v), f(w)\}$. Since $f(v) \geq \min\{f(u), f(w)\}$, the geodesic resulting from adding edge $\{u, v\}$ to that geodesic, is a $u - w$ geodesic with all interior vertices having f values no less than $\min\{f(u), f(w)\}$, and the result follows. □

6. Conclusion and future work

In this paper we adjust Tukey depth, which is a well-known data depth function that measures the centrality of an arbitrary point w.r.t. a finite set of points in R^d, to assess the centrality of the vertices of a graph. Tukey centrality scores vertex v to be all the more central as the smaller is a maximum cardinality convex set not including v. In graphs there are two established concepts of convexity. A proper subset S of vertices of a graph G is strong [weak] convex if every pair of vertices of S are connected in the subgraph induced by S by all [at least one of] the geodesics that connect the two vertices in G. Strong and weak convexity give rise to different measures of centrality, which we call strong Tukey centrality and weak Tukey centrality. The former is more strict than the latter as it requires for a set to be convex that all geodesics in the graph between two vertices of the subset remain in the induced subgraph. Thus, strong centrality ranks vertices no lower than weak centrality.

The two centrality measures properly generalize to groups of vertices, and we show that the centrality of a group cannot be less than the centrality of any of its members and strong Tukey centrality of a group cannot exceed the sum of the individual strong centrality of its members. Interestingly these bounds also apply for the group Tukey data depth.

We show that computing Tukey centrality is NP-hard, but closed formulas exist that allow centrality to be easily calculated for certain classes of graphs, including trees. We give 0/1 linear formulations to compute the centrality indices and derive some valid inequalities that may be used to accelerate computation.

Tukey depth is a quasi-concave function, which means that the set of points with Tukey depth greater than or equal to any $\tau \in \mathbb{R}$ is convex. In graphs this reads as the set of vertices with centrality values greater than or equal to $\tau \in \mathbb{R}$ is strongly [weakly] convex. Equivalently, the centrality of all intermediate vertices of every [at least one] geodesic that links a pair of vertices u, w is at least the minimum between the centralities of u and w. We say that a centrality function that satisfies this property in a graph G is strongly [weakly] quasi-concave in G. We show that strong Tukey centrality is strongly quasi-concave in every graph. We also show that weak Tukey centrality and other well-known geodesic based centrality measures (closeness, betweenness and centroid centrality) are weak quasi-concave in certain classes of graphs, but not in general graphs.

We finish by listing some topics that we believe are worth addressing.

- In what concerns group Tukey centrality, it would be interesting to find, among groups of vertices having the same cardinality, which group has the maximum centrality. This gives an indication on how much centrality weight can be achieved by a group with a given number of members. In the Zachary's karate club network we found that the group formed by the instructor and the administrator has strong centrality which is almost equal to the order of the whole network.

- We saw that for certain classes of graphs Tukey centrality can be easily computed. Finding other classes of graphs where Tukey centrality can be calculated in polynomial time, and developing strategies to accelerate the computation of centrality, such as effective valid inequalities and heuristic procedures, would be interesting.

- There are graphs where the weak Tukey centrality is not weakly quasi-concave. In such graphs (e.g., the Zachary's network) we found that there is a large number of vertices with high strong Tukey centrality, which reproduces the achievement of Tukey depth on multi-modal data where, the highest depth values are assigned to a vast interior region. In such cases it may be appropriate to consider a partition of the vertex set in such a way that each cluster induces a subgraph where the weak Tukey is strongly or weakly quasi-concave.

These are issues that we intend to address in future work.

Acknowledgements

All figures were produced using the package igraph [28] of the R Statistical Software (R Core Team, 2020) [29]. To compute strong and weak Tukey centrality we ran CPLEX (CPLEX Optimization Studio V12.7.1) [30] in RStudio (R version 3.4.1) environment.

This research was funded by NOVA MATH and by the Forest Research Centre, research units funded by the Fundação para a Ciência e a Tecnologia, I.P. (FCT), Portugal (UIDB/00297/2020 (NOVA MATH) and UIDB/00239/2020 (CEF)).

References

[1] John W. Tukey. Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), Vol. 2*, pages 523–531, 1975.

[2] Subhajit Dutta, Anil K. Ghosh, and Probal Chaudhuri. Some intriguing properties of tukey's half-space depth. *Bernoulli*, 17(4):1420 – 1434, 2011.

[3] Xiaohui Liu. Fast implementation of the tukey depth. *Computational Statistics*, 32(4):1395–1410, 2017.

[4] David S. Johnson and Franco P. Preparata. The densest hemisphere problem. *Theor. Comput. Sci.*, 6:93–107, 1978.

[5] Regina Y. Liu. On a notion of data depth based on random simplices. *Ann. Stat.*, 18(1):405–414, 1990.

[6] Robert Serfling and Yijun Zuo. General notions of statistical depth function. *The Annals of Statistics*, 28(2):461 – 482, 2000.

[7] M. Everett, S.P. Borgatti, PJ Carington, J Scott, and S Wasserman. *Extending Centrality*, pages 57–75. Cambridge University Press, United Kingdom, 2005.

[8] Juan Wang, Chao Li, and Chengyi Xia. Improved centrality indicators to characterize the nodal spreading capability in complex networks. *Applied Mathematics and Computation*, 334:388–400, 2018.

[9] Frank Harary and Juhani Nieminen. Convexity in graphs. *Journal of Differential Geometry*, 16(2):185 – 190, 1981.

[10] Norbert Polat. On isometric subgraphs of infinite bridged graphs and geodesic convexity. *Discrete Mathematics*, 244(1):399–416, 2002.

[11] Magdalena Lemańska. Weakly convex and convex domination numbers. *Opuscula Mathematica*, 24:181–188, 2004.

[12] Komal Batool and Muaz A. Niazi. Towards a methodology for validation of centrality measures in complex networks. *PLOS ONE*, 9(4):1–14, 2014. URL `https://doi.org/10.1371/journal.pone.0090283`.

[13] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99 (12):7821–7826, 2002.

[14] Wayne W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33(4):452–473, 1977.

[15] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486 (3):75–174, 2010.

[16] R. Karp. Reducibility among combinatorial problems. In R. Miller and J. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972.

[17] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, New York, 1979.

[18] Jože Marinček and Bojan Mohar. On approximating the maximum diameter ratio of graphs. *Discrete Mathematics*, 244(1):323–330, 2002.

[19] M. G. Everett and S. P. Borgatti. The centrality of groups and classes. *Journal of Mathematical Sociology*, 23(3):181–201, 1999.

[20] Jennifer A. Vu and Jill J. Locke. Social network profiles of children in early elementary school classrooms. *Journal of Research in Childhood Education*, 28(1):69–84, 2014.

[21] Yu-Xiang Yao, Zhi-Tong Bing, Liang Huang, Zi-Gang Huang, and Ying-Cheng Lai. A network approach to quantifying radiotherapy effect on cancer: Radiosensitive gene group centrality. *Journal of Theoretical Biology*, 462:528–536, 2019.

[22] Vinko Bandelj, Cosimo Solidoro, Célia Laurent, Stefano Querin, Sara Kaleb, Fabrizio Gianni, and Annalisa Falace. Cross-scale connectivity of macrobenthic communities in a patchy network of habitats: The mesophotic biogenic habitats of the northern adriatic sea. *Estuarine, Coastal and Shelf Science*, 245:106978, 2020.

[23] Takayasu Fushimi, Kazumi Saito, Tetsuo Ikeda, and Kazuhiro Kazama. A new group centrality measure for maximizing the connectedness of network under uncertain connectivity. In Luca Maria Aiello, Chantal Cherifi, Hocine Cherifi, Renaud Lambiotte, Pietro Lió, and Luis M. Rocha, editors, *Complex Networks and Their Applications VII*, pages 3–14, Cham, 2019. Springer International Publishing.

[24] Linton C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 1978.

[25] Jozef Kratica, Vladimir Filipovic, Dragan Matic, and Aleksandar Kartelj. An integer linear programming formulation for the convex dominating set problems. arXiv 1904.02541, 2019.

[26] Bohdan Zelinka. Medians and peripherians of trees. *Archivum Mathematicum*, 004(2):87–95, 1968. URL `http://eudml.org/doc/15833`.

[27] Stephen P. Borgatti and Martin G. Everett. A graph-theoretic perspective on centrality. *Social Networks*, 28(4):466–484, 2006.

[28] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006. URL `https://igraph.org`.

[29] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

[30] Cplex, IBM ILOG. *V12.7: Users Manual for CPLEX*. IBM ILOG CPLEX Division, Incline Village, NV, 2018.