The Strong Maximum Circulation Algorithm: A New Method for Aggregating Preference Rankings

Nathan Atkinson
University of Wisconsin
natkinson@wisc.edu

Scott C. Ganz
Georgetown MSB and AEI
scott.ganz@georgetown.edu

Dorit S. Hochbaum
UC Berkeley
dhochbaum@berkeley.edu

James B. Orlin
MIT Sloan
jorlin@mit.edu

Abstract

We present a new optimization-based method for aggregating preferences in settings where each decision maker, or voter, expresses preferences over pairs of alternatives, or votes. The challenge is to determine a ranking that summarizes the voters' preferences in cases when some of the votes conflict. Only a collection of votes that contains no cycles represents a consensus partial order over alternatives. Our approach to identifying a consensus partial order is motivated by the observation that a collection of votes that form a cycle can be treated as collective ties. The approach is then to remove unions of cycles of votes, or circulations, from the vote graph and determine aggregate preferences from the remainder.

We study the removal of *maximal circulations* attained by any union of cycles the removal of which leaves an acyclic graph. We establish that the removal of a maximal circulation results in non-unique and conflicting partial orders. Moreover, we prove that a *minimum maximal* circulation—i.e., a maximal circulation containing smallest number of votes—is an NP-hard problem.

Instead, we introduce the *strong maximum circulation*, the removal of which guarantees a **unique** outcome in terms of the induced partial order, called the *strong partial order*. The strong partial order includes the aggregate preferences induced by the removal of any maximum circulation. In that sense, conditional on removing a circulation containing as many votes as possible, the strong partial order maximizes the number of induced aggregate preferences over pairs of alternatives. The strong maximum circulation also satisfies strong complementary slackness conditions, and is shown here to be solved efficiently as a network flow problem. We further establish the relationship between the dual of the maximum circulation problem and Kemeny's method, a popular optimization-based approach for preference aggregation.

1 Introduction

The problem of aggregating preferences is of central interest to a broad set of problems in economics, political science, psychology, operations research, and computer science. Given a set of (possibly conflicting) pairwise comparisons of alternatives, how can we determine non-conflicting aggregate pairwise comparisons that best summarize the original data? This question is fundamental to topics as diverse as the analysis of democratic voting systems, the design of online search and recommendation algorithms, and the construction of nonparametric statistical tests.

It is well-known that no preference aggregation method satisfies every potential desirable normative property—e.g., non-dictatorship, independence of irrelevant alternatives, and Pareto efficiency (Arrow, 1951; Pini et al., 2009). Further, existing approaches that are based on maximizing the alignment of an aggregate ranking with a set of pairwise comparisons face issues related to computational complexity that are driven by the exponential size of the set of possible orderings of alternatives (Bartholdi, Tovey and Trick, 1989; Fischer, Hudry and Niedermeier, 2016). As a result, there has been significant interest in applying approaches from optimization and machine learning in order to discover new aggregation methods that are both well-justified normatively and computationally efficient (see, e.g., Agarwal and Lim, 2010).

We introduce a new optimization-based approach for aggregating preferences that is based on the logic of removing cycles of votes, i.e., sets of votes for which every vote in favor of an alternative is counterbalanced by a vote against that alternative. Given a cyclic set of votes, there is no principled way of saying that any alternative is preferred to any other. This insight points to the possibility of treating cycles of votes as uninformative ties and using the non-conflicting remainder to determine an aggregate preference ranking.

We represent a set of voters' pairwise preferences over alternatives, or votes, in a vote graph. The nodes of the vote graph are the alternatives. For each pair of alternatives i and j, we let q_{ij} denote the number of voters who prefer i to j. An arc (i, j) is in the vote graph provided that $q_{ij} > 0$. The collection of votes is said to be *conflicting* if there is a directed cycle in the vote graph. A set of votes is *non-conflicting* if there is no directed cycle in the vote graph. For non-conflicting sets of votes, there is an induced partial order in which alternative i is preferred to alternative j if and only if there is a directed path from i to j in the vote graph. This partial order is the smallest one that is consistent with the votes.

For each pair i, j of alternatives, we refer to the number of votes removed in which i is preferred to j as the flow in (i, j). Our method removes votes that correspond to a flow circulation in the vote graph. That is, for each alternative i, the flow out of i is equal to the flow into i. It is well known that any circulation can be decomposed into flows around cycles (see, for example, Chapter 3 of Ahuja, Magnanti and Orlin (1993)).

A necessary condition for our preference aggregation approach to induce a consensus partial order is that after the removal of a circulation, the remaining votes are non-conflicting; equivalently, the remaining vote graph is acyclic. We refer to such circulations as *maximal*. The arc (i, j) is in the remaining vote graph whenever the maximal circulation sends strictly less than q_{ij} units of flow in (i, j). In this case, we say that there is an *aggregate preference* for i over j following the removal of the circulation. From this partial order, one can also determine a complete ordering of the alternatives (also known as a full ranking) or a weak ordering (also known as a weak ranking) in which ties are permitted.

There are many possible ways of removing maximal circulations from a vote graph, and thus there are many possible acyclic vote graphs that remain after their removal. When considering approaches for removing cycles, we emphasize two important properties. First, the procedure has to result in a unique partial order. Second, the procedure has to be efficient. These properties are especially important in settings where equitable treatment of alternatives is an important consideration. Suppose that there was an optimal partial order in which alternative i was preferred to alternatives j and k and another optimal partial order in which j was preferred to i and k. Then it would not be clear whether the method should report an aggregate preference for i over j or j over i. And, further, in settings with many alternatives or voters, it could be computationally prohibitive to determine whether there is also a third optimal partial order for which k was preferred to i and j.

We first consider removing a *maximum circulation*, which is a circulation containing a maximum total number of votes. We show that this procedure permits the efficient computation of a consensus partial order, but does not result in a unique partial order. We subsequently introduce the *strong maximum circulation* as a maximum circulation whose removal returns as many aggregate preferences as possible. In other words, the remaining vote graph has as many distinct arcs as possible. We refer to the partial order induced by this acyclic graph as a *strong partial order*. The strong partial order satisfies the properties of uniqueness and efficiency that we wanted. In particular, we establish the following:

1. There is a unique strong partial order.

2. The unique strong partial order is the union of the partial orders induced by removing any maximum circulation.

3. One can obtain a strong maximum circulation by solving a single minimum cost flow problem.

4. There are optimality conditions for the strong maximum circulation based on a relaxation of strict complementary slackness conditions for the maximum circulation problem.

A natural alternative approach to removing votes on cycles would be to remove as few votes as possible. We refer to this type of maximal circulation as a *minimum maximal circulations.* Despite its intuitive appeal, this approach does not satisfy the properties that we were seeking. First, we show that it is NP-hard to find a minimum maximal circulation. Second, we show that it does not lead to a unique solution. Worse yet, there may be two minimum maximal circulations that lead to conflicting partial orders. That is, after removing the first minimum maximal circulation from the vote graph, alternative i is preferred to alternative j. After removing the second one, j is preferred to i.

We also relate our cycle removal-based approach to preference aggregation to existing optimization-based approaches, focusing on Kemeny's method (Kemeny, 1959). Kemeny's method is to remove the fewest number of votes so that the remaining set of votes is non-conflicting. While Kemeny's method also has some desirable theoretical properties, e.g., it can be justified in terms of minimizing the swap distance from a set of individual rankings (Fischer, Hudry and Niedermeier, 2016), it also suffers similar drawbacks to the problem of removing a minimum maximal circulation from the vote graph in terms of computational complexity (Gary and Johnson, 1979; Bartholdi, Tovey and Trick, 1989) and the multiplicity of (conflicting) optimal solutions (Muravyov, 2014; Yoo and Escobedo, 2021).

The objective in Kemeny's model appears to be diametrically opposed to the objective of the strong maximum circulation model. Kemeny's method removes as few votes as possible. The strong maximum circulation removes as many votes as possible. Nevertheless, there appears to be an interesting type of duality between the two approaches as revealed by the Hochbaum and Levin (HL) separation-deviation model (see Hochbaum and Levin, 2006) Specifically, Kemeny's model can be transformed into a special case of the HL model with a concave, separable objective function. If this concave objective is approximated by a convex objective, the resulting optimization problem is the dual of the maximum circulation problem. So, the maximum circulation model can be viewed as the dual of the "convexification" of Kemeny's model.

The paper proceeds as follows. In the next section we restate the problem of preference aggregation in terms of the isolation of an acyclic subgraph in a capacitated directed network, or vote graph. In Section 3, we introduce the concept of a strong maximum circulation as a method for eliminating cycles. In Section 4 we provide an efficient algorithm for finding a strong maximum circulation. In Section 5, we describe our method in terms of a convex relaxation of Kemeny's model and demonstrate its relationship to Hochbaum and Levin (2006). Finally, we consider the removal of minimum maximal circulations in Section 6. We show that an aggregation approach based on eliminating minimum maximal circulations faces many of the same computational and normative problems as Kemeny's method.

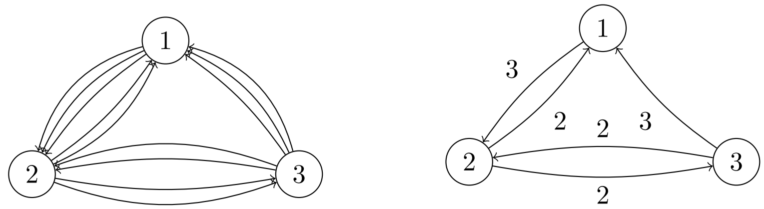
2 Background and Preliminaries

Vote graph, circulations, and acyclic graphs. A vote graph is a directed graph $G = (V, A)$ with the set of alternatives V represented by the nodes (or vertices) of the graph. Let $n = |V|$ and $m = |A|$. Each arc $(i, j) \in A$ represents the (strict) preference of i over j by at least one voter. If there are $q_{ij} \geq 1$ voters that express the same preference (i, j), we represent it by associating a capacity q_{ij} with arc (i, j). Alternatively, one could represent the vote graph as a *multi-graph* with q_{ij} arcs going from i to j. An example of a vote graph represented as multi-graph is given in Figure 1a, and the same vote graph represented as a weighted, or capacitated, simple graph is given in Figure 1b. Note that it is possible for both q_{ij} and q_{ji} to be positive, meaning that there are voters that prefer i to j and voters that prefer j to i. In that, case we have $\min\{q_{ij}, q_{ji}\}$ 2-cycles of votes that contradict each other. A directed graph $G = (V, A)$ is called *acyclic* if there are no directed cycles in G.

We denote vectors by boldface characters. Let the vector of capacities of the arcs of A in the vote graph be $\mathbf{q}$, where $\mathbf{q} \in \mathbb{R}^m_+$. The vector $\mathbf{q}$ is initially non-negative and integer, representing the count of the votes with each preference over pairs of alternatives. However, our method may leave a fractional number of votes. Accordingly, we permit capacities to be non-integral.

An alternative, compact representation of the vote graph is denoted by $G[V, \mathbf{q}]$ with the interpretation that this graph is $G = (V, A)$ where $A = A(\mathbf{q})$ defined as $A(\mathbf{q}) = \{(i, j) : q_{ij} > 0\}$.

A (flow) vector $\mathbf{x} \in \mathbb{R}^m_+$ is said to be a *circulation* in $G[V, \mathbf{q}]$ if it satisfies that $0 \leq x_{ij} \leq q_{ij}$ for all $(i, j) \in A(\mathbf{q})$, and $\sum_{j:(i,j)\in A} x_{ij} - \sum_{j:(j,i)\in A} x_{ji} = 0$ for all $i \in V$. That is, the incoming flow into each node is equal to the outgoing flow from that node.



(a) (b)

Figure 1: The representation of the vote graph as a: (a) multi-graph, (b) capacitated graph.

A circulation $\mathbf{x}'$ is *maximal* if there is no circulation vector $\mathbf{x}'' \neq \mathbf{x}'$ such that $\mathbf{x}'' \geq \mathbf{x}'$. We observe that,

Observation 1. *For a maximal circulation $\mathbf{x}'$ in $G[V, \mathbf{q}]$, the graph $G[V, \mathbf{q} - \mathbf{x}']$ is acyclic. Equivalently, $G' = (V, A(\mathbf{q} - \mathbf{x}'))$ is acyclic.*

The observation means that the removal of a maximal circulation leaves a graph that is acyclic. This follows since if G' contains a cycle, then $\mathbf{x}'$ can be appended by that cycle, contradicting its being maximal.

We next define three specific types of maximal circulations: (1) minimum maximal, (2) maximum, and (3) strong maximum.

1. A *minimum maximal circulation* is a maximal circulation that contains the smallest number of arcs counted with their multiplicity. That is, it is a circulation $\mathbf{x}'$ that minimizes $\sum_{(i,j)\in A} x_{ij}$ among all maximal circulations.

2. A *maximum circulation* is a circulation that contains the largest number of arcs counted with their multiplicity. That is, it is a circulation $\mathbf{x}'$ that solves the following:

$$\begin{aligned} & \max \sum_{(i,j)\in A} x_{ij} && (1) \\ \text{(Maximum Circulation)} \quad & \text{subject to} \sum_{j:(i,j)\in A} x_{ij} - \sum_{j:(j,i)\in A} x_{ji} = 0 \quad \forall i \in V \\ & 0 \leq x_{ij} \leq q_{ij} \quad \forall (i,j) \in A. \end{aligned}$$

   By construction, all maximum circulations are also maximal.

3. A *strong maximum circulation*, $\mathbf{x}^*$, is a maximum circulation for which the minimum number of arcs in G are at capacity. That is, if for any other maximum circulation $\mathbf{x}'$, $|A(\mathbf{q} - \mathbf{x}^*)| \geq |A(\mathbf{q} - \mathbf{x}')|$.

Topological order, transitive closure, and partial order. A graph is acyclic if and only if it admits a *topological ordering*. A topological ordering is an assignment of distinct labels in the set $\{1, 2, \ldots, n\}$ to the nodes so that for every arc $(i, j) \in A$, $i < j$. Given an acyclic graph, the time to find a topological order is $O(m + n)$. In general, there are multiple (perhaps exponentially many) topological orders of an acyclic graph.

For an acyclic graph $G = (V, A(\mathbf{q}))$, the *transitive closure* of G contains all pairs (i, j) such that there is a directed path in G from i to j. We denote the arc set of the transitive closure as $A^{TC}(\mathbf{q})$ and denote the transitive closure of G as $G^{TC} = (V, A^{TC}(\mathbf{q}))$. When $\mathbf{q}$ is clear from context, we refer to the transitive closure compactly as A^{TC}.

A *partial order* is a transitive and antisymmetric relation among the elements of a set. We denote the relation using the symbol "$\succ$." Associated with each acyclic graph $G = (V, A)$ is a *partial order*, obtained as follows: for elements $i, j \in V$, $i \succ j$ if and only if $(i, j) \in A^{TC}$.

If G is acyclic, we sometimes use A^{TC} to represent both the transitive closure and the corresponding partial order. Let $A^{TC}(\mathbf{q} - \mathbf{x}')$ represent the transitive closure of the arcs that remain following the elimination of maximal circulation $\mathbf{x}'$.

Partial orders P and Q defined on the same set of elements are said to *conflict* if there are elements i and j such that $i \succ_P j$ and $j \succ_Q i$. Two partial orders defined on the same set of elements are said to be *non-conflicting* if they do not conflict.

Partial orders and rankings. A partial order is obtained by determining the transitive closure of the arcs of an acyclic graph. A partial order is transitive and antisymmetric, but not necessarily complete. In some cases it may be desirable to construct a ranking from the partial order so that every alternative is comparable to every other alternative. We say that a ranking is *strict* if ties are not permitted and *weak* if ties are permitted. Constructing a ranking from a partial order imposes additional structure that is not initially present. Indeed, there may be an exponential number of rankings consistent with a partial order (e.g. every topological order that is consistent with the partial order). A natural choice of (weak) ranking is to assign each alternative the most favorable possible rank that is consistent with the strong partial order. This is equivalent to assigning a rank equal to the length of the longest path to i in the acyclic subgraph (where lower ranks are better).

Strong maximum circulations, strong partial order, strong arcs, and strong subgraph. We will show that for any two strong maximum circulations $\mathbf{x}'$ and $\mathbf{x}''$, $A(\mathbf{q} - \mathbf{x}') = A(\mathbf{q} - \mathbf{x}'')$, and thus $A^{TC}(\mathbf{q} - \mathbf{x}') = A^{TC}(\mathbf{q} - \mathbf{x}'')$. That is, the set of arcs that remains following the elimination of a strong maximum circulation $\mathbf{x}^*$ is independent of the strong maximum circulation.

We refer to $A^{TC}(\mathbf{q} - \mathbf{x}^*)$ as the *strong partial order*, and denote it more simply as A^{SP}. We say that an arc (i, j) is *strong* if there is a maximum circulation $\mathbf{x}'$ such that $\mathbf{x}'_{ij} < \mathbf{q}_{ij}$, and thus $(i, j) \in A(\mathbf{q} - \mathbf{x}')$. The *strong subgraph* of A is the set $A^* = \{(i, j) : (i, j) \text{ is strong}\}$. We will show that the transitive closure of A^* is equal to A^{SP}.

Network Flow. We next introduce some well-known results from the network flow literature that will be used in the computational analysis of our algorithm. A more detailed treatment is available in Ahuja, Magnanti and Orlin (1993).

The minimum cost network flow problem (MCNF). The minimum cost network flow problem is defined on a directed graph $G = (V, A)$ where every arc $(i, j) \in A$ has a nonnegative arc capacity q_{ij} and cost c_{ij} per unit flow. Every node $i \in V$ has a *supply* value b_i associated with it that could be positive, negative or zero, and such that $\sum_{i \in V} b_i = 0$. A vector $\mathbf{x} \in \mathbb{R}^m_+$ is said to be a *feasible flow* if $0 \le x_{ij} \le q_{ij}$ for all $(i, j) \in A(\mathbf{q})$, and for all $i \in V$, $\sum_{j:(i,j)\in A} x_{ij} - \sum_{j:(j,i)\in A} x_{ji} = b_i$. The formulation of MCNF is given in (2):

$$
\begin{aligned}
& & \min & \sum_{(i,j)\in A} c_{ij} x_{ij} & & (2)\\
& \text{(MCNF)} & \text{subject to} & \sum_{j:(i,j)\in A} x_{ij} - \sum_{j:(j,i)\in A} x_{ji} = b_i \quad \forall i \in V & & \\
& & & 0 \le x_{ij} \le q_{ij} \quad \forall (i,j) \in A. & &
\end{aligned}
$$

The maximum circulation problem is a special case of MCNF in which $b_i = 0$ for all $i \in V$, the minimization is replaced by maximization, and each cost is equal to 1. (This is equivalent to a minimization problem with costs of -1.)

Complexity of solving MCNF and maximum circulation. Using the successive approximation algorithm of Goldberg and Tarjan (1987), the MCNF problem is solved in running time $O(nm \log n \log(nC))$, where $n = |V|$ $m = |A|$ and $C = \max_{(i,j)\in A} |c_{ij}|$. Goldberg and Tarjan also provide the running time bound of $O(n^{5/3} m^{2/3} \log nC)$, which improves on $O(nm \log n \log(nC))$ when $\frac{n^2}{m} < \log^3 n$.

For the maximum circulation problem, $C = 1$. Therefore the running time for solving the maximum circulation is $O(nm \log^2 n)$.

3 Maximum Circulations and Strong Maximum Circulations

In this section, we analyze preference aggregation methods that are based on eliminating maximum circulations. Specifically, the section includes the following results:

1. For A^*, a strong subgraph, if $\mathbf{x}^*$ is a strong maximum circulation, then $A(\mathbf{q}-\mathbf{x}^*) = A^*$, (Subsection 3.1). It follows that A^* is **unique**.
2. The strong subgraph A^* is acyclic.
3. An algorithm that computes a strong maximum circulation by solving $m+1$ maximum circulation problems. (Subsection 3.2).

3.1 Generating a strong maximum circulation based on the acyclic set A^*

Recall that the strong subgraph A^* includes each arc that is not at capacity in some maximum circulation. Here we show how to obtain a strong maximum circulation based on A^*. We also establish that A^* is acyclic.

We first start with the special case in which the vote graph is Eulerian.

The vote graph is *Eulerian* if $\sum_{j:(i,j)\in A} q_{ij} = \sum_{j:(j,i)\in A} q_{ji}$. The following lemma is a well known result about Eulerian networks.

Lemma 1. *If the vote graph $G = (V, A(\mathbf{q}))$ is Eulerian, then the flow $\mathbf{x}^* = \mathbf{q}$ is the unique maximum circulation, and $A^* = \emptyset$. If the vote graph is not Eulerian, then $\mathbf{x}^* = \mathbf{q}$ is not a circulation, and $A^* \neq \emptyset$.*

Corollary 1. *Let A^* denote the strong subgraph of the vote graph $G = (V, A(\mathbf{q}))$. Then $A^* = \emptyset$ if and only if the vote graph is Eulerian.*

Lemma 2. *Let $G = (V, A(\mathbf{q}))$ be a non-Eulerian vote graph. Let A^* denote the strong subgraph of G. For each arc $a \in A^*$, let $\mathbf{x}^a$ be a maximum circulation for which the flow in arc a is less than q_a. Let $\mathbf{x}^* = \frac{\sum_{a\in A^*} \mathbf{x}^a}{|A^*|}$. Then*

1. *$A^* = A(q - \mathbf{x}^*)$;*
2. *$\mathbf{x}^*$ is a strong maximum circulation, and;*
3. *A^* is acyclic.*

Proof. $\mathbf{x}^*$ is the convex combination of maximum circulations, and is thus a maximum circulation. No arc of A^* is at its upper bound in $\mathbf{x}^*$. Therefore, $A^* \subseteq A(q - \mathbf{x}^*)$. Every arc in $A \setminus A^*$ is at its upper bound in every maximum circulation. Therefore, $(A \setminus A^*) \cap A(q - \mathbf{x}^*) = \emptyset$. Hence, $A(q - \mathbf{x}^*) \subseteq A^*$. It follows that $A^* = A(q - \mathbf{x}^*)$. Since $\mathbf{x}^*$ is a maximum circulation, $A(q - \mathbf{x}^*)$ must be acyclic. And since $A^* = A(q - \mathbf{x}^*)$ it follows that $\mathbf{x}^*$ is a strong maximum circulation. □

We now show that the arc set A^* is equal to the set of arcs remaining after removing any strong maximum circulation from the graph. It is therefore unique.

Corollary 2. *A maximum circulation $\mathbf{x}'$ in the vote graph $G = (V, A(\mathbf{q}))$ is strong if and only if $A(\mathbf{q}-\mathbf{x}') = A^*$.*

Proof. Let x^* be defined as in Lemma 2. Suppose that $\mathbf{x}'$ is a strong maximum circulation. It follows that $|A(\mathbf{q} - \mathbf{x}')| = |A^*| = |A(\mathbf{q} - \mathbf{x}^*)|$. Moreover, $A(\mathbf{q} - \mathbf{x}') \subseteq A^*$. Therefore, $A(\mathbf{q} - \mathbf{x}') = A^*$. □

Corollary 3. *For any two maximum circulations $\mathbf{x}'$ and $\mathbf{x}''$, there the partial orders induced by $A(\mathbf{q} - \mathbf{x}')$ and $A(\mathbf{q} - \mathbf{x}'')$ are non-conflicting.*

Proof. Suppose that for two maximum circulations $\mathbf{x}'$ and $\mathbf{x}''$ there is a conflict. Therefore there is a pair i and j such that $i \succ_{\mathbf{x}'} j$ and $j \succ_{\mathbf{x}''} i$. By definition, A^* contains both arcs (i, j) and (j, i) which contradicts that A^* is acyclic. □

Recall that A^{SP} is the strong partial order, which is the partial order induced by A^*, or equivalently by $A(\mathbf{q} - \mathbf{x}')$, where $\mathbf{x}'$ is any strong maximum circulation.

3.2 Constructing a Strong Maximum Circulation

The development of the strong maximum circulation $\mathbf{x}^*$ in Lemma 2 is not a constructive algorithm because it does not construct the flows x^a for $a \in A^*$. The next algorithm efficiently constructs these flows one arc a at a time checking whether a maximum circulation flow has the flow on the arc strictly less than q_a. For simplicity, it is assumed that all data is integral. However, the algorithm can be easily modified to work for data that is not integral.

Algorithm 1:

01. $z^* :=$ optimum objective value for the maximum circulation problem (1).
02. $A^* := \emptyset$
03. **for each** arc $(i, j) \in A$ **do**;
04. let z_{ij} be the optimal objective value for the maximum circulation problem
. with the added constraint that $x_{ij} \leq q_{ij} - 1$.[1]
05. **if** $z_{ij} = z^*$, **then** $A^* := A^* \cup \{(i, j)\}$
06. **endfor**
07. A^{SP} is the transitive closure of A^*

If the data are not integral, then to determine A^*, the constraint in line 04 of Algorithm 1 would be modified to "$x_{ij} \leq q_{ij} - \epsilon$", where ϵ is chosen to be sufficiently close to 0. Additionally, note that the algorithm may deliver A^* empty if the vote graph is Eulerian.

Algorithm 1 determines A^* by solving $m + 1$ maximum circulation problems, including one for determining the optimal objective value. Each maximum circulation problem is solvable in $O(nm \log^2 n)$ time. In Section 4, we show how to determine A^* by solving just one minimum cost flow problem.

Using Lemma 2, the output of Algorithm 1 can deliver a strong maximum circulation. For each $a \in A^*$, let $\mathbf{x}^a$ be the maximum circulation found by the algorithm in which the flow in a is less than its capacity. Assuming that the vote graph is not Eulerian, and thus $A^* \neq \emptyset$, the average of these $|A^*|$ maximum circulations, $\mathbf{x}^* = \frac{\sum_{a \in A^*} \mathbf{x}^a}{|A^*|}$, is a strong maximum circulation.

Even though A^* and A^{SP} are uniquely determined, there may be an infinite number of different strong maximum circulations. For example, consider any flow $\mathbf{x}'$ obtained as follows:

$$\mathbf{x}' = \sum_{a \in A^*} \lambda_a \mathbf{x}^a. \tag{3}$$

where $\{\lambda_a\}_{a \in A^*}$ is any strictly positive vector whose components sum to 1. Then $\mathbf{x}'$ is a strong maximum circulation.

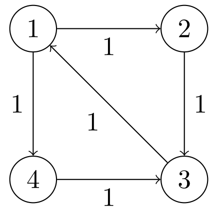
3.2.1 An Illustrative Example

We next illustrate the construction of the strong partial order and the strong maximum circulation for a graph $G[V, \mathbf{q}]$. Recall that the arcs in the graph G are $A = A(\mathbf{q}) = \{(i, j) : i, j \in V \text{and } q_{ij} > 0\}$. We represent the capacities vector $\mathbf{q}$ as an $n \times n$ matrix where only the positive entries correspond to the arcs of $G = [V, \mathbf{q}]$.

[1] Subtracting 1 works because the vector $\mathbf{q}$ is integral, and there is always an integer optimum flow if all data is integral.

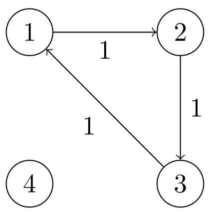
$$q = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$G[V, \mathbf{q}] =$

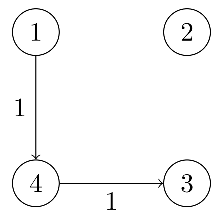


In this graph there are only 5 arcs and two integer-valued maximum circulations $\mathbf{x}'$ and $\mathbf{x}''$, each of value 3. The circulation $\mathbf{x}'$ is obtained by sending one unit of flow on the cycle 1-2-3-1; the circulation $\mathbf{x}''$ is obtained by sending one unit of flow on the cycle 1-4-3-1. This circulations and the remaining (residual flows) graphs, $G[V, \mathbf{q} - \mathbf{x}']$ and $G[V, \mathbf{q} - \mathbf{x}'']$ are:

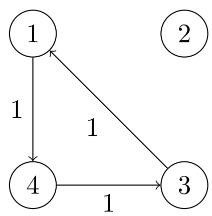
$G[V, \mathbf{x}'] =$



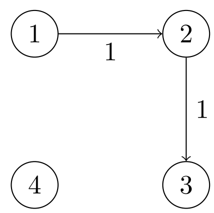
$G[V, \mathbf{q} - \mathbf{x}'] =$



$G[V, \mathbf{x}''] =$

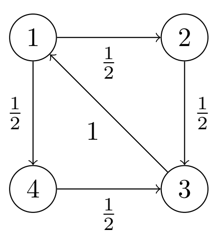


$G[V, \mathbf{q} - \mathbf{x}''] =$

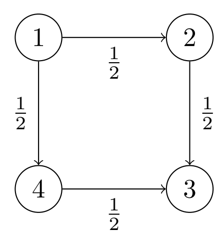


Consider next the circulation $\mathbf{x}^* = (\mathbf{x} + \mathbf{y})/2$ and the respective remaining graph after removing $\mathbf{x}$, $G[V, \mathbf{q} - \mathbf{x}^*]$:

$G[V, \mathbf{x}^*] =$

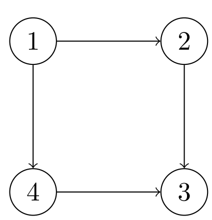


$G[V, \mathbf{q} - \mathbf{x}^*] =$

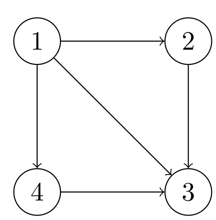


Note that the graph $G[V, \mathbf{q} - \mathbf{x}^*]$ is acyclic, and therefore the set of arcs $A(\mathbf{q} - \mathbf{x}^*)$ form an acylic graph. That follows since $\mathbf{x}^*$ is a maximal circulation. Note, however, that whereas the number of arcs in $A(\mathbf{q} - \mathbf{x}')$ is 2 and the number of arcs in $A(\mathbf{q} - \mathbf{x}'')$ is 2, the number of arcs in $A(\mathbf{q} - \mathbf{x}^*)$ is 4 and hence strictly greater than the number of arcs in the acyclic graph induced by the integer valued circulations. Following Lemma 2, $A(\mathbf{q} - \mathbf{x}^*) = A(\mathbf{q} - \mathbf{x}') \cup A(\mathbf{q} - \mathbf{x}'')$. Finally, by adding the transitive closure arcs to $A(\mathbf{q} - \mathbf{x}^*)$, we get the strong partial order A^{SP}.

$A(\mathbf{q} - \mathbf{x}^*) =$



$A^{SP}(\mathbf{q}) =$



4 Computing a Strong Circulation with One Maximum Circulation Flow

Algorithm 1, described in Section 3, computes the acyclic graph A^*, which in turn can be used to construct a strong maximum circulation and a strong partial order. Algorithm 1 requires solving at most $m+1$ maximum circulation problems. In this section, we introduce the *perturbation algorithm*, which determines A^* by solving a single minimum cost network flow (MCNF) problem. Specifically, the preturbation algorithm returns a strong maximum circulation, which is then used to find A^* by removing the arcs at capacity.

The perturbation algorithm relies on the optimality conditions that are shown to characterize strong maximum circulations—the *strong complementary slackness conditions*, as defined in Subsection 4.1. The formulation of the associated MCNF problem and the perturbation algorithm are subsequently given in Subsection 4.2.

4.1 Strong (Rather than Strict) Complementary Slackness

We define here the "strong complementary slackness" conditions that are proved to characterise strong maximum circulations. These conditions are more restrictive than the usual complementary slackness conditions, but less restrictive than the conditions that are known as strict complementary slackness. We will subsequently show that the strong maximum circulation derived from the MCNF solution satisfies strong complementary slackness.

Among all optimal solutions for the maximum circulation problem 1, we want a strong maximum circulation; that is, we want a maximum circulation with as many arcs as possible with flow less than than their capacity. Suppose that we were given a circulation $\mathbf{x}'$. One can identify that $\mathbf{x}'$ is maximum using a form of LP duality such as the complementary slackness conditions. It is shown next that one can identify that $\mathbf{x}'$ is strong if it satisfies the "strong complementary slackness conditions."

The dual of the maximum circulation problem (1) is:

$$\begin{aligned} \min \sum_{(i,j)\in A} q_{ij} z_{ij} \\ \text{subject to } y_i - y_j + z_{ij} \geq 1 \ \forall (i,j) \in A \\ z_{ij} \geq 0 \ \forall (i,j) \in A. \end{aligned} \tag{4}$$

Suppose that $\mathbf{x}'$ is a feasible circulation for (1), and that $(\mathbf{y}', \mathbf{z}')$ is feasible for (**??**. We will express the strong optimality conditions based entirely on the vectors $\mathbf{x}'$ and $\mathbf{y}'$. We may do so without loss of generality because we assume that for all $(i,j) \in A$, $z'_{ij} = \max\{y'_j - y'_i + 1, 0\}$. These conditions are satisfied by any optimal solution to (4).

We use the notation $\mathbf{c}$ for the vector of cost coefficients in a MCNF problem, which for the primal maximum circulation problem (1) satisfy $c_{ij} = 1$ for all $(i,j) \in A$. We say that $\mathbf{x}'$ and $\mathbf{y}'$ satisfy *strong complementary slackness* for the MCNF problem if the following is true:

1. For each $(i,j) \in A$, if $x'_{ij} = 0$, then $c_{ij} - y'_i + y'_j \leq 0$.
2. For each $(i,j) \in A$, if $0 < x'_{ij} < q_{ij}$, then $c_{ij} - y'_i + y'_j = 0$.
3. For each $(i,j) \in A$, if $x'_{ij} = q_{ij}$, then $c_{ij} - y'_i + y'_j > 0$.

The above conditions are more restrictive than the standard complementary slackness conditions because of (3). The usual complementary slackness condition is: "if $x'_{ij} = q_{ij}$, then $c_{ij} - y'_i + y'_j \geq 0$". The above conditions are also less restrictive than *strict* complementary slackness conditions, which require, in addition to (2) and (3), to replace (1) by: "For each $(i,j) \in A$, if $x'_{ij} = 0$, then $c_{ij} - y'_i + y'_j < 0$".

If $\mathbf{x}'$ and $\mathbf{y}'$ satisfy strong complementary slackness, then $\mathbf{y}'$ "certifies" that $\mathbf{x}'$ is a strong maximum circulation, as stated in the following lemma.

Theorem 1. *Suppose that* $\mathbf{x}'$ *is a feasible circulation for (1), and that* $\mathbf{y}'$ *is feasible for (4)—i.e., the dual of (1). If* $\mathbf{x}'$ *and* $\mathbf{y}'$ *satisfy strong complementary slackness, then* $\mathbf{x}'$ *is a strong maximum circulation.*

Proof. Assume that $\mathbf{x}'$ and $\mathbf{y}'$ satisfy strong complementary slackness. Then they also satisfy complementary slackness, and are thus optimal for the respective primal (1) and dual (4) problems. Therefore $\mathbf{x}'$ must be a maximum circulation and $\mathbf{y}'$ is a dual optimal solution. To prove that $\mathbf{x}'$ is a strong maximum circulation, we show next that for any other maximum circulation $\mathbf{x}''$, if $x'_{ij} = q_{ij}$ for any arc $(i,j) \in A$, then $x''_{ij} = q_{ij}$.

By the strong complementary slackness satisfied by $\mathbf{x}'$ and $\mathbf{y}'$: if $x'_{ij} = q_{ij}$, then $c_{ij} - y'_i + y'_j > 0$. Since $\mathbf{x}''$ is a maximum circulation and $\mathbf{y}'$ is a dual optimal solution, it follows that $\mathbf{x}''$ and $\mathbf{y}'$ satisfy complementary slackness. Therefore $c_{ij} - y'_i + y'_j > 0$, implies that $x''_{ij} = q_{ij}$. □

4.2 The "Perturbation" Algorithm

We present here the "Perturbation" algorithm that finds a strong maximum circulation $\mathbf{x}^*$ as well as the set A^* and the induced strong partial order A^{SP} by solving a single maximum (or minimum) cost flow problem derived from a "perturbation" of the maximum circulation problem. That solution also produces a dual vector $\mathbf{y}^*$ such that $\mathbf{x}^*$ and $\mathbf{y}^*$ satisfy strong complementary slackness.

The perturbation algorithm relies on solving the following maximum utility (negative cost) flow problem with two sets of decision variables $\mathbf{w}, \mathbf{v} \in \mathbb{R}^m$. Recall that $m = |A| = |A(\mathbf{q})|$, the number of arcs in the vote graph $G = (V, A)$. Our network flow formulation (5) is defined on a graph $G = (V, A^{(2)})$ where $A^{(2)}$ contains 2 copies of each arc $(i,j) \in A$. The flow in the first copy of (i,j) will be denoted as w_{ij} with capacity upper bound $q_{ij} - \epsilon$, and the flow in the second copy of (i,j) will be denoted as v_{ij} with capacity upper bound ϵ. If q_{ij} is integral for all arcs $(i,j) \in A$, then we can choose $\epsilon = 1/(m+1)$. Otherwise, we can choose $\epsilon = 1/((m+1) \cdot LCD)$, where LCD is the least common denominator of the values in the vector q.

We let $c_{ij} = 1$ for all arcs $(i,j) \in A$.

$$
\begin{aligned}
\max \quad & f(\mathbf{w}, \mathbf{v}) = \sum_{(i,j)\in A} c_{ij}(w_{ij} + v_{ij}) - \frac{1}{m+1} \sum_{(i,j)\in A} v_{ij} \\
\text{subject to} \quad & \sum_{j:(i,j)\in A} (w_{ij} + v_{ij}) - \sum_{j:(j,i)\in A} (w_{ji} + v_{ji}) = 0 \quad \forall i \in V, \\
& 0 \le w_{ij} \le q_{ij} - \epsilon \quad \text{for } (i,j) \in A, \\
& 0 \le v_{ij} \le \epsilon \quad \text{for } (i,j) \in A.
\end{aligned}
\tag{5}
$$

We note that model 5 is a maximum cost flow problem. It can be converted to a minimum cost flow problem by replacing the objective function by its negative.

The perturbation algorithm consists of first finding an optimal solution $(\mathbf{w}', \mathbf{v}')$ for problem (5), and then finding an optimal solution $(\mathbf{y}', \mathbf{z}')$ for the dual of (5). The complexity of finding an optimal solution to MCNF, using the successive approximation algorithm of Goldberg and Tarjan (1987), is $O(nm \log(n^2/m) \log(nC))$, where $n = |V|$ $m = |A|$ and $C = \max_{(i,j)\in A} |c_{ij}|$. Here we scale the objective function by $m+1$ so all coefficients are integer, and $C = m+1$. Observing that $m = O(n^2)$ the complexity of solving (5) is $O(nm \log(n^2/m) \log n)$ – the same complexity as that of solving one maximum circulation problem. To derive the optimal dual solution, the standard procedure is to construct a graph where the arcs are all the residual arcs with respect to the optimal flow, with their respective costs, and to add a source node with arcs of length 0 to all nodes in the graph. In this constructed graph the length of the shortest paths from the source node, correspond to the optimal dual values. Hence, the optimal dual solution $\mathbf{y}'$ can be obtained in $O(mn)$ time, using, e.g., the Bellman-Ford algorithm. This time to determine the optimal dual solution is dominated by the run time required to find the optimal flow.

The perturbation algorithm outputs $\mathbf{x}' = \mathbf{w}' + \mathbf{v}'$. We show next that $\mathbf{x}'$ is a strong maximum circulation and that $(\mathbf{x}', \mathbf{y}')$ satisfies strong complementary slackness for the maximum circulation problem.

Lemma 3. *Let $(\mathbf{w}', \mathbf{v}')$ be an optimal basic solution for the maximum cost network flow problem (5) and $(\mathbf{y}', \mathbf{z}')$ optimal for the dual linear program. Then $\mathbf{x}' = \mathbf{w}' + \mathbf{v}'$ is a strong maximum circulation and $A' = \{(i,j) : v'_{ij} = 0\}$ satisfies $A' = A^*$. Moreover, $\mathbf{x}'$ and $\mathbf{y}'$ satisfy strong complementary slackness for the maximum circulation problem (1).*

Proof. We first consider a solution to (5) that is derived from a strong maximum circulation. Let $\mathbf{x}^*$ be the strong maximum circulation calculated by averaging the flows from Algorithm 1. Let $A^* = A(q - \mathbf{x}^*)$ denote the arcs whose flow are not at capacity in $\mathbf{x}^*$. By Lemma 2, A^* is the set of strong arcs. Let $K = |A^*|$. Thus $K < m$. Note that $\mathbf{x}^*$ is obtained as the average of K basic flows. Assuming that all capacities are integral, for each arc $(i,j) \in A^*$, $x^*_{ij} \leq q_{ij} - \frac{1}{K} < q_{ij} - \epsilon$.

Let $(\mathbf{w}^*, \mathbf{v}^*)$ be derived from $\mathbf{x}^*$ as follows. If $x^*_{ij} < q_{ij} - \epsilon$, then $w^*_{ij} = x^*_{ij}$ and $v^*_{ij} = 0$. If $x^*_{ij} = q_{ij}$, then $w^*_{ij} = x^*_{ij} - \epsilon$ and $v^*_{ij} = \epsilon$. Note that $(\mathbf{w}^*, \mathbf{v}^*)$ is feasible for the maximum cost flow problem (5). Let η^* be the optimal objective for the maximum circulation problem. Then $f(\mathbf{w}^*, \mathbf{v}^*) = \eta^* - \frac{(m-K)\epsilon}{m+1} > \eta^* - \epsilon$.

We next establish the optimality of $\mathbf{x}'$ for (1). Note that $f(\mathbf{w}', \mathbf{v}') \geq f(\mathbf{w}^*, \mathbf{v}^*) > \eta^* - \epsilon$. Therefore, $\mathbf{c} \cdot \mathbf{x}' > \eta^* - \epsilon$. Since $(\mathbf{w}', \mathbf{v}')$ is a basic solution, it follows that all flows are integral multiples of $1/\epsilon$, and thus $\mathbf{c} \cdot \mathbf{x}' \geq \eta^*$, implying that $\mathbf{x}'$ is optimal for the maximum circulation problem.

We next show that $A' = A^*$, and thus, according to Corollary 2, $\mathbf{x}'$ is a strong maximum circulation. Since $f(\mathbf{w}', \mathbf{v}') = \eta^* - \frac{(m-|A'|)\epsilon}{m+1} \geq f(\mathbf{w}^*, \mathbf{v}^*) = \eta^* - \frac{(m-|A^*|)\epsilon}{m+1}$, it follows that $|A'| \geq |A^*|$. And because $\mathbf{x}'$ is a maximum circulation and $\mathbf{x}^*$ is a strong maximum circulation, it follows that $A' \subseteq A^*$. We conclude that $A' = A^*$, and therefore $\mathbf{x}'$ is a strong maximum circulation.

Because $A' = A^*$, it follows that $f(\mathbf{w}^*, \mathbf{v}^*) = f(\mathbf{w}', \mathbf{v}')$. Therefore, $(\mathbf{w}^*, \mathbf{v}^*)$ is optimal for (5).

We next show that $\mathbf{x}^*$ and $\mathbf{y}'$ satisfy strong complementary slackness conditions for (1). We see why as follows. Because $(\mathbf{w}^*, \mathbf{v}^*)$ is optimal for (5), $(\mathbf{w}^*, \mathbf{v}^*)$ and $\mathbf{y}'$ satisfy complementary slackness conditions for (5). Therefore, the following is true:

1. For each $(i,j) \in A$, if $x^*_{ij} = 0$, then $w^*_{ij} = 0$, and $c_{ij} - y'_i + y'_j \leq 0$.
2. For each $(i,j) \in A$, if $0 < x^*_{ij} < q_{ij}$, then $0 < w^*_{ij} < q_{ij} - \epsilon$, and $c_{ij} - y'_i + y'_j = 0$.
3. For each $(i,j) \in A$, if $x^*_{ij} = q_{ij}$, then $v^*_{ij} = \epsilon$, and $(c_{ij} - \frac{1}{m+1}) - y'_i + y'_j \geq 0$, which implies that $c_{ij} - y'_i + y'_j > 0$.

Thus, $\mathbf{x}^*$ and $\mathbf{y}'$ satisfy strong complementary slackness.

Finally, we show that $\mathbf{x}'$ and $\mathbf{y}'$ satisfy strong complementary slackness, which will complete the proof of this lemma: Clearly, $\mathbf{x}'$ and $\mathbf{y}'$ satisfy the first and third conditions of strong complementary slackness. As for the second condition,

$$\text{If } 0 < x'_{ij} < q_{ij}, \text{ then } 0 < x^*_{ij} < q_{ij} - \epsilon, \text{ and } c_{ij} - y'_i + y'_j = 0.$$

□

5 Dual Optimal Rankings and the Hochbaum-Levin Model

We now discuss rankings that are consistent with the dual linear program of the maximum circulation problem. We begin by introducing the separation-deviation model of Hochbaum and Levin (2006), which links the Kemeny model to maximum cycle removal methods, and we show that the convex relaxation of Kemeny's model is the dual of the maximum circulation problem (but not the strong maximum circulation problem).

The Hochbaum and Levin model, henceforth called the HL model, determines a score for each alternative so that the vector of scores minimizes a loss function. The voting system permitted in the HL model is more general than the preference votes studied here: Each voter may provide a score for each alternative, as well as a pairwise preference, that is expressed with intensity of preference for each pair of alternatives (e.g., the differences in ranks among pairs of alternatives in a voter's ranked list).

Given the more general voting system in the HL model, we will refer to voters in this section as *reviewers*. For each alternative i, each reviewer r can submit a score denoted as s_i^r. For each pair of alternatives i and j, reviewer r can submit a preference of i over j with intensity p_{ij}^r.

The decision variables for the HL model are the components of the vector $\mathbf{y}$, where $y(i)$ is the score of alternative i; the higher the score, the more preferred the alternative.

For each reviewer r and alternative i, there is a loss $g_i^r(y(i) - s_i^v)$, which is a function of the *deviation* of the final score from the score that reviewer r provides for alternative i. For each reviewer r and for each pair i, j of alternatives, there is a loss *separation* function $f_{i,j}^r(p_{ij}^r - (y(i) - y(j)))$, which is a function of the solution's difference in scores as compared to the intensity of the preference of reviewer r for this pair i, j. Both the separation and deviation functions take the value 0 for zero valued arguments.

As before, we let V be the set of alternatives, and A the set of pairs (i, j) where at least one reviewer expresses a preference of i over j. We assume that $g_i^r(0) = 0$ for all $i \in V$ and $r \in R$. We also assume that $f_{ij}^r(0) = 0$ for all $i, j \in V$ and $r \in R$. Accordingly, if the voter's score or preference is identical to the solution's, then the loss penalty is zero. The HL model is to determine a ranking that minimizes the total deviation and separation loss for all voters.

Let R_{ij} be the set of reviewers that express a preference of i over j. Let R_i be the set of reviewers that assigned a score to alternative i. Recall that $q_{ij} = |R_{ij}|$ is the number of votes expressing a preference of i over j. The variables $y(i)$ can take (continuous) values in an interval, or they can be restricted to take integer values in a given range.

The overall model is to minimize the aggregate loss function:

$$\text{(HL)} \quad \min Loss(\mathbf{y}) = \sum_{i \in V} \sum_{r \in R_i} g_i^r(y(i) - s_i^r) + \sum_{(i,j) \in A} \sum_{r \in R_{ij}} f_{ij}^r(y(i) - y(j) - p_{ij}^r).$$

Hochbaum and Levin showed how to solve the optimization problem (HL) in polynomial time when the functions $g_i^r()$ and $f_{ij}^r()$ are convex. They showed how to solve this convex problem in polynomial time for integer variables (and also for continuous variables). This problem was shown to be a special case of the *convex dual of minimum cost network flow*, which is solved by the algorithm of Ahuja, Hochbaum and Orlin (2003) in $O(nm \log \frac{n^2}{m} \log nU)$, where n is the number of alternatives $|V|$, m is the number of different f_{ij} functions, and U is the range of the variables. For example, if the components of $\mathbf{y}$ are constrained to fall in the range $[-n, n]$, then $U = 2n$. When the functions are non-convex, then the problem is NP-hard, see, e.g., Gupta and Tardos (2000).

The preference voting system considered here is a special case of that in HL in which $R_i = \emptyset$ for all $i \in V$, and where $p_{ij}^r = 1$ for all $r \in R_{ij}$. That is, whenever reviewer r prefers alternative i to alternative j it is associated with an intensity of 1 unit. In addition, all decision variables (scores) are required to be integer valued. For each $r \in R_{ij}$, there is an associated loss in the objective function of $f_{ij}^r(y(i) - y(j) - 1)$.

We observe that Kemeny's model can be represented as a special case of the HL model. Let $\delta^+(z)$ denote the function that takes the value 1 when $z > 0$ and zero otherwise. With this notation, Kemeny's model can be expressed as a special case of the HL model as follows:

$$\text{(KEM)} \quad \min \sum_{(i,j) \in A} q_{ij} \delta^+(y(j) - y(i) + 1).$$

This objective function is non-convex, and there is no known efficient algorithm for solving it. Recall that finding an optimal solution to the Kemeny model is NP-hard.

Taking a convex relaxation of the objective in Kemeny's model, one obtains the following convex loss function for the resulting optimization problem:

$$\text{(Relax-KEM)} \quad \min \sum_{(i,j) \in A} q_{ij} \max\{(y(j) - y(i) + 1), 0\}.$$

Gupte et al. (2011) created a model equivalent to Relax-KEM. where they referred to the loss function as the "agony" loss function. They also showed that Relax-KEM is the dual of the problem that we refer to as the maximum circulation problem. This result is a special case of the fact that convex HL and the convex separation-deviation model is a convex cost dual network flow problem (Hochbaum and Levin, 2006; Hochbaum, 2001). We provide their result as well as a more concise proof next.

Lemma 4. *(Relax-KEM) is the equivalent to the dual of the maximum circulation problem (1).*

Proof. The dual to the maximum circulation problem is LP (4). Recall that any optimal solution $(\mathbf{y}, \mathbf{z})$ to 4 satisfies the following: $z_{ij} = \max\{y_j - y_i + 1, 0\}$. Therefore, if $(\mathbf{y}, \mathbf{z})$ is optimal for (4), then $y(i) = y_i$ minimizes (Relax-KEM). □

The maximum circulation problem has the property that all cost coefficients are 1. If the optimum vector of dual variables $\mathbf{y}'$ is calculated based on an optimal spanning tree solution (see for example, Ahuja, Magnanti and Orlin (1993)), then $\mathbf{y}'$ is integral, and $|y'_i - y'_j| \leq n - 1$ for all $i, j \in V$. By subtracting $1 + \max\{y'_i : i \in V\}$ from all coefficients in $\mathbf{y}'$ one obtains an optimal dual solution in which all coefficients are in the range $\{-n, ..., -1\}$.

Consequently, the approach of the maximum circulation can be viewed as the dual of Kemeny's method's convex relaxation of the penalty minimization formulation (KEM).

6 Minimum Maximal Circulations

A natural counterpart of Kemeny's method in our setting is to find a maximal circulation that removes a minimum number of votes. We refer to this problem as the *minimum maximal circulation problem.* While this problem does have the desirable property of having as many votes as possible included in the acyclic graph, it also comes with the following drawbacks, which are established next.

1. **Non-uniqueness**: There may be more than one minimum maximal circulation. Further, these minimum maximal circulations may induce conflicting partial orders. That is, there may be minimum maximal circulations $\mathbf{x}'$ and $\mathbf{x}''$ and a pair of alternatives i and j such that $(i, j) \in A^{TC}[V, \mathbf{q} - \mathbf{x}']$ and $(j, i) \in A^{TC}[V, \mathbf{q} - \mathbf{x}'']$, as shown in Subsection 6.1.

2. **NP-hardness**: Finding a minimum maximal circulation is NP-hard, as shown in Subsection 6.2.

6.1 Minimum Maximal Circulations Can Induce Conflicting Partial Orders

In this subsection, we show that there may be two different minimum maximal circulations that lead to conflicting partial orders. Cycle cancelling approaches based on minimum maximal circulations thus face similar normative challenges as Kemeny's method. Alternatives receive more or less favorable treatment depending on which set of votes is eliminated.

Consider the vote graph $G = (V, A)$, illustrated in the top panel of Figure 2. Here $V = \{1, 2, ..., 8\}$, and $A = \{(1, 2), (2, 3), (2, 5), (3, 4), (4, 1), (4, 7), (5, 6), (6, 1), (6, 5), (7, 8), (8, 3), (8, 7)\}$ where each arc in A has a capacity of 1.

There are three maximal circulations denoted $\mathbf{x}^*$, $\mathbf{x}'$, and $\mathbf{x}''$ in the second, third, and bottom rows of the figure. $\mathbf{x}^*$ is a (strong) maximum circulation containing 8 votes. $\mathbf{x}'$ and $\mathbf{x}''$ are minimum maximal circulations, which contain 6 votes each. $\mathbf{x}'$ consists of directed cycles 3-4-7-8-3 and 5-6-5. $\mathbf{x}''$ consists of directed cycles 1-2-5-6-1 and 7-8-7. After eliminating $\mathbf{x}'$, there remains a directed path from node 1 to node 3. Therefore, 1 is preferred to 3 in $A^{TC}(\mathbf{q} - \mathbf{x}')$. After eliminating $\mathbf{x}''$, there remains a directed path from node 3 to node 1. Therefore, 3 is preferred to 1 in $A^{TC}(\mathbf{q} - \mathbf{x}'')$. Thus, the two partial orders conflict.

Note that $A^{TC}(\mathbf{q} - \mathbf{x}^*)$ (or A^{SP}) does not include a dominance relationship between 1 and 3. Therefore, rankings for which 1 is better-ranked than 3, for which 3 is better-ranked than 1, and for which 1 and 3 receive the same rank are all consistent with the strong partial order induced by the strong maximum circulation $\mathbf{x}^*$. In the minimum consistent ranking, in which the rank of each alternative is the smallest number possible consistent with the partial order, 1 and 3 are both second-ranked, because 1 is dominated by 4 in A^{SP}, and 3 is dominated by 2.

6.2 Computational Complexity of Finding a Minimum Maximal Circulation

We now turn towards the computational complexity of finding a minimum maximal circulation. We first note that there is an efficient algorithm for determining a maximal (although not necessarily minimum maximal) circulation. A maximal circulation is closely connected to the concept of "blocking flow" for

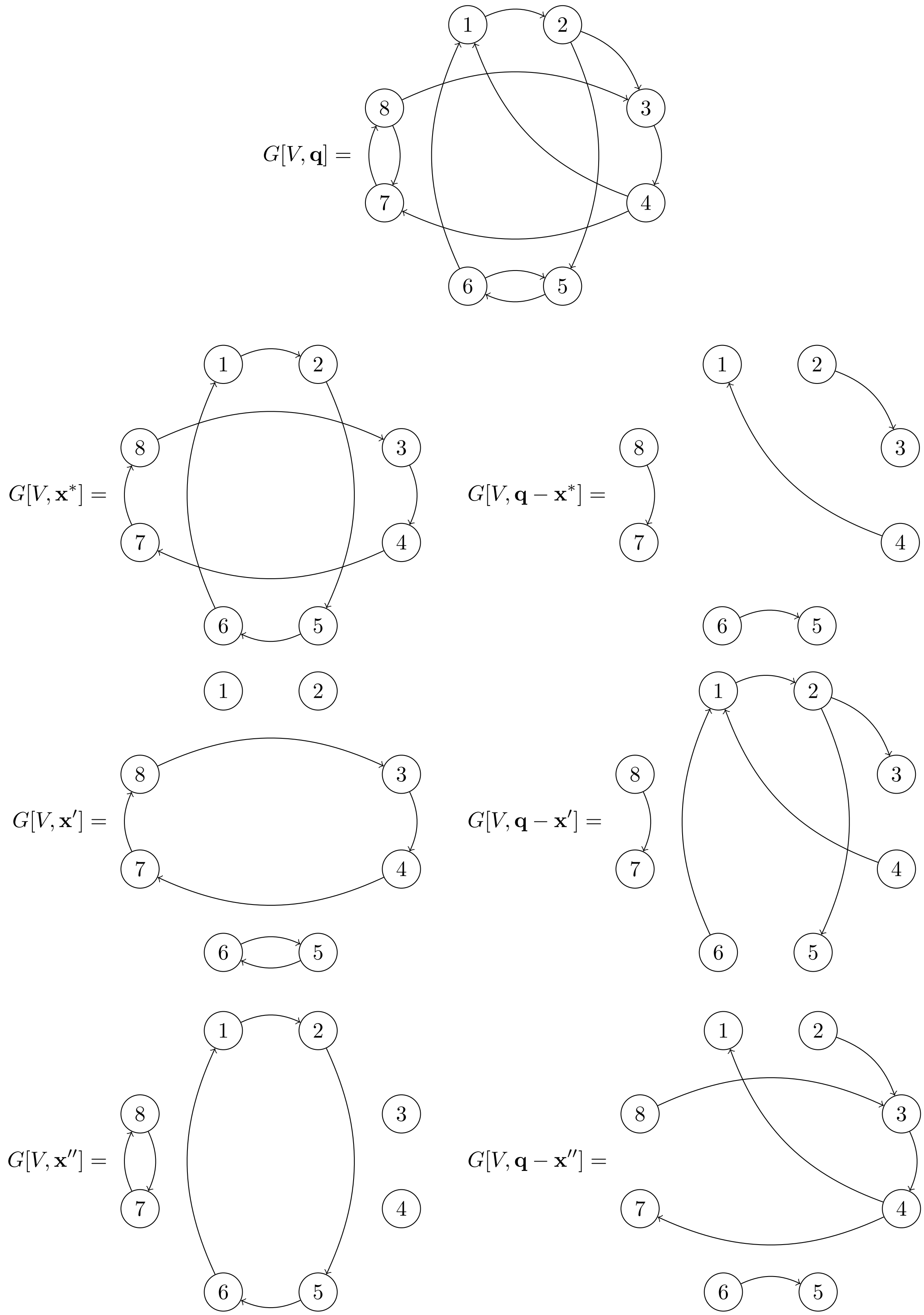


Figure 2: Example of a vote graph where minimum maximal circulations induce conflicting partial orders.

the maximum flow problem. (A blocking flow is a maximal flow from a source node to a sink node.) Sleator and Tarjan (1981) show how to find a blocking flow in $O(m \log n)$ time using the dynamic tree data structure that they developed. Their algorithm can be easily modified to find a maximal circulation (but not necessarily a minimum maximal circulation) in $O(m \log n)$ time. We restate this as a theorem.

Theorem 2. *There is an algorithm to find a maximal circulation in $O(m \log n)$ time.*

We next demonstrate the NP-hardness of finding a minimum maximal circulation, which we refer to as the *minmax circulation problem*. To prove that the problem is NP-complete we formalize this problem as a *decision problem*. If a decision problem is NP-complete, then the respective optimization version is said to be NP-hard.

Recall that the minimum FAS problem, which we restate here as a decision problem, is NP-complete. As a result, finding a minimum FAS is NP-hard (Gary and Johnson, 1979).

Minimum Feedback Arc Set (G, K)
INSTANCE: A directed graph $G = (V, A)$ and integer K.
QUESTION: Is there a subset of at most K arcs whose removal from G leaves an acyclic graph?

We next state the minmax circulation problem as a decision problem and prove that it is also NP-complete:

The Minmax Circulation Problem (G, K)
INSTANCE: A directed multi-graph $G = (V, A)$ and integer K.
QUESTION: Is there a circulation with at most K arcs whose removal from G leaves an acyclic graph?

Theorem 3. *The Minmax Circulation Problem is NP-complete.*

Proof. It is clearly in the class NP. Let $G = (V, A)$ and K be an instance of the Feedback Arc Set Problem. We create a graph $G' = (V', A')$ for the Minmax Circulation Problem as follows.

1. $V' = V \cup \{s, t\} \cup V^*$, where V^* is defined as follows. For every arc $(i, j) \in A$, we replace (i, j) by a path $P[i, j]$ of length 3. The first and last nodes of $P[i, j]$ are i and j. The second and third nodes are additional nodes added to V^* that we label ij_1 and ij_2.

2. A' is defined as follows: For each arc $(i, j) \in A$, A' includes the arcs of the path $P[i, j] = (i, ij_1), (ij_1, ij_2), (ij_2, j)$. In addition, A' includes an arc (s, ij_1) and an arc (ij_2, t). (A' does not include the arc (i, j).)

3. A' also includes K copies of arc (t, s). (A' contains no other arc.)

4. $K' = 4K$.

Note that we have introduced a number of cycles of length four in the graph G'. For each $(i, j) \in A$, let $C[i, j] = s\text{-}ij_1\text{-}ij_2\text{-}t\text{-}s$. Thus $C[i, j]$ has four arcs. There is no cycle in A' with fewer than four arcs.

We claim that there is a feasible solution for the minmax circulation problem (G', K') if and only if there is a feasible solution for the feedback arc set problem instance (G, K). Suppose first that there is a subset $S \subseteq A$ of K arcs whose deletion from G results in an acyclic graph. Let $S' \subseteq A'$ be obtained as follows: For each $(i, j) \in A$, if $(i, j) \in S$, then S' contains the cycle $C[i, j]$. Accordingly, S' contains K cycles, each with four arcs. The graph with arc set $A' \setminus S'$ is acyclic.

Suppose conversely that $S' \subseteq A'$ has at most K' arcs and suppose that $A' \setminus S'$ is acyclic. S' must contain all K copies of (t, s). Moreover, each simple cycle in A' contains at least 4 arcs. Since $K' = 4K$, each of the cycles in S' contain exactly four arcs. The only cycles in A' with four arcs are the sets $C[i, j]$ for $(i, j) \in A$. Let $S = \{(i, j) : C[i, j] \subset S'\}$. Then $A \setminus S$ is acyclic. $\square$

Guruswami et al. (2011) proved that if the Unique Games Conjecture is true, then it is hard to approximate the minimum feedback arc set to within any constant factor in polynomial time. Because of the nature of our reduction (factor 4, constant, used in the reduction), we have the following corollary:

Corollary 4. *If the Unique Games Conjecture is true, then it is hard to approximate the minimum maximal circulation to within any constant factor in polynomial time.*

Preference aggregation based on the removal of minimum maximal circulations therefore faces some of the same normative and computational issues as Kemeny's model, in contrast to the strong partial order, which is unique, computationally efficient, and consistent with a convex relaxation of Kemeny's model.

7 Conclusion

It is self evident that a cycle of preference votes, each requiring that one alternative ranks higher than the next, cannot conform to any partial order, or consistent ranking. Such a cycle represents a tie, or contradictory votes for the alternatives on the cycle, and must be disregarded in order to obtain a consensus partial order that aggregates the other votes.

We explore various methods for removing unions of cycles of votes, i.e., maximal circulations in the vote graph, and inferring aggregate preferences from the non-conflicting remainder. We begin by examining methods that are based on the removal of a maximum circulation. A major contribution of our paper is the introduction of the *strong* maximum circulation that guarantees the *uniqueness* of the induced *strong partial order*. If there is *any* maximum circulation that keeps at least one vote for a certain preference, then so does the strong maximum circulation. A second major contribution is that we show how to obtain a strong maximum circulation by solving a single minimum cost flow problem.

We then compare the maximum circulation-removal approach with the well-known Kemeny method. Kemeny's method does not return a unique solution and may result in contradictory rankings among the different solutions. Further, there is no polynomial time algorithm to produce an optimal solution for Kemeny's method (unless P = NP). The strong maximum circulation problem, in contrast, is solved in strongly polynomial time using combinatorial network flow algorithms. However, we also demonstrate a close relationship between the two methods, as the dual of the maximum circulation problem can be viewed as a convex relaxation of Kemeny's model. We also explore the possibility of removing minimum maximal circulations. We show that this approach inherits the drawbacks associated with Kemeny's model related to the multiplicity of inconsistent solutions and computational complexity.

We conclude by noting that there are alternative maximal cycle removal approaches that remove fewer votes than a strong maximum circulation does that also inherit many of the desirable properties of the strong maximum circulation algorithm. For example, it is straightforward to first eliminate all cycles of length 2, and then find a strong maximum circulation on the resulting graph. Doing so will frequently eliminate fewer votes than the strong maximum circulation and will also guarantee that any alternative that beats every other alternative in pairwise comparison is the unique winner. Exploring other efficient cycle removal approaches that also induce unique partial orders—and exploring the properties of these inferred aggregate preferences—is a worthwhile avenue for future research.

Acknowledgement

Dorit Hochbaum's research is supported in part by AI institute NSF award 2112533.

References

Agarwal, Shivani, and Lek-Heng Lim. 2010. "The Mathematics of Ranking." American Institute of Mathematics, Palo Alto.

Ahuja, Ravindra K, Dorit S Hochbaum, and James B Orlin. 2003. "Solving the convex cost integer dual network flow problem." *Management Science*, 49(7): 950–964.

Ahuja, Ravindra K., Thomas L. Magnanti, and James B. Orlin. 1993. *Network flows: theory, algorithms, and applications.* USA:Prentice-Hall, Inc.

Arrow, Kenneth J. 1951. *Social Choice and Individual Values.* Yale University Press.

Bartholdi, J., C. A. Tovey, and M. A. Trick. 1989. "Voting schemes for which it can be difficult to tell who won the election." *Social Choice and Welfare*, 6(2): 157–165.

Fischer, Felix, Olivier Hudry, and Rolf Niedermeier. 2016. "Weighted tournament solutions." In *Handbook of Computational Social Choice.* , ed. Felix Brandt, Vincent Conitzer, Ulle Endriss, Jerome Lang and Ariel D. Procaccia, 85–102. Cambridge:Cambridge University Press.

Gary, Michael R, and David S Johnson. 1979. "Computers and Intractability: A Guide to the Theory of NP-completeness."

Goldberg, A., and R. Tarjan. 1987. "Solving minimum-cost flow problems by successive approximation." 7–18. New York, New York, United States:ACM Press.

Gupta, Anupam, and Éva Tardos. 2000. "A constant factor approximation algorithm for a class of classification problems." 652–658.

Gupte, Mangesh, Pravin Shankar, Jing Li, S. Muthukrishnan, and Liviu Iftode. 2011. "Finding Hierarchy in Directed Online Social Networks." 557–566. Hyderabad, India:ACM.

Guruswami, Venkatesan, Johan Hårstadrstad, Rajsekar Manokaran, Prasad Raghavendra, and Moses Charikar. 2011. "Beating the random ordering is hard: Every ordering CSP is approximation resistant." *SIAM Journal on Computing*, 40(3): 878–914.

Hochbaum, Dorit S. 2001. "An efficient algorithm for image segmentation, Markov random fields and related problems." *Journal of the ACM (JACM)*, 48(4): 686–701.

Hochbaum, Dorit S., and Asaf Levin. 2006. "Methodologies and Algorithms for Group-Rankings Decision." *Management Science*, 52(9): 1394–1408. Publisher: INFORMS.

Kemeny, John G. 1959. "Mathematics without Numbers." *Daedalus*, 88(4): 577–591. Publisher: The MIT Press.

Muravyov, Sergey V. 2014. "Dealing with chaotic results of Kemeny ranking determination." *Measurement*, 51: 328–334.

Pini, Maria Silvia, Francesca Rossi, Kristen Brent Venable, and Toby Walsh. 2009. "Aggregating partially ordered preferences." *Journal of Logic and Computation*, 19(3): 475–502.

Sleator, Daniel D., and Robert Endre Tarjan. 1981. "A data structure for dynamic trees." 114–122. Milwaukee, Wisconsin, United States:ACM Press.

Yoo, Yeawon, and Adolfo R. Escobedo. 2021. "A New Binary Programming Formulation and Social Choice Property for Kemeny Rank Aggregation." *Decision Analysis*, 18(4): 296–320.