Realizing Degree Imbalances in Directed Graphs

Dhruv Mubayi[*], Todd G. Will[†], Douglas B. West[‡]

March 30, 2003

Abstract

In a directed graph, the *imbalance* of a vertex v is $b(v) = d^+(v) - d^-(v)$. We characterize the integer lists that can occur as the sequence of imbalances of a simple directed graph. For the realizable sequences, we determine the maximum number of arcs in a realization and provide a greedy algorithm that constructs realizations with the minimum number of arcs.

1 Introduction

A sequence of integers is *graphic* if it is the degree sequence of a simple undirected graph; there are well-known characterizations of graphic sequences. We obtain analogous results for "imbalance sequences" of digraphs without repeated arcs. The *imbalance* $b(v)$ of a vertex v in a digraph is $d^+(v) - d^-(v)$, where $d^+(v)$ is the out-degree of v and $d^-(v)$ is the in-degree. A sequence $b = b_1, b_2, \ldots, b_n$, is *realizable* if there exists a simple digraph G (no repeated arcs) with vertices $v_1, v_2, \ldots, v_n$ such that $b(v_i) = b_i$.

If we allow repeated arcs, then the trivial necessary condition $\sum b_i = 0$ is sufficient, using only arcs from vertices with positive imbalance to vertices with negative imbalance. This is analogous to the observation that when $\sum d_i$ is even, the sequence d of nonnegative integers

[*]School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, mubayi@math.gatech.edu
[†]Department of Mathematics, Davidson College, Davidson, NC 28036, towill@davidson.edu
[‡]Department of Mathematics, University of Illinois, Urbana, IL 61801, west@math.uiuc.edu

is the degree sequence of some undirected graph, allowing loops and multiple edges. Hence we forbid repeated arcs. Loops and pairs of opposed arcs have no effect on the imbalance sequence, so their presence is irrelevant.

Our results are analogous to the known conditions for graphic sequences. Havel [4] and Hakimi [5] independently showed that for $n > 1$, a nonincreasing sequence $d = d_1, d_2, \ldots, d_n$ is graphic if and only if $\hat{d} = \hat{d}_2, \ldots, \hat{d}_n$ is graphic, where $\hat{d}$ is formed from d by deleting d_1 and subtracting 1 from the d_1 largest remaining elements of d. In this context, we define d_i to be *larger* than d_j if $d_i > d_j$ or if $d_i = d_j$ and $i > j$; this ensures that $\hat{d}$ is also nonincreasing. The Havel-Hakimi result tests recursively whether a sequence is graphic. When it is graphic, retracing the computation produces a realization; if $\hat{G}$ realizes $\hat{d}$, then adding a vertex adjacent to vertices whose degrees are the d_1 largest entries of $\hat{d}$ forms a graph G that realizes d.

Erdős and Gallai [2] characterized graphic sequences explicitly: the nonnegative integer sequence $d = d_1, d_2, \ldots, d_n$ is graphic if and only if 1) $\sum d_i$ is even, and 2) $\sum_{i=1}^{k} d_i \leq k(k-1) + \sum_{j=k+1}^{n} \min\{k, d_j\}$ for $1 \leq k \leq n$. The inequalities are necessary conditions because the summation on the right bounds the number of edges joining the set of k vertices with largest degrees to the remaining vertices. Aigner and Triesch [1] gave a short proof of sufficiency using order ideals in a natural partial order on nonincreasing nonnegative integer sequences with fixed sum. This order has been called the *dominance* or *majorization* order, putting $a \leq b$ if $\sum_{i=1}^{k} a_i \leq \sum_{i=1}^{k} b_i$ for all k.

A greedy algorithm for realizing an imbalance sequence directs arcs from the vertex with greatest imbalance (b_1) to the b_1 vertices with smallest imbalance. This heuristic is natural in that vertices with positive imbalance require high out-degree, and vertices with negative imbalance require high in-degree. This operation suggests a modified sequence $\hat{b}$ and a recursive test analogous to the Havel-Hakimi test.

We also prove an explicit characterization analogous to the Erdős-Gallai result. If $b = b_1, b_2, \ldots, b_n$ is realizable, then $\sum b_i = 0$, because each edge contributes positively and negatively to the total imbalance. Also, $\sum_{i=1}^{k} b_i \leq k(n-k)$; letting S be the set of k vertices with largest imbalance, the edges within S contribute nothing to $\sum_{i=1}^{k} b_i$, and the pairs $S \times (V(G) - S)$ contribute at most one each.

In section 2, we discuss three proofs that these obvious necessary conditions are also sufficient. The first shows that the greedy algorithm suggested above produces a realization when one exists. The second proof, based on the method of ideals used by Aigner and Triesch, is much shorter but does not yield such a fast construction algorithm. The third proof, provided by Uri Peled, obtains the result from the famous supply-demand theorem of Gale [3]. The constructive proof of the supply-demand theorem uses the augmenting path methods of network flows and thus does not directly yield as efficient an algorithm to produce a realization as the greedy algorithm provided by the first proof.

In section 3, we consider realizations of an imbalance sequence with the fewest and most arcs (here we forbid loops and pairs of opposed arcs). We prove that the greedy algorithm suggested above produces a realization with the fewest arcs. We also determine the maximum number of arcs in a realization of an imbalance sequence.

2 Necessary and Sufficient Conditions

In this section, we provide necessary and sufficient conditions for a sequence of integers to be the imbalance sequence of a simple directed graph. We also provide an algorithm for constructing a realization when one exists.

Definition 1 *A nonincreasing integer sequence* $a = a_1, a_2, \ldots, a_n$ *is* feasible *if it has sum zero and satisfies* $\sum_{i=1}^{k} a_i \le k(n-k)$ *for* $1 \le k < n$.

From a feasible sequence $a = a_1, a_2, \ldots, a_n$, we form another sequence $\hat{a} = \hat{a}_2, \ldots, \hat{a}_n$ by deleting a_1 and adding 1 to the a_1 smallest elements of a, defining a_i to be *smaller* than a_j if $a_i < a_j$ or if $a_i = a_j$ and $i < j$. The sequence $\hat{a}$ is the result of applying one step of the greedy algorithm to the sequence a. The complication in defining $\hat{a}$ arises when $a_{n-a_1+1} = a_{n-a_1}$; in this case there is a gap consisting of elements to which we add zero. We endure this complication to ensure that $\hat{a}$ is nonincreasing. The following example with $n = 9$ and $a_1 = 5$ produces such a gap, since $a_5 = a_4 = 2$.

$$\begin{array}{rcccccccccc} a: & 5 & 3 & 2 & 2 & 2 & 2 & -5 & -5 & -6 \\ & \cdot & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ \hat{a}: & \cdot & 3 & 3 & 3 & 2 & 2 & -4 & -4 & -5 \end{array}$$

The values of a_i in the gap and to the left of the gap are all equal. When this occurs in a feasible sequence, a stronger statement can be made about the partial sums.

Lemma 2 *If a is a feasible sequence and $a_k = a_{k+1} = \cdots = a_{k+m}$, then $\sum_{i=1}^{k} a_i \leq k(n-k)-m$.*

Proof: If $a_k \leq n - 2k - m$, then

$$\sum_{i=1}^{k} a_i = (\sum_{i=1}^{k-1} a_i) + a_k \leq (k-1)(n-k+1) + (n-2k-m) = k(n-k) - m - 1.$$

If $a_k > n - 2k - m$, then

$$\sum_{i=1}^{k} a_i = (\sum_{i=1}^{k+m} a_i) - ma_k \leq (k+m)(n-k-m) - m(n-2k-m+1) = k(n-k) - m. \qquad \square$$

Theorem 3 *If a is feasible, then $\hat{a}$ is feasible.*

Proof: Because $\sum_{i=1}^{n} a_i = 0$ and $a_1 \leq 1 \cdot (n-1) = n-1$, the sequence $\hat{a}$ is well defined. By the definition of the a_1 "smallest" elements, $\hat{a}$ is non-increasing. The construction of $\hat{a}$ distributes a_1 among the other entries, so the sum is still 0. It remains to verify the condition on partial sums, which is $\sum_{i=2}^{k} \hat{a}_i \leq (k-1)(n-k)$ for $2 \leq k < n$.

Let $a_1 = n - r$, and suppose that the construction of $\hat{a}$ involves a gap of s entries to which we add 0, where $s \geq 0$. Beginning with the leftmost entry to which 1 is added, there are $(n-r)+s$ positions in the sequence; hence $s < r$ and a_{r-s+1} is the leftmost entry to which 1 is added. If $k \leq r - s$, then

$$\sum_{i=2}^{k} \hat{a}_i = \sum_{i=2}^{k} a_i \leq (k-1)(n-r) \leq (k-1)(n-k).$$

Now consider $k > r - s$. Let $\alpha = \sum_{i=2}^{k} \hat{a}_i - \sum_{i=2}^{k} a_i$; this is the number of added 1's in the sum. Let t be the number of positions in the sequence that are to the left of the gap and receive an augmentation of 1 ($t = 0$ if $s = 0$). The rightmost entry to which 0 is added is a_{r+t}.

$$\begin{array}{cccccccccccccc}
 & & & & \multicolumn{3}{c}{\overbrace{\hspace{8em}}^{t}} & \multicolumn{3}{c}{\overbrace{\hspace{8em}}^{s}} & & & \\
a: & a_1 & a_2 & \cdots & a_{r-s} & a_{r-s+1} & \cdots & a_{r-s+t} & a_{r-s+t+1} & \cdots & a_{r+t} & a_{r+t+1} & \cdots & a_n \\
 & & 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & \cdots & 0 & 1 & \cdots & 1 \\
\hat{a}: & & \hat{a}_2 & \cdots & \hat{a}_{r-s} & \hat{a}_{r-s+1} & \cdots & \hat{a}_{r-s+t} & \hat{a}_{r-s+t+1} & \cdots & \hat{a}_{r+t} & \hat{a}_{r+t+1} & \cdots & \hat{a}_n
\end{array}$$

If $r-s<k<r+t$, then $a_k=\cdots=a_{r+t}$. By the Lemma, $\sum_{i=1}^k a_i \le k(n-k)-(r+t-k)$. In this case, $\alpha\le t$. If $k\ge r+t$, then $\sum_{i=1}^k a_i\le k(n-k)$ and $\alpha=k-r$. In either case, we have $(\sum_{i=1}^k a_i)+\alpha\le k(n-k)+(k-r)$. Hence

$$\sum_{i=2}^{k}\hat{a}_i=(\sum_{i=1}^{k}a_i)+\alpha-a_1\le k(n-k)+(k-r)-(n-r)=(k-1)(n-k). \qquad \square$$

A simple inductive argument now completes the proof of the sufficiency of the condition and shows that the greedy algorithm produces a realization if one exists.

Theorem 4 *A sequence is realizable as an imbalance sequence if and only if it is feasible.*

Proof: We have already argued the necessity of the condition. For sufficiency, we use induction on the length n of a feasible sequence a. If $n=1$, then $a_1=0$ and K_1 has imbalance sequence a. If $n>1$, then Theorem 3 implies that $\hat{a}$ is a feasible sequence of length $n-1$. By the induction hypothesis, there is a graph G' with vertices $v_2,v_3,\ldots,v_n$ such that $b(v_i)=\hat{a}_i$. Form G by adding the vertex v_1 and arcs from v_1 to v_j for each $j\in V(G')$ such that $\hat{a}_j=a_j+1$. The graph G realizes a. $\square$

Kleitman and Wang [6] observed that the Havel-Hakimi argument works with the deletion of *any* element d_k of the sequence to be tested, subtracting 1 from the d_k largest other elements. The analogous statement about imbalance is false. Consider the imbalance sequence $(3,1,-1,-3)$ of a transitive tournament. Deleting the element 1 and adding 1 to the smallest imbalance leaves us trying to realize $(3,-1,-2)$, which has no realization by a simple digraph. In a transitive tournament, we must start by deleting an extreme element. The difference in the directed case is that the reduction step works with high imbalance and low imbalance separately, while in the undirected case no high values are preserved.

Our second proof of sufficiency uses the method of Aigner and Triesch [1]. This proof is much shorter but does not directly yield a fast algorithm. The idea is to define a partial order P on sequences such that 1) the realizable sequences form an ideal (down-set), and 2) the maximal sequences in P among those satisfy the desired condition (here feasibility) are realizable. This implies that all sequences satisfying the condition are realizable, because they are dominated in P by a realizable sequence.

Second Sufficiency Proof. Consider the majorization order P on nonincreasing integer sequences with sum 0. In the subposet of P induced by the feasible sequences, we claim that only the sequence $n-1, n-3, \ldots, -(n-3), -(n-1)$ is maximal. It achieves each constraint with equality, and therefore by the definition of the order it dominates all other feasible sequences. Furthermore, this is the imbalance sequence of the transitive tournament with n vertices and hence is realizable.

We must also show that the realizable sequences in P form an ideal. If $b > b'$ in P, then b' is obtained from b by a sequence of unit shifts in which some position i decreases by 1 and some later position j increases by one; it suffices to show that every such shift maintains realizability. Consider a realizable b with $b_i > b_j$, realized by G. The inequality and the prohibition of multiple edges from G yield a vertex $z \notin \{v_i, v_j\}$ satisfying one of the three cases below. In each case, the listed action produces a new digraph G' with the same imbalances, except that the imbalance of v_i decreases by 1 and the imbalance of v_j increases by 1.

case	action
$i \rightarrow z \not\leftrightarrow j$	$i \not\leftrightarrow z \leftarrow j$
$i \rightarrow z \rightarrow j$	$i \not\leftrightarrow z \not\leftrightarrow j$
$i \not\leftrightarrow z \rightarrow j$	$i \leftarrow z \not\leftrightarrow j$

□

Third Sufficiency Proof (Uri Peled). The characterization of realizable sequences also follows from Gale's Supply-Demand Theorem [3]. We sketch this argument. Gale's Theorem considers a network with edge capacities $c(e)$ and vertices partitioned into three sets S, R, T. We associate with each $x \in S$ a positive *supply* $\sigma(x)$ and with each $y \in T$ a positive *demand* $\partial(y)$. A *feasible flow* f assigns a flow $f(e)$ to each edge e such that $0 \leq f(e) \leq c(e)$ and such that the net outflow from vertex v is at most $\sigma(v)$ if $v \in S$, at least $-\partial(v)$ if $v \in T$, and 0 otherwise. Gale proved that a feasible flow exists if and only if for each vertex subset U, the total demand of sinks in U minus the total supply from sources in U is at most the total capacity of edges entering U. Also, there is an integer feasible flow when the constraints are integral and there is a feasible flow.

This applies to realizability of an imbalance sequence a by considering the complete digraph on n nodes with unit capacities. Let S be the set of nodes with positive imbalance, putting

$\sigma(x) = a(x)$ for $x \in S$. Let T be the set of nodes with negative imbalance, putting $\partial(y) = -a(y)$ for $y \in T$. Since the total demand equals the total supply, we may assume equality in the supply/demand constraints, and an integer feasible flow produces a realization of a consisting of the edges with flow 1. In Gale's condition $\partial(T \cap U) - \sigma(S \cap U) \leq c(\bar{U}, U)$ for a set U of size $n-k$, the left side is the negative of the total imbalance of vertices in U (which equals the total imbalance of vertices in $\bar{U}$), and the right side is $k(n-k)$. The left side is maximized by letting $\bar{U}$ be the set of k vertices with largest desired imbalance, so Gale's condition reduces to our condition for feasible sequences.

3 Extremal Realizations

In this section we prove that the greedy algorithm produces a realization with the fewest arcs. We also determine the maximum number of arcs in a realization of an imbalance sequence.

For a realizable sequence b, let D_b be the digraph with imbalance sequence b produced by the greedy algorithm described in Section 2.

Theorem 5 *Among realizations of $b = b_1, \ldots, b_n$, the digraph D_b has the fewest arcs.*

Proof: We use induction on n. The claim holds trivially when $n = 1$. Consider $n > 1$. Consider realizations of b on vertex set $v_1, \ldots, v_n$ with $b(v_i) = b_i$. Among realizations with fewest arcs, let D be one such that $N^+(v_1) \cap N^+_{D_b}(v_1)$ is as large as possible. If we can show that $N^-(v_1) = N^-_{D_b}(v_1) = \emptyset$ and $N^+(v_1) = N^+_{D_b}(v_1)$, then we can remove v_1 from D to obtain a digraph D' that realizes b' and apply the induction hypothesis. We show first that $N^-(v_1) = \emptyset$ and then obtain $N^+(v_1) = N^+_{D_b}(v_1)$.

Step 1: Suppose that $N^-(v_1)$ and $N^+(v_1)$ are nonempty. They are disjoint, since opposed arcs can be deleted. Let w be a vertex of maximum imbalance in the subdigraph induced by $N^-(v_1)$. If there is an edge from $N^+(v_1)$ to w, then a 3-cycle can be deleted. If some $y \in N^+(v_1)$ is not adjacent to w, then the w, y-path through v_1 can can replaced with the edge wy. Otherwise, $b(v_1) \geq b(w)$ requires the existence of a vertex z not adjacent to v_1 such that zw is an edge. Now the z, v_1-path through w can be replaced with the edge zv_1. In each case, we have reduced the number of arcs without changing any imbalances.

Step 2: We may assume that $N^-(v_1)$ is empty but that $N^+(v_1) \neq N^+_{D_b}(v_1)$. It suffices to enlarge $N^+(v_1) \cap N^+_{D_b}(v_1)$ without increasing the number of arcs. Recall that $N^+_{D_b}(v_1)$ consists of the b_1 vertices of smallest imbalance. Thus we may assume that $v_i \in N^+(v_1)$ and $v_j \notin N^+(v_1)$, with $b_i \geq b_j$. Also $v_j v_1$ is not an edge, so $b_i \geq b_j$ implies the existence of a vertex z that contributes more to the imbalance of v_i than to that of v_j. Since the digraph is simple, we have the three cases listed below. Combined with the replacement of $v_1 v_i$ with $v_1 v_j$, the listed actions enlarge $N^+(v_1) \cap N^+_{D_b}(v_1)$ without changing imbalances, violating simplicity, or increasing the number of arcs.

case	action
$v_i \to z \not\leftrightarrow v_j$	$v_i \not\leftrightarrow z \leftarrow v_j$
$v_i \to z \to v_j$	$v_i \not\leftrightarrow z \not\leftrightarrow v_j$
$v_i \not\leftrightarrow z \to v_j$	$v_i \leftarrow z \not\leftrightarrow v_j$

□

Finally, we determine the maximum number of arcs in a realization of b, among simple digraphs with no opposed arcs.

Theorem 6 *For a realizable imbalance sequence $b = b_1, \ldots, b_n$, the maximum number of arcs in a realization without opposed arcs is $\sum_i \lfloor (n-1+b_i)/2 \rfloor$.*

Proof: We show first that in a realization with the most arcs, each vertex has at most one nonneighbor. If v is nonadjacent to both x and y, then without changing imbalances we can add a 3-cycle on these vertices (if $x \not\leftrightarrow y$) or replace an edge between x and y with a path between them through v.

To obtain the desired formula, we count the edges by summing the out-degrees. From $b(v) = d^+(v) - d^-(v)$ and $n-2 \leq d^+(v) + d^-(v) \leq n-1$, we obtain $d^+(v) = \lfloor (n-1+b(v))/2 \rfloor$. □

We can also express the number of edges in the maximum realization as $\binom{n}{2}$ minus half the number of entries in b having opposite parity from $n-1$. This holds because the parity of the number of neighbors of v is the same as the parity of $b(v)$; hence v has a nonneighbor exactly when $b(v)$ and $n-1$ have opposite parity.

One might hope that there is a realization of b with any number of arcs between the minimum and maximum, but the sequence of all zeros (with length at least three) shows that this is not true.

4 Acknowledgments

We thank the referees for suggesting the inclusion of the last section and Uri Peled for showing us how [3] can be used to give an alternative proof of our main result.

References

[1] M. Aigner and E. Triesch, *Realizability and uniqueness in graphs*, Discrete Math. **136** (1994), 3–20.

[2] P. Erdős and T. Gallai, *Graphs with prescribed degrees of vertices*, Mat. Lapok **11** (1960), 264–274 (in Hungarian).

[3] D. Gale, *A theorem on flows in networks*, Pacific J. Math. **7** (1957), 1073–1082.

[4] V. Havel, *A remark on the existence of finite graphs*, Casopis Pest. Mat. **80** (1955), 477–480.

[5] S. L. Hakimi, *On the realizability of a set of integers as degrees of the vertices of a graph.*, SIAM J. Appl. Math. **10** (1962), 496–506.

[6] D. J. Kleitman and D. L. Wang, *Algorithms for constructing graphs and digraphs with given valences and factors*, Discrete Math. **6** (1973), 79–88.