Inclusion-Exclusion, Cancellation, and Consecutive Sets

D. R. Shier
Department of Mathematics
The College of William and Mary
Williamsburg, VA USA 23187

M. H. McIlwain
Department of Mathematics
Furman University
Greenville, SC USA 29613

Abstract

The principle of inclusion and exclusion has been applied to numerous areas of discrete and combinatorial mathematics. One manifestation of this principle occurs in expressing the probability of the union of events $E_1, E_2, \ldots, E_k$ as an alternating sum of probabilities of various intersections of these events. If the constituent events are themselves sufficiently well structured, then predictable cancellation occurs in this expansion. We discuss the special case in which each of the underlying sets is "consecutive": namely, its elements are consecutive integers. For such consecutive systems the inclusion-exclusion expansion assumes a particularly simple form, in which all reduced coefficients in the expression equal ± 1. Moreover, the appropriate sign of each noncancelling term is dictated by the length of a certain path in a graph derived from the incidence structure of the given sets.

1. Introduction

To motivate the subsequent discussion, consider the problem of sending a message from one specified vertex s to another specified vertex t in a communication system defined by the directed graph $G = (V, E)$. Edges $e \in E$ are known to fail randomly, and independently, with probability $q_e = 1 - p_e$; that is, p_e represents the edge reliability or edge availability at a random instant. A fundamental problem in stochastic network analysis involves calculating the *two-terminal reliability* $R_{st}(G)$ of G: namely, the probability that at a random instant there is an available s-t path in G.

One approach to calculating $R_{st}(G)$ requires first enumerating the simple s-t paths $P_1, P_2, \ldots, P_k$ of G. Let E_j denote the event in which all edges in path P_j are operating. By independence, $\Pr[E_j] = \Pi\{p_e: e \in P_j\}$ is easy to calculate. Then the two-terminal reliability $R_{st}(G)$ can be expressed as

$$R_{st}(G) = \Pr[E_1 \cup E_2 \cup \cdots \cup E_k].$$

Application of the inclusion-exclusion principle yields

$$R_{st}(G) = \sum_i \Pr[E_i] - \sum_{i<j} \Pr[E_i E_j] + \cdots + (-1)^{k+1} \Pr[E_1 E_2 \cdots E_k]. \qquad (1.1)$$

By independence each individual term of (1.1) is easy to calculate; however there are $2^k - 1$ terms to calculate in the expression.

For example, in the graph of Figure 1.1, there are four simple s-t paths, identified by their edge sets:

$$P_1\text{: 1-5,} \quad P_2\text{: 2-6,} \quad P_3\text{: 1-3-6,} \quad P_4\text{: 2-4-5.}$$

If p_i indicates the reliability of edge i, then application of (1.1) yields

$$\begin{aligned} R_{st}(G) &= \Pr[E_1 \cup \cdots \cup E_4] \\ &= p_1p_5 + p_2p_6 + p_1p_3p_6 + p_2p_4p_5 - p_1p_2p_5p_6 - p_1p_3p_5p_6 \\ &\quad - p_1p_2p_4p_5 - p_1p_2p_3p_6 - p_2p_4p_5p_6 + p_1p_2p_3p_5p_6 + p_1p_2p_4p_5p_6. \end{aligned}$$

Notice that only 11 terms of the 15 possible terms appear in this expression. In particular, the two terms producing $-p_1p_2p_3p_4p_5p_6$ cancel the two terms producing $+p_1p_2p_3p_4p_5p_6$. A more dramatic illustration of this phenomenon is illustrated by the graph of Figure 1.2, in which there are 70 simple s-t paths and thus $2^{70} - 1 \approx 10^{21}$ terms potentially appearing in (1.1). However, after appropriate cancellation has been carried out, the reduced inclusion-exclusion expression involves only 34,983 terms. Moreover, each term appearing in this reduced expression has a coefficient of ± 1.

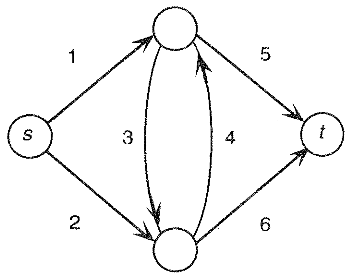

Figure 1.1

The phenomenon of cancellation was first studied by Satyanarayana and Prabhakar (1978), who showed that this ± 1 property holds for $R_{st}(G)$ in any directed graph G. Furthermore, noncancelling terms of the expansion (1.1) correspond to certain acyclic subgraphs H of G, and the reduced coefficient for H in the expansion is precisely $(-1)^{|E(H)|-|V(H)|+1}$, where $|E(H)|$ and $|V(H)|$ refer, respectively, to the number of edges and vertices of subgraph H.

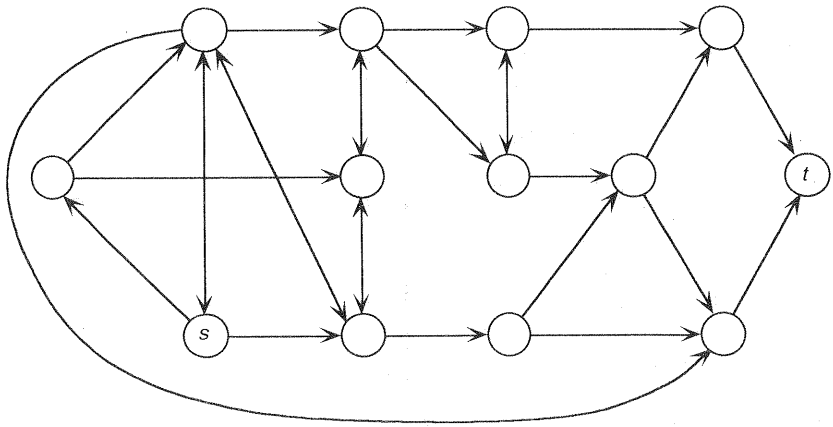

Figure 1.2

The two-terminal reliability of a graph provides one example of the more general notion of the reliability of a coherent system. Let $N = \{1, 2, \ldots, n\}$ denote a set of components, each of which fails independently; $q_i = 1 - p_i$ designates the failure probability of component i. In addition, a structure function Φ is defined on subsets $X \subseteq N$, with $\Phi(X) = 1$ if the system operates when all components of X operate and all components of $N - X$ fail, and $\Phi(X) = 0$ otherwise. In a coherent system, the structure function is monotone: $X \subseteq Y \Rightarrow \Phi(X) \leq \Phi(Y)$. That is, repairing a failed component in a coherent system cannot degrade the overall performance of the system. A coherent system is completely described by its minimal operating sets (or *pathsets*): namely, minimal sets $S \subseteq N$ such that $\Phi(S) = 1$. It is also assumed in a coherent system that each component i appears in some pathset S; that is, component i is "relevant" to the operation of the system. The fundamental problem for a coherent system is to calculate its overall reliability R_Φ. Let the system have pathsets $S_1, S_2, \ldots, S_k$, and let E_j denote the event in which all components of pathset S_j operate. Also let the event that the system as a whole operates be denoted by E_Φ. Then the overall reliability of the system is given by

$$R_\Phi = \Pr[E_\Phi] = \Pr[E_1 \cup E_2 \cup \cdots \cup E_k].$$

In this more general setting, the question again arises of when there is significant cancellation in the inclusion-exclusion expansion for R_Φ. As a first example, consider the coherent system on $N = \{1, \ldots, 4\}$ defined by the pathsets $S_1 = \{1, 2, 4\}$, $S_2 = \{2, 3\}$, and $S_3 = \{1, 3, 4\}$. Then application of (1.1) yields

$$R_\Phi = p_1p_2p_4 + p_2p_3 + p_1p_3p_4 - 2p_1p_2p_3p_4.$$

In this case, the reduced form for R_Φ contains a coefficient other than ± 1, so this does not qualify in our view as significant cancellation.

On the other hand, for the system with components $N = \{1, \ldots, 6\}$ and pathsets $S_1 = \{1, 2\}$, $S_2 = \{2, 3, 4\}$, $S_3 = \{3, 4, 5\}$, and $S_4 = \{5, 6\}$, the reduced form of (1.1) is

$$\begin{aligned} R_\Phi = {} & p_1p_2 + p_2p_3p_4 + p_3p_4p_5 + p_5p_6 - p_1p_2p_3p_4 \\ & - p_1p_2p_5p_6 - p_2p_3p_4p_5 - p_3p_4p_5p_6 + p_1p_2p_3p_4p_5p_6. \end{aligned} \tag{1.2}$$

This latter system has all reduced coefficients in (1.2) equal to ± 1, with only 9 of the 15 possible terms appearing. Systems such as this, whose pathsets display a certain type of consecutive structure, will be the focus of the present paper. Section 2 defines such consecutive systems and establishes certain fundamental relations for the coefficients of its inclusion-exclusion expansion. In Section 3, the ± 1 property is shown to hold for consecutive systems and an interpretation is provided for the occurrence of $0, +1, -1$

coefficients in the expansion (1.1). A generalization of these results to column consecutive systems is briefly outlined in the final section.

2. Consecutive Systems

Suppose that $(N, \mathcal{S})$ is a coherent system given by components $N = \{1, 2, \ldots, n\}$ and pathsets $\mathcal{S} = \{S_1, S_2, \ldots, S_k\}$. The system $(N, \mathcal{S})$ is called *consecutive* if each set S_j contains elements consecutively appearing in N. That is to say, each S_j can be represented as the interval $[\ell_j, r_j] = \{i \in N: \ell_j \le i \le r_j\}$. Our main result will be to show that the ± 1 property holds for consecutive systems and to provide a graph-theoretic interpretation of the significance of these signed coefficients.

It will suffice to study the values of particular coefficients occurring in the expansion (1.1) for a consecutive system $\mathcal{S}$. Namely, let $d(i, i+1, \ldots, n)$ denote the coefficient of $p_i p_{i+1} \cdots p_n$ in the reduced inclusion-exclusion expansion for $\mathcal{S}$. The basic tool for calculating these coefficients derives from the theorem of total probability (Feller 1968), applied to the two states assumed by component m — either working (indicated by the event m) or failed (indicated by the event $\bar{m}$):

$$\Pr[E_\Phi] = (1 - p_m) \Pr[E_\Phi | \bar{m}] + p_m \Pr[E_\Phi | m]. \qquad (2.1)$$

Equation (2.1) will now be applied repeatedly to express the coefficient $d(i, i+1, \ldots, n)$ in terms of similar coefficients occurring in related systems.

Before indicating the general result, we first illustrate using the consecutive system $(N, \mathcal{S})$ defined by components $N = \{1, \ldots, 11\}$ and pathsets $\mathcal{S} = \{S_1, \ldots, S_6\}$ described as follows:

$$
\begin{aligned}
S_6 &= \{1, 2, 3\} \\
S_5 &= \{3, 4, 5, 6\} \\
S_4 &= \{4, 5, 6, 7\} \\
S_3 &= \{6, 7, 8\} \\
S_2 &= \{7, 8, 9\} \\
S_1 &= \{9, 10, 11\}
\end{aligned}
\tag{2.2}
$$

The reason for this numbering scheme will soon become apparent. For the system $\{S_1\}$, the inclusion-exclusion expansion gives one term $+p_9p_{10}p_{11}$ so that $d(9, \ldots, 11) = +1$ in $\{S_1\}$. For the larger system $\{S_1, S_2\}$, R_Φ is given by $p_7p_8p_9 + p_9p_{10}p_{11} - p_7p_8p_9p_{10}p_{11}$, so that $d(7, \ldots, 11) = -1$ in $\{S_1, S_2\}$. To determine the coefficient $d(6, \ldots, 11)$ for the system $\{S_1, S_2, S_3\}$ we apply (2.1) to the first component (6) of set S_3:

$$R_\Phi = (1 - p_6) \Pr[E_\Phi|\overline{6}] + p_6 \Pr[E_\Phi|6].$$

We can now apply (2.1) to the final quantity $\Pr[E_\Phi|6]$ above, relative to the second component (7) of S_3:

$$\Pr[E_\Phi|6] = (1 - p_7) \Pr[E_\Phi|6\overline{7}] + p_7 \Pr[E_\Phi|67].$$

Finally, the application of (2.1) to $\Pr[E_\Phi|67]$ relative to component 8 of S_3 yields

$$\Pr[E_\Phi|67] = (1 - p_8) \Pr[E_\Phi|67\overline{8}] + p_8 \Pr[E_\Phi|678].$$

Combining the above three equations and using $\Pr[E_\Phi|678] \equiv 1$ yields

$$
\begin{aligned}
R_\Phi = {} & (1 - p_6) \Pr[E_\Phi|\overline{6}] + p_6(1 - p_7) \Pr[E_\Phi|6\overline{7}] \\
& + p_6p_7(1 - p_8) \Pr[E_\Phi|67\overline{8}] + p_6p_7p_8.
\end{aligned}
\tag{2.3}
$$

To obtain $d(6, \ldots, 11)$ we now need only equate the coefficients of $p_6 p_7 \cdots p_{11}$ on both sides of (2.3) yielding

$$d(6, \ldots, 11) = -\{d(7, \ldots, 11|\overline{6}) + d(8, \ldots, 11|6\overline{7}) + d(9, \ldots, 11|67\overline{8})\}, \qquad (2.4)$$

where for example $d(7, \ldots, 11|\overline{6})$ is the coefficient of $p_7 p_8 \cdots p_{11}$ in $\{S_1, S_2, S_3\}$ conditioned on the failure of component 6. Examination of the collection $\{S_1, S_2, S_3\}$ shows that when component 6 fails the pathset S_3 fails and so $d(7, \ldots, 11|\overline{6})$ is precisely the coefficient $d(7, \ldots, 11)$ for $\{S_1, S_2\}$; this coefficient has already been found to be -1. Similarly, when component 8 fails, the only viable pathset in $\{S_1, S_2, S_3\}$ is S_1 so that $d(9, \ldots, 11|67\overline{8})$ is the same as $d(9, \ldots, 11)$ in $\{S_1\}$, previously found to be $+1$. When component 7 fails, the only viable pathset is S_1 so that $d(8, \ldots, 11|6\overline{7}) = 0$. Using this information in (2.4) produces

$$d(6, \ldots, 11) = -\{-1 + 0 + 1\} = 0,$$

meaning that $p_6 p_7 \cdots p_{11}$ does not appear in the reduced expansion for $\{S_1, S_2, S_3\}$.

More generally, suppose that we are interested in the system $\mathcal{S} = \{S_1, \ldots, S_j\}$ where $S_j = [\ell_j, r_j]$ and $1 \le j \le k$. Then the coefficient $d(\ell_j, \ldots, n)$ in the expansion of $\mathcal{S}$ can be found by repeatedly applying (2.1), yielding

$$\begin{aligned} d(\ell_j, \ldots, n) = -\{&d(\ell_j + 1, \ldots, n|\overline{\ell_j}) + d(\ell_j + 2, \ldots, n|\ell_j\overline{\ell_j+1}) + \ldots \\ &+ d(r_j + 1, \ldots, n|\overline{\ell_j}, \ldots, \overline{r_j})\}. \end{aligned} \qquad (2.5)$$

Notice that the above relation provides a recursion over the components of S_j. However, as we have seen in the previous example, certain of the terms appearing on the right-hand side of (2.5) are automatically 0, while others are coefficients $d(\ell_i, \ldots, n)$ for certain subsystems $\mathcal{S}' \subseteq \mathcal{S}$. It will now be shown that the relation (2.5) can in fact be further simplified to a recursion over a smaller number of objects: namely, over sets rather than components.

To this end, it is convenient to define a directed graph based on the structure of the given sets $S_1, \ldots, S_k$. It is assumed that the sets S_j are ordered so that if $i < j$ then $\ell_i > \ell_j$. (Notice that since the sets S_j are pathsets, none can contain another and so $\ell_i \neq \ell_j$ for distinct i and j.) The *consecutive union graph* $U(\mathcal{S})$ has a vertex j corresponding to each set S_j and has the directed edge (i, j), $i > j$, if $S_i \cup S_j$ is a consecutive set: i.e., $r_i + 1 \geq \ell_j$. The outdegree of vertex i in $U(\mathcal{S})$ is denoted δ_i^+. To illustrate this construction, the consecutive union graph for the system (2.2) is shown in Figure 2.1.

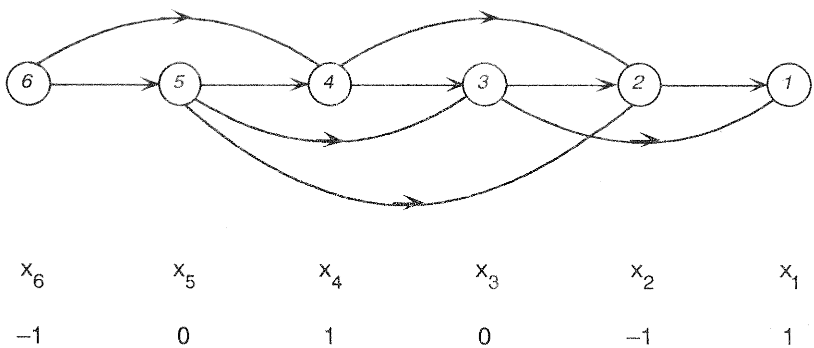

Figure 2.1

Of interest here are the specific coefficients $x_i = d(\ell_i, \ldots, n)$ relative to the subsystem $\{S_1, \ldots, S_i\} \subseteq \mathcal{S}$, for $i = 1, \ldots, k$. By virtue of (2.5) and the fact that certain coefficients in (2.5) are automatically zero, it is straightforward to obtain the recursion

$$x_1 = 1$$

$$x_i = -\sum_{r=1}^{\delta_i^+} x_{i-r}, \quad i = 2, \ldots, k. \tag{2.6}$$

Notice that relation (2.6) involves a recursion over sets, rather than over components as in (2.5). When applied to the example of Figure 2.1, the recursion produces in turn the values $x_1, \ldots, x_6$ shown in that figure. In particular, the coefficient of $p_1p_2 \cdots p_{11}$ in the inclusion-expansion for $\mathcal{S}$ must be -1, since $x_6 = -1$.

An inductive argument based on (2.6) can be used to show that all $x_i \in \{-1, 0, 1\}$, giving the promised result for the coefficients $d(\ell_i, \ldots, n)$. Section 3 shows that from this result all noncancelling terms appearing in the reduced inclusion-exclusion expansion for $\mathcal{S}$ have coefficients ± 1. Rather than describing this inductive proof, we will explore in Section 3 an alternative representation for the coefficients that quickly demonstrates this result and provides additional insight into the occurrence of ± 1 coefficients.

It should be emphasized that it is only the structure of the sets S_j captured in the consecutive union graph that has any bearing on the resulting inclusion-exclusion coefficients. That is, any two systems $\mathcal{S}_1$ and $\mathcal{S}_2$ with $U(\mathcal{S}_1) = U(\mathcal{S}_2)$ have the same coefficients appearing on corresponding terms in their inclusion-exclusion expansions. A natural next question to ask concerns what directed graphs G can arise as the consecutive union graphs of consecutive systems. We now briefly discuss this characterization.

Notice that if $G = (V, E)$ is the consecutive union graph for a consecutive system $\mathcal{S}$, then

$$\begin{aligned} (i, j) \in E &\Rightarrow (i, m) \in E, \quad i > m \geq j \\ (i, j) \in E &\Rightarrow (m, j) \in E, \quad i \geq m > j. \end{aligned} \tag{2.7}$$

This follows since the sets are ordered by their left-hand endpoints, and so they are also ordered by their right-hand endpoints. (Recall that the sets of a coherent system must be incomparable with respect to set inclusion.) Consequently, if $(i, j) \in E$ and $i > m \geq j$, then $r_i + 1 \geq \ell_j \geq \ell_m$, so $(i, m) \in E$. Also, if $i \geq m > j$ then $r_m + 1 \geq r_i + 1 \geq \ell_j$, so $(m, j) \in E$.

In other words, the vertices adjacent from (or adjacent to) a given vertex in G must be consecutive. Moreover, the relevance of components ensures that $(i, i-1) \in G$ for each $i > 1$ and so vertex $i > 1$ is adjacent to vertices $i-1, \ldots, i-\delta_i^+$. Thus, it suffices to prescribe the outdegrees δ_i^+, $i > 1$, to completely specify G; of course, $\delta_1^+ = 0$ always holds. It follows from (2.7) that the outdegrees δ_i^+ satisfy, for $i > 1$:

$$
\begin{aligned}
&1 \leq \delta_i^+ \leq i-1 \\
&\delta_{i+1}^+ \leq \delta_i^+ + 1.
\end{aligned}
\tag{2.8}
$$

In fact, these conditions completely characterize the consecutive union graphs. Therefore, the set of consecutive union graphs G on k vertices is determined by those outdegree sequences $\delta_2^+, \delta_3^+, \ldots, \delta_k^+$ satisfying (2.8). A straightforward but tedious calculation shows that the number of such graphs G on k vertices is the $(k-1)$st Catalan number: namely,

$$\frac{1}{k}\binom{2k-2}{k-1}.$$

For example, the possible outdegree sequences for $k = 4$ are shown in Table 2.1, and there are $\frac{1}{4}\binom{6}{3} = 5$ such sequences.

δ_2^+	1	1	1	1	1
δ_3^+	1	1	2	2	2
δ_4^+	1	2	1	2	3

Table 2.1

3. Linear Algebraic Viewpoint

The recursion (2.6) provides a rapid method for calculating the coefficient $x_i = d(\ell_i, \ldots, n)$ in the system $\{S_1, \ldots, S_i\}$. In order to gain an increased understanding of the occurrence of −1, 0, +1 coefficients, the equations (2.6) will be rewritten as

$$x_1 = 1$$
$$\sum_{r=0}^{\delta_i^+} x_{i-r} = 0, \quad i = 2, \ldots, k.$$

Equivalently this can be viewed as the linear system $Ax = e_1$, in which $A = (a_{ij})$ is a $k \times k$ unit lower triangular matrix with $a_{ij} = 1$ if $i - \delta_i^+ \leq j \leq i$ ($a_{ij} = 0$ otherwise) and $e_1 = (1, 0, \ldots, 0)^T$. Thus $A = I + M$, where M is the standard adjacency matrix for $U(\mathcal{S})$. Notice that the coefficient matrix A has a very special form: it has consecutive 1's in each row and in each column. Consequently, the matrix A is *totally unimodular* (Nemhauser and Wolsey 1989), from which it directly follows that the solution $x = (x_1, \ldots, x_k)^T$ has components satisfying $x_i \in \{-1, 0, +1\}$.

Moreover, when viewed in the context of linear systems, the coefficients x_i can be regarded as weights applied to the columns of A so that the resulting linear combination produces the unit vector e_1. Because each nonzero weight x_i is ± 1, we are seeking a "positive" set of columns of A whose sum yields

$$e^{+} = (1, 1, \ldots, 1, 0, \ldots, 0)^T$$
$$\uparrow$$
$$r$$

and a "negative" set of columns of A whose negated sum yields

$$e^{-} = (0, -1, \ldots, -1, 0, \ldots, 0)^T.$$
$$\uparrow$$
$$r$$

The index $r \leq k$ indicates the last position containing a nonzero element for both e^+ and e^-.

Because the columns of A have the consecutive 1's property, the only way of achieving a positive combination of columns to produce e^+ is for each column in the combination to have its consecutive 1 entries begin where the consecutive 1 entries of its predecessor column leave off. A similar remark applies to the negative combination of columns that yields e^-.

To capture the way the positive (or negative) columns fit together, it is first convenient to append a new row and column to A with the only nonzero entry being $a_{k+1,k+1} = 1$. We then define the undirected graph T(A) by associating a vertex with each column of this augmented matrix. An edge (i, j) is placed in this graph whenever column j could follow column i in the appropriate positive (or negative) linear combination: namely, whenever $j = i + \delta_i^- + 1$, where δ_i^- is the indegree of vertex i in $U(\mathcal{S})$. Addition of the new row and column to A ensures that each vertex $i \in T(A)$, $i \neq k + 1$, has a unique *successor* vertex $j = i + \delta_i^- + 1$. Thus T(A) is an undirected tree, rooted at vertex k + 1. Moreover, as the following result indicates, it is the character of the (unique) path joining vertex 1 and vertex 2 in T(A) that determines the coefficient $x_k = d(\ell_k, \ldots, n) = d(1, \ldots, n)$ in the system $\mathcal{S} = \{S_1, \ldots, S_k\}$ with components $N = \{1, \ldots, n\}$.

THEOREM 3.1: *Let* P *be the path joining* 1 *and* 2 *in* T(A). *Then path* P *contains the edge* (k, k + 1) *if and only if* $x_k \neq 0$. *Moreover, in this case* $x_k = (-1)^{|P|+1}$ *where* |P| *denotes the length (number of edges) of path* P.

Proof: (⇒) The proof is by contradiction. Suppose then that $x_k = 0$ in the solution x to $Ax = e_1$, and let r be the position of the last nonzero element occurring in both e^+ and e^-. Corresponding to e^+ is a path P^+ in T(A) connecting all the positive columns, corresponding to columns i with $x_i = +1$. Likewise, P^- denotes the path connecting all the negative columns in the linear combination. There are two possibilities for r. If $r \leq k - 1$ then the last vertex on P^+ and the last vertex on P^- are connected to vertex $r + 1 \leq k$, so that vertex k + 1 is not on the unique path P joining 1 and 2. If r = k then the last vertex of P^+ and the last vertex of P^- are connected to vertex k + 1. Since $x_k = 0$, neither P^+ nor P^- contains vertex k, so that the path P joining 1 and 2 does not include vertex k. In either case, P cannot contain the edge (k, k + 1).

(⇐) If $x_k \neq 0$ then k must be the last vertex on P^+ or P^-. Suppose it is the last vertex on P^+, so that $x_k = +1$. Let j be the last vertex on P^-. Then both j and k are joined to k + 1, whence the unique path P between 1 and 2 contains edge (k, k + 1). Moreover, because the + and – signs alternate in the linear combination $x = (x_1, \ldots, x_k)$, $|P^+| = |P^-| + 1$. As a result, $|P| = |P^+| + |P^-| + 2 = 2|P^-| + 3$, and so $(-1)^{|P|+1} = +1$, as required. For the case when k is the last vertex on P^-, we have $x_k = -1$. If j denotes the last vertex on P^+, then both j and k connect to k + 1, whence P contains (k, k + 1). Moreover $|P^+| = |P^-|$ and $|P| = 2|P^-| + 2$, so that $(-1)^{|P|+1} = -1$, as required.

While the tree T(A) has been defined relative to the consecutive union graph $U(\mathcal{S})$, it should be clear that T(A) can be directly constructed from $\mathcal{S} = \{S_1, \ldots, S_k\}$, using the prescription that $(i, i + \delta_i^- + 1)$ is an edge of T(A) for each i = 1, . . . , k. Notice that δ_i^- simply counts the number of sets S_j, j > i, for which $r_j + 1 \geq \ell_i$.

To illustrate the construction of T(A) = T($\mathcal{S}$) and the application of Theorem 3.1, consider the consecutive system on components $\{1, \ldots, 6\}$ with pathsets

$$\begin{aligned} S_4 &= \{1, 2\} \\ S_3 &= \{2, 3, 4\} \\ S_2 &= \{3, 4, 5\} \\ S_1 &= \{5, 6\}. \end{aligned} \tag{3.1}$$

The graph T(A) has $k + 1 = 5$ vertices and is shown in Figure 3.1. Since the path joining vertices 1 and 2 in T(A) does contain the edge (4, 5) and has length 3, Theorem 3.1 assures us that $x_4 = +1$ and so the term $p_1 p_2 \cdots p_6$ appears with coefficient +1 in the reduced inclusion-exclusion expansion of $\{S_1, \ldots, S_4\}$.

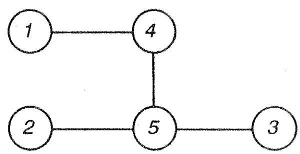

Figure 3.1

As a matter of fact, it is easy to derive other coefficients from the structure of T(A). Namely, the path joining j and j + 1 in T(A) determines the coefficient $d(\ell_k, \ldots, r_j)$ in the inclusion-exclusion expansion of $\{S_j, \ldots, S_k\}$ and hence $\{S_1, \ldots, S_k\}$. We record this fact as the following generalization of Theorem 3.1; it is proved in the same manner, concentrating instead on the linear system $Ax = e_j$.

THEOREM 3.2: *Let* P *be the path joining* j *and* j + 1 *in* T(A). *Then path* P *contains the edge* (k, k + 1) *if and only if* $d(\ell_k, \ldots, r_j) \neq 0$. *Moreover, in this case* $d(\ell_k, \ldots, r_j) = (-1)^{|P|+1}$.

To illustrate this theorem, consider again the system $\mathcal{S}$ in (3.1). Since the 2-3 path in T(A) does not contain edge (4, 5), then the term $p_1p_2 \cdots p_5$ does not appear in the expansion of $\mathcal{S}$. However, the 3-4 and 4-5 paths in T(A) do contain the required edge, yielding the terms $-p_1p_2p_3p_4$ and $+p_1p_2$ in the inclusion-exclusion expansion.

The above result allows us to deduce certain coefficients relative to the union of all components in sets $\{S_j, \ldots, S_k\}$. In order to deduce the coefficients relative to (say) $\{S_j, \ldots, S_{k-1}\}$, a new T(A) must be constructed for the system $\{S_1, \ldots, S_{k-1}\}$. By virtue of the structure (2.7) of $U(\mathcal{S})$, the new T(A) can be constructed directly from the current T(A) by simply coalescing the two vertices k and k + 1 into the new vertex k. For example, the system $\{S_1, S_2, S_3\}$ in (3.1) has the rooted tree shown in Figure 3.2(a), obtained by coalescing vertices 4 and 5 in Figure 3.1. Consequently, the application of Theorem 3.2 produces the terms $-p_2p_3p_4p_5$ and $+p_2p_3p_4$. By coalescing vertices 3 and 4 in Figure 3.2(a), we obtain the tree of Figure 3.2(b), appropriate for the system $\{S_1, S_2\}$. Likewise, coalescing vertices 2 and 3 in Figure 3.2(b) produces Figure 3.2(c), appropriate for $\{S_1\}$. The terms arising from these latter two trees are $-p_3p_4p_5p_6$ and $+p_3p_4p_5$, and $+p_5p_6$, respectively. Altogether the current analysis has accounted for 8 of the 9 terms in the reduced inclusion-exclusion expansion of $\{S_1, \ldots, S_4\}$ given in (1.2). To complete the analysis, an additional result is needed.

Thus far, the construction of T(A) has enabled the determination of coefficients d(u, u + 1, . . . , v) of terms involving consecutive components. Any other term can be decomposed into (maximal) sets A_i, each of which involves consecutive components. For example, the coefficient d(1 2 3 5 6 8 9) involves maximal consecutive sets $A_1 = \{1, 2, 3\}$,

$A_2 = \{5, 6\}$, and $A_3 = \{8, 9\}$. The following theorem shows that any such coefficient can be determined from the coefficients for its sets A_i in a multiplicative fashion.

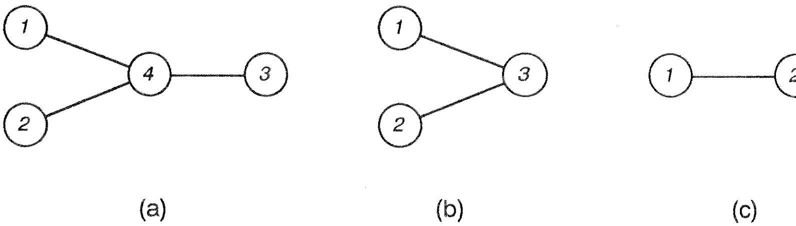

Figure 3.2

THEOREM 3.3: $d(A_1 A_2 \cdots A_r) = (-1)^{r+1} d(A_1) d(A_2) \cdots d(A_r)$.

Proof: It suffices to consider sets A_1 and A_2 and show $d(A_1 A_2) = -d(A_1)d(A_2)$. This will be done by induction on the size (number of components) of set A_1. We can assume that $A_1 = \{1, \ldots, k, \ldots, j\}$ where the first set of the system $\mathcal{S}$ is given by $\{1, \ldots, k\}$. Since the base case is easy to establish, we assume that $d(BC) = -d(B)d(C)$ holds for sets B of size less than k. By (2.5),

$$d(A_1 A_2) = -\sum_{m=1}^{k} d(m+1, \ldots, j; A_2 | 1\ 2 \cdots m-1\ \bar{m}).$$

Each subsystem on the right (defined by failed component m) is also a consecutive system, to which the inductive hypothesis can be applied, yielding

$$d(A_1 A_2) = -\sum_{m=1}^{k} -d(m+1, \ldots, j) d(A_2)$$

$$= -\left[-\sum_{m=1}^{k} d(m+1, \ldots, j) \right] d(A_2)$$

$$= -\, d(A_1)d(A_2),$$

where we have again invoked (2.5).

Since we have already established that $d(A_i) \in \{-1, 0, 1\}$, for A_i a consecutive set, Theorem 3.3 shows that every coefficient $d(A_1 A_2 \cdots A_r) \in \{-1, 0, 1\}$ as well. To complete the analysis of the example system (3.1), recall that we have earlier determined, via Theorem 3.2, the coefficients

d(1 2) = +1,	d(1 2 3 4) = −1,	d(1 2 3 4 5 6) = +1,
d(2 3 4) = +1,	d(2 3 4 5) = −1,	
d(3 4 5) = +1,	d(3 4 5 6) = −1,	
d(5 6) = +1.		

The only nonconsecutive set having a nonzero coefficient is {1, 2, 5, 6} so that

$$d(1\ 2\ 5\ 6) = -\, d(1\ 2)d(5\ 6) = -1.$$

Therefore, the reduced inclusion-exclusion expansion is precisely as given in (1.2).

4. Discussion

The objective of this paper has been to study the cancellation properties that occur upon applying the inclusion-exclusion principle to the pathsets of a coherent system. One motivation for this study is the groundbreaking work of Satyanarayana and Prabhakar (1978), who established a ± 1 property for the two-terminal reliability of directed graphs G, when expressed in terms of the s-t paths. The same cancellation phenomenon also occurs for K-terminal reliability in directed graphs. For example, Boesch et al. (1990) establish the corresponding K-terminal result, which again involves certain acyclic subgraphs of the original graph G.

The present work considers the more general context of coherent systems and establishes a complementary ± 1 property for consecutive systems $\mathcal{S}$. This result in turn generalizes the work of Kossow and Preuss (1989) which examines the case of "consecutive k-out-of-n" systems, special types of consecutive systems in which every pathset S_j has the same cardinality. A key ingredient in analyzing the consecutive system $\mathcal{S}$ is the directed graph $U(\mathcal{S})$, which captures the essential incidence structure of the given sets $S_j \in \mathcal{S}$. Interestingly enough, we have seen that methods for calculating the coefficients of the inclusion-exclusion expansion can be based on either the outdegrees of vertices in $U(\mathcal{S})$ or on their indegrees.

The consecutive systems considered here should more precisely be called *row consecutive systems*. That is, the components of such a system can be ordered so that each set consists of consecutively numbered components. Alternatively, one could consider the notion of a *column consecutive system*, in which the sets can be ordered so that each component occurs in consecutively numbered sets. It can be seen that every row consecutive system is a column consecutive system; in fact the ordering of sets $S_j = [\ell_j, r_j]$ by their left-hand endpoints ℓ_j places the sets in the required order. To verify this, suppose that $u \in S_i$ and $u \in S_j$, but $u \notin S_m$ for $i > m > j$. Then $\ell_m < \ell_j \leq u$ and so $r_m < u$; also $\ell_i < \ell_m$ and $r_m < u \leq r_i$. This implies $S_m \subset S_i$, contradicting the coherence of $\mathcal{S}$. The following system is a column consecutive system but is not a row consecutive system, showing that the former concept is indeed more general than the latter.

$$\begin{aligned} S_5 &= \{1, 2, 4\} \\ S_4 &= \{2, 3, 4, 5\} \\ S_3 &= \{4, 5, 6\} \\ S_2 &= \{6, 7, 9\} \\ S_1 &= \{7, 8, 9\} \end{aligned} \tag{4.1}$$

We now briefly indicate how the results obtained for row consecutive systems can be extended with slight modification to column consecutive systems. For a column consecutive system $\mathcal{S}$, let $I_u = [a_u, b_u]$, with $a_u \le b_u$, denote the interval of indices j such that S_j contains component u. It is convenient to order the components $u \in N = \{1, \ldots, n\}$ by the endpoints a_u: namely, $u < v \Rightarrow a_u \ge a_v$. Under this ordering of components, each set S_j consists of a union of certain subintervals of N; let the first such subinterval (containing the smallest numbered consecutive components) be denoted $[\ell_j, r_j]$. Then the consecutive union graph $U(\mathcal{S})$ for the column consecutive system $\mathcal{S}$ is defined as before, but relative to this "first" subinterval: i.e., (i, j) is an edge of $U(\mathcal{S})$ if $i > j$ and $r_i + 1 \ge \ell_j$. Once the graph $U(\mathcal{S})$ has been so defined, the development of the previous sections is directly applicable. In particular, the ± 1 property still holds and Theorems 3.1-3.3 can be employed to deduce the signs of the appropriate coefficients.

To illustrate the case of a column consecutive system, consider the system defined by (4.1). Notice that the sets S_j have been ordered so that sets containing a given component u occur consecutively. The corresponding intervals $I_u = [a_u, b_u]$ are thus

$$I_1 = [5, 5], \quad I_2 = [4, 5], \quad I_3 = [4, 4],$$
$$I_4 = [3, 5], \quad I_5 = [3, 4], \quad I_6 = [2, 3],$$
$$I_7 = [1, 2], \quad I_8 = [1, 1], \quad I_9 = [1, 2].$$

The components u have been numbered so that the sequence $\{a_u\}$ is nonincreasing. The graph $U(\mathcal{S})$ associated with this system is shown in Figure 4.1, together with the coefficients x_j. Since, for example, $x_5 = -1$ then the term $-p_1 p_2 \cdots p_9$ occurs in the simplified inclusion-exclusion expansion of (4.1).

Figure 4.2 shows the tree $T(\mathcal{S})$ derived from this system. Since the 1-2 path in the tree contains edge (5, 6), we conclude that $x_5 \ne 0$; moreover, this path has length 4 and Theorem 3.1 gives $x_5 = -1$, as expected. Since the 3-4, 4-5, and 5-6 paths in $T(\mathcal{S})$ contain edge (5, 6), application of Theorem 3.2 produces the additional terms $+p_1p_2p_3p_4p_5p_6$, $-p_1p_2p_3p_4p_5$, and $+p_1p_2p_4$, respectively.

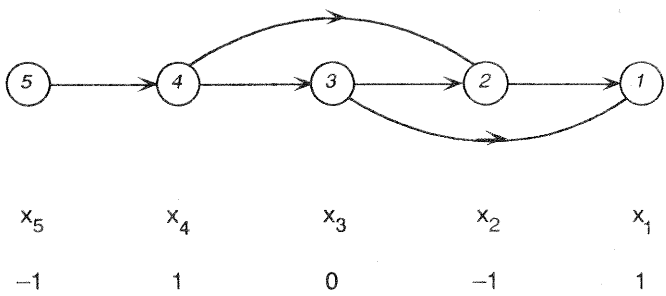

Figure 4.1

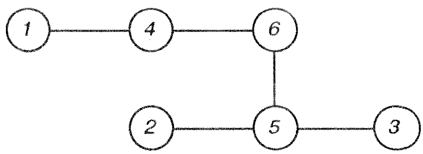

Figure 4.2

Acknowledgements

This work has been supported by the Air Force Office of Scientific Research under grant AFOSR-89-0071 and the National Science Foundation under grant DMS-9000645.

5. References

Boesch, F. T., Satyanarayana, A., and Suffel, C. L. (1990). Some alternate characterizations of reliability domination. *Probability in the Engineering and Informational Sciences* **4**, 257-276.

Feller, W. (1968). An Introduction to Probability Theory and Its Applications, Vol. 1. John Wiley, New York.

Kossow, A. and Preuss, W. (1989). Reliability of consecutive-k-out-of-n:F systems with nonidentical component reliabilities. *IEEE Trans. Reliability* **38**, 229-233.

Nemhauser, G. L. and Wolsey, L. A. (1988). Integer and Combinatorial Optimization. John Wiley, New York.

Satyanarayana, A. and Prabhakar, A. (1978). New topological formula and rapid algorithm for reliability analysis of complex networks. *IEEE Trans. Reliability* **R-27**, 82-100.