



## An Exponential Family of Probability Distributions for Directed Graphs: Comment

Ronald L. Breiger

*Journal of the American Statistical Association*, Vol. 76, No. 373. (Mar., 1981), pp. 51-53.

Stable URL:

<http://links.jstor.org/sici?sici=0162-1459%28198103%2976%3A373%3C51%3AAEFOPD%3E2.0.CO%3B2-M>

*Journal of the American Statistical Association* is currently published by American Statistical Association.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/astata.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

RONALD L. BREIGER\*

Holland and Leinhardt have long been among the most innovative and most productive researchers in the field of social network analysis. I and my colleagues were gratified to have some aspects of the empirical findings derived from our network theory supported rather strikingly in their accompanying article pioneering new approaches to binary matrices. At the same time we were surprised, given our theoretical goals, at the awkwardness of certain aspects of Holland and Leinhardt's approach.

In the "empirical example" section of their paper, Holland and Leinhardt's  $p_1$  distribution provides a null model of an idealized structure. A null model for social networks should be well defined, plausible, and clearly capable of rejection in applications where definite structure exists. The  $p_1$  model fulfills these criteria admirably, as demonstrated in Holland and Leinhardt's Section 4. On the other hand, if one's goal is an explicit model of social structure, a modeling context that goes beyond the null hypothesis is required. Some specific directions are provided below.

It seems anomalous to generate predictions for square binary arrays from a statistical framework that in the given instance is unable to predict any one of the possible array events with a "probability" over one-half (see Holland and Leinhardt's Table 4). As the authors point out, "most of the [observed] relational ties in this example are 'unexpected,'" in the sense that the 38 *residuals* greater than +.70 are a subset of the 56 observed data ties. (Each residual greater than +.70 results, of course, from matching an observed value of "1" with a predicted value less than .30). I have no doubt that the Holland-Leinhardt stance on what is interesting will pay off in drawing generalizations across the enormous array of square binary matrices. But I do not see why a sociologist, theorist or field worker, is much concerned for social network analysis unless it leads to models of *explicit* social structure.

This brings me to my second point. Their pioneering development of statistical tools leads Holland and Leinhardt away from a focus on *patterns* of relationships and toward a framework within which a structure of persons and relations is decomposed into a heap of pairs, which I argue is inappropriate for social networks. Indeed, the frequent collaborator with Holland and Leinhardt in past

publications, the sociologist James A. Davis, writes in his survey article on the previous Davis-Holland-Leinhardt studies of his "regret [over] our slide from global structure to microanalysis. . . I wish we had been able to move upward to say things about groups as a whole instead of retreating to a perspective from which a *triad* looks as large and complicated as intergalactic space" (Davis 1979, p. 60).

Put these points together: Holland and Leinhardt's goal is not to replicate entries in the network proper (the interior cell values and the structure of present and absent ties), and no attempt is made to discern an overall contextual pattern. I now illustrate that simultaneous reversal of these two weaknesses (from my point of view) can yield results.

Blockmodel analysis "requires the aggregation of a population's members into distinct sets, termed *blocks*, such that each set is treated homogeneously across all networks simultaneously" (Breiger 1979, p. 23). From this perspective, the basic point is that each of the nine submatrices resulting from our three-block partition should be internally homogeneous if the blockmodel analysis makes sense. Provisionally adopting Holland and Leinhardt's  $p_1$  model as a characterization of a homogeneous structure, I will therefore apply their model separately to *each* of the nine submatrices identified in White, Boorman, and Breiger (1976, p. 750) and reported in Holland and Leinhardt's Figure 5. If the  $p_1$  model fits the *full* matrix poorly, but fits each of the blocked submatrices well, then the improvement in fit is due to the explicit structure uncovered by the blockmodel. (Such improvement should not be confused with a "test" of the blockmodel; see the last paragraph of this Comment.) The underlying "principle of internal homogeneity" as a criterion for assessing the fit of a simultaneous partition of the rows and columns of a square array is proposed in Breiger (1981) and is implemented there in log-linear models for counted data in social mobility tables.

Fitted values obtained from applying Holland and Leinhardt's  $p_1$  model separately to each of the nine blocked submatrices are reported in the accompanying Table 1. Treating only those entires greater than .50 in Table 1 as predictions of ties, one obtains the binarized fitted values of the accompanying Table 2. Comparing these with the observed data (Holland and Leinhardt's Figure 5), the

\* Ronald L. Breiger is Associate Professor of Sociology, Harvard University, Cambridge, MA 02138. This research was supported by National Science Foundation Grants SOC76-24394 and SES80-08658.

Table 1. Fitted Expected Values From the Model Described in the Text<sup>a</sup>

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	$\hat{p}_{i+}$
1	0	84	65	29	65	29	29	0	0	0	0	0	0	0	100	0	0	0	4
2	0	0	78	48	81	45	48	0	0	0	0	0	0	0	0	0	0	0	3
3	0	78	0	23	58	19	23	50	0	0	50	0	0	0	0	0	0	0	3
4	0	88	69	0	73	33	37	0	0	0	0	0	0	0	0	0	0	0	3
5	0	91	76	46	0	41	46	0	0	0	0	0	0	0	0	0	0	0	3
6	0	72	44	17	50	0	17	50	0	0	50	0	0	0	0	0	0	0	3
7	0	88	69	37	73	33	0	0	0	0	0	0	0	0	0	0	0	0	3
8	0	0	0	0	0	0	0	0	58	77	22	43	77	22	0	0	0	0	3
9	0	0	0	0	0	0	0	71	0	48	8	18	48	8	0	100	0	0	3
10	0	0	0	0	0	0	0	89	53	0	21	41	76	21	0	0	0	0	3
11	0	0	0	0	0	0	0	83	45	65	0	29	65	13	0	0	0	0	3
12	0	0	0	0	0	0	0	85	46	69	16	0	69	16	0	0	0	0	3
13	0	0	0	0	0	0	0	89	53	76	21	41	0	21	0	0	0	0	3
14	0	0	0	0	0	0	0	83	45	65	13	29	65	0	0	0	0	0	3
15	0	100	0	0	0	0	0	0	5	58	0	0	30	0	0	1	1	99	3
16	0	0	0	0	0	0	0	0	100	15	0	0	6	0	99	0	100	100	4
17	0	0	0	0	0	0	0	0	5	58	0	0	30	0	0	100	0	100	3
18	0	0	0	0	0	0	0	0	5	58	0	0	30	0	1	99	100	0	3
$\hat{p}_{+j}$	0	6	4	2	4	2	2	6	4	6	2	2	5	1	2	3	2	3	56 = $\hat{p}_{++}$

<sup>a</sup> The decimal points have been left off the entries in the body of the table. The marginals have not been so altered.

reader will find 22 errors of prediction: 8 cells predicted to be "1" are actually "0," and 14 cell predictions err conversely.

The reader may compare these predictions directly with those of Holland and Leinhardt (their Table 4). To aid intuition, I report (in Table 3 of this Comment) Holland and Leinhardt's predicted values above and below .30 (of which 56 are greater than .30; no predicted value is greater than .50; my choice of .30 is arbitrary, but the reader may experiment with other choices). Treating only those entries greater than .30 in Table 3 as predictions of ties, the reader will find 74 errors of prediction (37 cells predicted to be "1" are actually "0," and 37 cell predictions err conversely). What is more, Table 3 emphasizes the patterning of ties predicted from Holland

and Leinhardt's model. Of these 56 cells with the highest expected values, 51 appear in just three columns (each column consisting entirely of fitted values above .30, except for self-choices), and 33 of these correspond to the observed absence of ties. The contrast of Tables 2 and 3 suggests the implications of a model of definite structure.

Narrowing the gap between a "suggestion" such as this one and a rigorous statistical procedure is a goal that Holland and Leinhardt's paper advances. Clearly, several entire submatrices in the accompanying Table 1 are fitted exactly under the model, which (even apart from this fact) exhibits fewer degrees of freedom than Holland and Leinhardt's  $p_1$  model applied to the *full* matrix. I am not prepared to assess the goodness of fit of the model for the blocked data. The model illustrated here (Table

Table 2. Fitted Expected Values Greater Than .50 in Table 1 of This Comment<sup>a</sup>

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
2	0	0	1	0	1	[0]	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	[1]	0	[0]	[0]	0	0	0	0	0	0	0	0	0	0
4	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	1	[1]	[0]	0	[0]	0	0	0	0	0	0	0	0	0	0	0	0
6	0	1	0	0	0	0	[0]	0	0	0	0	[0]	0	0	0	0	0	0
7	0	[1]	1	[0]	1	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0
9	0	0	0	0	0	0	0	1	0	0	[0]	0	0	0	0	1	0	0
10	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0
11	0	0	0	0	0	0	0	1	[0]	[1]	0	[0]	[1]	0	0	0	0	0
12	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0
13	0	0	0	0	0	0	0	1	[1]	1	0	0	0	[0]	0	0	0	0
14	0	0	0	0	0	0	0	[1]	0	1	0	[0]	1	0	0	0	0	0
15	0	1	0	0	0	0	0	0	0	[1]	0	0	[0]	0	0	0	0	1
16	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	1
17	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1
18	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0

<sup>a</sup> Errors of prediction (compare the observed data in Holland and Leinhardt's Figure 5) are enclosed in square brackets.

Table 3. Fitted Expected Values Greater Than .30 in Table 4 of Holland and Leinhardt<sup>a</sup>

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0	1	[0]	0	[0]	0	0	[1]	0	[1]	0	0	[1]	0	[0]	0	0	0
2	0	0	[0]	0	[0]	[0]	0	[1]	0	[1]	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	[0]	1	0	[1]	0	0	0	0	0	0	0	0
4	0	1	[0]	0	[0]	0	0	[1]	0	[1]	0	0	0	0	0	0	0	0
5	0	1	0	[0]	0	[0]	0	[1]	0	[1]	0	0	0	0	0	0	0	0
6	0	1	0	0	0	0	[0]	[1]	0	[1]	0	0	0	0	0	0	0	0
7	0	[1]	[0]	[0]	[0]	0	0	[1]	0	[1]	0	0	0	0	0	0	0	0
8	0	[1]	0	0	0	0	0	0	[0]	1	0	0	[0]	0	0	0	0	0
9	0	[1]	0	0	0	0	0	1	0	[1]	[0]	0	0	0	0	[0]	0	0
10	0	[1]	0	0	0	0	0	1	[0]	0	0	0	[0]	0	0	0	0	0
11	0	[1]	0	0	0	0	0	1	[0]	[1]	0	[0]	0	0	0	0	0	0
12	0	[1]	0	0	0	0	0	1	0	1	0	0	[0]	0	0	0	0	0
13	0	[1]	0	0	0	0	0	1	0	1	0	0	0	[0]	0	0	0	0
14	0	[1]	0	0	0	0	0	[1]	0	1	0	[0]	[0]	0	0	0	0	0
15	0	1	0	0	0	0	0	[1]	0	[1]	0	0	[0]	0	0	0	0	[0]
16	0	[1]	[1]	0	[1]	0	0	[1]	1	[1]	0	0	[1]	0	[0]	0	[0]	[0]
17	0	[1]	0	0	0	0	0	[1]	0	1	0	0	0	0	0	[0]	0	[0]
18	0	[1]	0	0	0	0	0	[1]	0	1	0	0	0	0	0	[0]	[0]	0

<sup>a</sup> Errors of prediction (compare the observed data in Holland and Leinhardt's Figure 5) are enclosed in square brackets.

1) appears to be a natural generalization of the Holland-Leinhardt model that may prove useful in the analysis of matrices whose rows and columns have been partitioned, and where homogeneity of structure is hypothesized to be confined *within* the resulting submatrices. (The reader's attention is also called to the similarly motivated but distinct work of Fienberg and Wasserman in their important paper, 1979, Sec. 6.) The major technical considerations raised by these observations notwithstanding, there is also a crucial substantive point to be made. Despite the utility of measures of deviation from a well-defined null structure, sociological analysis is most in need of *explicit* models of the *overall* structures of observed populations.

One further main theoretical issue deserves recognition. Structural equivalence, an algebraic homomorphism concept rather than a labeling of each node with a derived statistical measure, is needed to provide an overall structural perspective and complement local concepts such as cliques. Holland and Leinhardt seem to me to be so literal about adopting standard parametric techniques that they confound any possibility of introducing such ideas (cf. White 1977). None of their techniques seem to have any relation to the semigroup-algebraic side of our theory-based analysis of multiple networks (Boorman and White 1976, Breiger and Pattison 1978, Mandel 1978, Pattison 1980). Nonetheless, this paper by Holland and Leinhardt constitutes a major contribution to the development of

statistical models for digraph data, and will serve as an enduring foundation for future work.

[Received August 1980.]

## REFERENCES

- BOORMAN, S.A., and WHITE, H.C. (1976), "Social Structure From Multiple Networks: II. Role Structures," *American Journal of Sociology*, 81, 1384-1446.
- BREIGER, R.L. (1979), "Toward an Operational Theory of Community Elite Structures," *Quality and Quantity*, 13, 21-57.
- (1981), "The Social Class Structure of Occupational Mobility," *American Journal of Sociology*, 87, in press.
- BREIGER, R.L., and PATTISON, P.E. (1978), "The Joint Role Structure of Two Communities' Elites," *Sociological Methods and Research*, 7, 213-226.
- DAVIS, JAMES A. (1979), "The Davis/Holland/Leinhardt Studies: An Overview," in *Perspectives on Social Network Research*, eds. P.W. Holland and S. Leinhardt, New York: Academic Press, 51-62.
- FIENBERG, STEPHEN E., and WASSERMAN, STANLEY (1979), "Categorical Data Analysis of Single Sociometric Relations," Technical Report No. 362, University of Minnesota, Dept. of Applied Statistics.
- MANDEL, MICHAEL J. (1978), *Roles and Networks: A Local Approach*, AB honors thesis, Harvard University, Applied Mathematics.
- PATTISON, P.E. (1980), "Equating the 'Joint Reduction' With Block-model Common Role Structure: Reply to McConaghy," *Sociological Methods and Research*, 9, in press.
- WHITE, H.C. (1977), "Probabilities of Homomorphic Mappings From Multiple Graphs," *Journal of Mathematical Psychology*, 16, 121-134.
- WHITE, H.C., BOORMAN, S.A., and BREIGER, R.L. (1976), "Social Structure From Multiple Networks: I. Blockmodels of Roles and Positions," *American Journal of Sociology*, 81, 730-780.