

Appendix

SURVEY METHODOLOGY

Our survey was mailed to the faculty in the departments of biology, biochemistry, chemistry, computer science, engineering, and physics at a large private Midwestern University in May of 1997. Emeritus, adjunct, and visiting faculty members were not contacted because they are normally peripheral to department activity. Data collection was completed by September 1997. The survey was diskette-based and computer-administered and took approximately 30 minutes to complete. The electronic survey guided respondents through the items (with appropriate branching) and created a data file on the diskette, which was returned to us. This procedure reduced errors of data entry by respondents and data entry assistants, and ensured that all respondents answered all items. All faculty members were sent a cover letter that briefly described the purpose of the study, assured confidentiality, gave references to our previous work in scientific journals of the hard sciences and at the National Science Foundation, and described the instructions for the use of the survey. Included along with the cover letter were a diskette and a pre-addressed envelope for returning the diskette. Table A1 describes the characteristics of our sample.

Faculty members who did not respond in three weeks or who used Macintosh computers were asked to set up a time for a telephone interview by a trained research assistant who entered responses in real time into a PC. Unfortunately, the biology department used Macintosh computers exclusively, which lowered the response rate of this department. Table A1 displays the response rate, which averaged 56% and ranged from 38% to 76%. This response rate is similar to other network surveys of organizations (Podolny and Baron, 1997; Burt 1992; Granovetter, 1973).

Table A1: *Sample characteristics and response rate*

Department	Number of Faculty			Number of Respondents		
	Men	Women	% Women	Men	Women	Response rate
Biology	28	6	21.4	12	1	38 %
Biochemistry	23	13	57	14	9	63 %
Chemistry	25	2	8	17	2	76 %
Computer Science	8	0	0	4	0	50 %
Engineering	33	5	15	22	1	57 %
Physics	27	3	11	16	0	53 %
Total	144	29	20	84	13	

We explored whether there was non-response bias by examining differences in the demographic background (age, gender, rank, and income) of faculty members who responded immediately and those who were interviewed by phone. There appeared to be few differences across these categories using O'Brien's method (Stata, 1996). There also did not appear to be a difference between responders and non-responders in terms of prestige of Ph.D.-granting institution (most had received degrees from prestigious schools), although full professors were more likely to respond after the first mailing. Lastly, the feedback we received from faculty members who could not participate indicated that it was due to randomly distributed factors (deadlines, annual meetings, grants, travel, etc.).

REGRESSION ANALYSIS

We used an ordered logit model to estimate the effect of social capital on the rate of publication. This model estimates the relationship between a categorical and an ordered dependent variable – 'no output,' 'low output,' 'medium output,' and 'high output' – and a set of

response rate

Number of Respondents		
Men	Women	Response rate
12	1	38 %
14	9	63 %
17	2	76 %
4	0	50 %
22	1	57 %
16	0	53 %
84	13	

response bias by examining round (age, gender, rank, and ended immediately and those appeared to be few differences method (Stata, 1996). There between responders and non-granting institution (most had ols), although full professors t mailing. Lastly, the feedback could not participate indicated ed factors (deadlines, annual

rate the effect of social capital el estimates the relationship dependent variable – 'no output,' high output' – and a set of

independent variables (Greene, 1993). An ordinal variable is a variable that is categorical and ordered, for instance, 'no output,' 'low output,' 'medium output,' and 'high output.' In an ordered logit, an underlying probability score of how a one unit change in an independent variable affects the change in probability of intensity of output is estimated as a linear function of the independent variables and set of cut points. The probability of observing outcome i corresponds to the probability that the estimated linear function, plus random error, is within the range of the cut points estimated for the outcome:

$$\Pr(\text{outcome } j = i) = \Pr(K_{i-1} < B_1x_1 + \dots + B_kx_k + u_j \leq K_i)$$

One estimates the coefficients B_1, B_2, \dots, B_k along with the cut points K_1, K_2, \dots, K_{I-1} , where I is the number of possible outcomes. All of this is a generalization of the ordinary two-outcome logit model. The ordered logit predictions are then the probability that outcome $j + u_j$ lies between a pair of cut points K_{i-1} and K_i (Stata 1996).

The attractive modeling feature of the ordered logit is that the substantive numerical values of the dependent variable are unimportant (Greene, 1993). In ordinary regression, arbitrarily assigning the number values of 4 to the 'high output' category, 3 to the 'medium output' category, 2 to the 'low output' category, and so on is inappropriate because different numeric values (say 10 versus 8 for 'high output') would obtain different estimates. This is not true in an ordered logit model. All that is necessary is that larger numbers correspond to more intense outcomes or levels of usage.

DEPENDENT VARIABLE

The criterion variable, research productivity, was measured by asking respondents to indicate their number of publications using the following interval scale: (1) no publications, (2) 1–2, (3) 3–4, (4) 5–10, (5) 11–20, (6) 21–50, and (7) more than 50 publications. This question was repeated for (a) articles published, (b) book chapters published, and (c) books published, in order to capture the range of publications. The

value for each respondent on (a), (b), and (c) was summed and divided by three to derive an average productivity score. A simple sum of (a) to (c) produced similar results to those reported in terms of significance and size of coefficients. The seven preceding categories were united into four categories by joining adjacent categories and dropping the first two categories, which had no observations. The use of a four-category dependent variable made the results more intuitive and produced results that were similar to the more complex eight-category model. Table A2 displays the distribution of the dependent variable for all faculty and untenured faculty. The numerical value '0' corresponds to the low output category (Less than 4), '1' corresponds to the low-to-medium category (5 to 10), '2' corresponds to the medium-to-high category (11 to 20), and '3' corresponds to the high output category (more than 21). Table 12.4 displays three cut points because one of the four categories is a reference category, which in this analysis corresponds naturally to category 0 (see Table A2).

Table A2: *Distribution of research productivity of faculty*

Number of publications reported	All faculty		Untenured faculty	
	Men	Women	Men	Women
Less than 4	3%	7%	7%	10%
5 to 10	12%	43%	20%	45%
11 to 20	28%	25%	45%	25%
More than 21	57%	25%	28%	20%

INDEPENDENT VARIABLES

Token Overload, Power Imbalance, Number of Strong Ties, and Number of Bridge Ties were measured using the items and scales described in Chapter 12. We squared the number of strong ties to examine our hypothesis that an intermediate level of strong ties is

c) was summed and divided by core. A simple sum of (a) to (c) led in terms of significance and eight categories were united into three and dropping the first two. The use of a four-category more intuitive and produced a complex eight-category model. The dependent variable for all hierarchical value '0' corresponds to '1' corresponds to the low-to-mediums to the high output category cut points because one of the categories, which in this analysis (Table A2).

tivity of faculty

Untenured faculty	
Men	Women
7%	10%
20%	45%
45%	25%
28%	20%

Number of Strong Ties, and used using the items and scales the number of strong ties to mediate level of strong ties is

positively associated with research productivity. Number of Co-authors was simply the number reported by the respondent.

CONTROL VARIABLES

Following prior research, we controlled for human capital and demographic factors with the following measures (Seashore *et al.*, 1989; Cole, 1992). We created an indicator variable, Gender, which was coded 0 for female and 1 for male, and Tenured which was coded 1 for tenured and 0 for non-tenured. Professional Age measured number of years since Ph.D. and Age in Years measured age of respondent. Cole (1992) reported that Professional Age has been found to have a statistically significant but small positive effect on getting grants. Age in Years has similarly been the center of many studies of scientific productivity. Cole (1992) noted that scientific creativity is commonly believed to decline with age, but noted that the empirical evidence is mixed. Research Budget Level controls for the level of the faculty's research budget (in dollars divided by 100), a factor positively associated with research productivity (Seashore *et al.*, 1989; Cole, 1992). Finally, we created an indicator variable called Post-Doc (1=Yes) to control for the positive effect of a post-doctorate fellowship on research productivity (Long and McGinnis, 1981).

Two variables were added to control for the synchronicity problem, which is the condition that the number of publications reported was probably affected by ties that existed prior to the ties reported at the time of the survey. One way to deal with this problem is to assess the historical stability of the network over the period of publishing reported. Network Turnover measures the level of turnover in an individual's network over the past two years. Another control variable that attempts to mitigate the synchronicity problem is Average Tie Duration, a variable that measures the average time the respondent has known the contacts named. The assumption underlying these controls is that contacts that have endured more than three years are likely to be long-term ties and in existence over the course of the reported number of publications.