COCS 6323: Statistical Methods in Research

Group Project

Group 2

Department of Computer Science

University of Houston

April 4, 2019

Contents

List of Tables

List of Figures

1 Contribution

Member	**Contribution**
Bradley Macdonald	Analyze regression models of Figure 4
Tung Huynh	Preprocess Data, create regression models of Table S2, Table S3
Yifan Zhang	Preprocess Data, and draw plot of Figure 4

Table 1: Contribution of group members of the second milestone

2 Figure 4

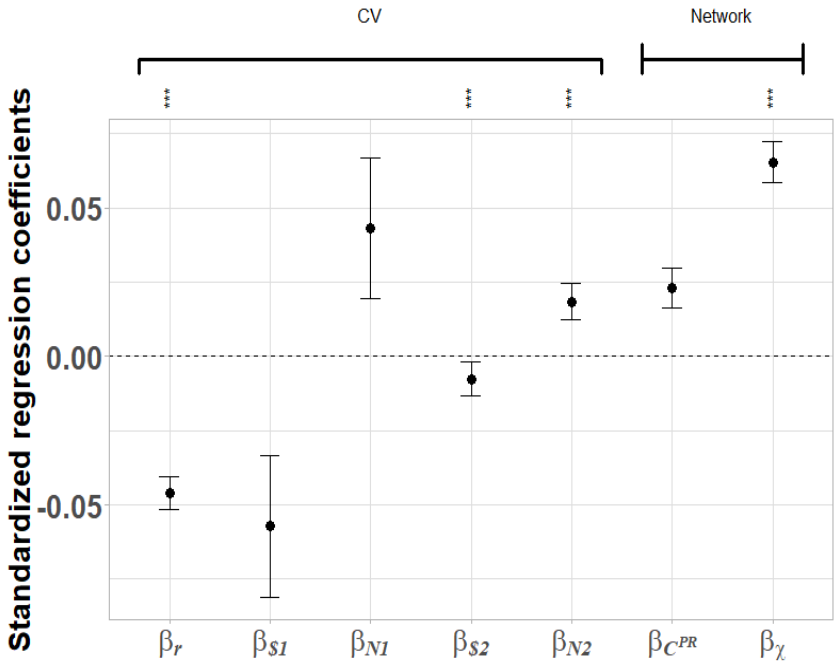

Figure 1: Career cross-sectional regression model

Regression coefficients from Equation 1 were standardized and are presented with error bars indicating standard error. Prior to regression analysis, dummy variables were created for researcher discipline, as well as the five year period in which they published their first genomic publication. Zero's in the dataset (except for in dummy variables) were replaced by 0.00001 in order to allow for logarithmic transformations to be applied to appropriate numeric variables. The location which the information was derived from is indicated by the labels above the plot, with "CV" indicating the information was collected from the researcher's CV and "Network" indicating that the information originated from the researcher's collaboration network. Standardized beta coefficients are shown, to aid with the comparison across covariates. The plot shows the impact of changes in covariate, by showing the change in the dependant variable if the covariate were alterred by one standard deviation. "***" symbols were inserted above the covariates which showed statistical significance where the p-value was less than or equal to 0.001. The significant results indicate a covariate which impacts the dependant variable to a degree of statistical significance when alterred by one standard deviation.

3 Supplementary Table S2

	CV		CV + Network		CV + Network [Standardized]	
CV parameters						
Department rank, β_r	-0.052***	(0.006)	-0.047***	(0.006)	-0.056***	(0.007)
Productivity (h-index), β_h	1.857***	(0.018)	1.866***	(0.018)	1.179***	(0.012)
Total NSF funding, $\beta_{\$1}$	-0.004*	(0.002)	-0.005**	(0.002)	-0.031**	(0.012)
# of NSF grants , β_{N1}	0.018	(0.012)	0.010	(0.012)	0.011	(0.013)
Total NIH funding, $\beta_{\$2}$	0.015***	(0.003)	0.018***	(0.003)	0.072***	(0.018)
# of NIH grants, β_{N2}	-0.062***	(0.016)	-0.054**	(0.017)	-0.060**	(0.018)
Network parameters						
PageRank Centrality, $\beta_{\zeta PR}$			0.041**	(0.014)	0.026**	(0.009)
Cross-displinary, β_χ			0.567***	(0.061)	0.085***	(0.009)
Discipline (O) dummy	Y		Y		Y	
5-year cohort ($y_{i,5^0}$) dummy	Y		Y		Y	
Constant	1.400***	(0.233)	1.708***	(0.271)	7.743***	(0.216)
n	4,190		3,900		3,900	
adj. R^2	0.883		0.882		0.882	

Standard errors in parentheses below estimate * p $\leq$ 0.05, ** p $\leq$ 0.01, *** p $\leq$ 0.0001

Table 2: Career data set: Poolel cross-sectional model

4 Supplementary Table S3

	(a) C^{PR}	(b) C^{B}	(c) C^{D}	(d) β_{N1}, β_{N2}	(e) β_{r}
CV parameters					
Department rank, β_r	-0.047***	-0.042***	-0.044***	-0.046***	
	(0.006)	(0.006)	(0.006)	(0.006)	
Productivity (h-index), β_h	1.866***	1.901***	1.848***	1.862***	1.892***
	(0.018)	(0.019)	(0.018)	(0.018)	(0.018)
Total NSF funding, $\beta_{\$1}$	-0.005**	-0.004*	-0.004*	-0.003**	-0.004*
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
# of NSF grants , β_{N1}	0.010	0.009	0.005		0.004
	(0.012)	(0.012)	(0.012)		(0.012)
Total NIH funding, $\beta_{\$2}$	0.018***	0.012***	0.012***	0.003*	0.012***
	(0.003)	(0.003)	(0.003)	(0.001)	(0.003)
# of NIH grants, β_{N2}	-0.054**	-0.056**	-0.055**		-0.052**
	(0.017)	(0.017)	(0.017)		(0.017)
Network parameters					
PageRank Centrality, $\beta_{\zeta PR}$	0.041**			0.042**	0.057***
	(0.014)			(0.014)	(0.014)
Betweeness Centrality, $\beta_{\zeta B}$		-0.0003			
		(0.005)			
Degree Centrality, $\beta_{\zeta D}$			0.052***		
			(0.010)		
Cross-displinary, β_{χ}	0.567***	0.560***	0.526***	0.579***	0.552***
	(0.061)	(0.062)	(0.061)	(0.061)	(0.061)
Discipline (O) dummy	Y	Y	Y	Y	Y
5-year cohort ($y_{i,5^0}$) dummy	Y	Y	Y	Y	Y
Constant	1.708***	1.204***	1.345***	1.711***	1.617***
	(0.271)	(0.225)	(0.226)	(0.270)	(0.272)
n	3,900	3,387	3,900	3,900	3,900
adj. R^2	0.882	0.873	0.883	0.882	0.881

Standard errors in parentheses below estimate * p $\leq$ 0.05, ** p $\leq$ 0.01, *** p $\leq$ 0.0001

Table 3: Career data set: Poolel cross-sectional model - Robustness check