Project - Part 3

Ryan Thackston

5/2/2021

Table S1:

X.XSACIP <chr>	XSA.1 <chr>	XCIP.1 <chr>	XSA.CIP <chr>
1.0461 *** (0.00160)	1.0327 *** (0.00085)	1.0205 *** (0.00150)	1.0609 *** (0.0021
1.415 *** (0.03325)	0.978 *** (0.01180)	1.2232 *** (0.02040)	1.3077 *** (0.0315
1.5616 *** (0.03400)	0.8968 *** (0.01010)	1.821 *** (0.02765)	1.6387 *** (0.0361
4.858 *** (0.15630)	4.6775 *** (0.07935)	0.9292 *** (0.02260)	4.8956 *** (0.1591
12.0039 *** (0.31475)	1.2108 *** (0.04920)	3.0284 *** (0.11255)	4.5694 *** (0.2750
NA	1.2944 *** (0.02295)	NA	NA
NA	NA	1.206 *** (0.01070)	NA
	0.9488 * (0.03800)	0.754 *** (0.02430)	0.7163 *** (0.0429
	0.9131 *** (0.04340)	0.38 *** (0.01655)	0.3974 *** (0.0274
	0.9422 * (0.04515)	0.313 *** (0.01405)	0.3448 *** (0.0241

1-10 of 16 rows | 4-7 of 7 columns Previous **1** 2 Next

Table S1 shows the results of a logistic regression model of the cross-domain activity at an article level. This means researchers from different backgrounds collaborating (SA) by co-authoring published papers in differing categories (CIP). Data was filtered for cross-domain activity by the years (Yp) from 1970 to 2018, the number of co-authors of a paper (Kp) ≥ 2, and the number Medical Subject Heading (MeSH) words ≥ 2 to distinguish mono-domain versus cross-domain activity. The parameters contribute to the total information of predicting papers that have cross domain activity with numbers reported in odds ratios (what are the odds of cross-domain activity based on a high value of certain parameters). The columns are each separate models that show what parameters contribute to SA, CIP, or SACIP in the logistic regressions. Robust error estimations were also made which act as confidence intervals for the estimated contribution of the parameters. From looking at the values, the log number of MeSH words (w) seems to relate more to the SA and SACIP. The region the paper was published (NRegp) appears to contribute highly to Classification of Instructional Program (CIP) and SACIP. This suggests that researchers in certain regions of the world historically publish papers ranging from a wide variety of subject fields from their own. The pseudo R^2 value also showed that the interactions of years papers were published contributed very little to the logistic regression model, suggesting that these trends are prevalent from the 1970s to 2018 rather than just 2014.

	XSA <chr>	XCIP <chr>	X.XS <chr>

	XSA	XCIP	X.XSA
	<chr>	<chr>	<chr>
y	1.03 *** (0.0008)	1.0019 ** (0.00130)	1.0255
$\\bar{z_j}$	1.4876 *** (0.01745)	1.3439 *** (0.02630)	1.7655
ln k	0.5304 *** (0.0058)	1.755 *** (0.03110)	1.132
ln w	1.756 *** (0.02865)	0.8891 *** (0.02570)	1.8158
N_R	1.7625 *** (0.02645)	6.2972 *** (0.11395)	8.4242
N_{CIP}	1.4289 *** (0.01975)		
N_{SA}		1.2301 *** (0.01285)	
I_{2014+}			
$I_{R_{NA}}$			
$I_{R_{EU}}$			

1-10 of 16 rows | 1-4 of 7 columns Previous **1** 2 Next

Table S2 shows the “neighboring” or shorter-distance cross-domain combinations. When a cross-domain article is published in a relatively “close” category (for example a neuroscience researcher publishes a paper in Biology), it is considered a “neighbor”. Data was filtered for neighborSA, neighborCIP, and neighborSACIP. neighborSA activity was found by finding if any research was in SA category 1 to 4, neighborCIP was found by finding id any research was published in any CIP category of 1 to 7. neighborSACIP was found by checking if the article had both neighborCIP and neighborSA. Data was filtered for cross-domain activity by the years (Yp) from 1970 to 2018, the number of co-authors of a paper (Kp) ≥ 2, and the number Medical Subject Heading (MeSH) words ≥ 2 to distinguish mono-domain versus cross-domain activity. The parameters contribute to the total information of predicting papers that have cross domain activity with numbers reported in odds ratios (what are the odds of cross-domain activity based on a high value of certain parameters). The columns are each separate models that show what parameters contribute to SA, CIP, or SACIP in the logistic regressions. Robust error estimations were also made which act as confidence intervals for the estimated contribution of the parameters. In SA columns, the w and NRegp seemed to relate weakly to the logistic regression (with a very low pseudo R^2 of ~ 0.05) while in CIP and SACIP, the nRegp appears to be much more strongly related and contributed the largest amount to the regression information (with pseudo-R^2 without year interactions of about 0.17 and 0.19).

The pseudo R^2 value also showed that the interactions of years papers were published contributed very little to the logistic regression model (only adding about 0.01 to the value), suggesting that these trends are prevalent from the 1970s to 2018 rather than just 2014.

	XSA	XCIP	X.XSACIP
	<chr>	<chr>	<chr>
y	1.0317 *** (0.0007)	NA *** NA	NA *** (NA)
$\\bar{z_j}$	0.9975 (0.01185)	NA *** NA	NA *** (NA)
ln k	0.8848 *** (0.0099)	NA *** NA	NA *** (NA)

	XSA <chr>	XCIP <chr>	X.XSACIP <chr>
ln w	4.6549 *** (0.0788)	NA *** NA	NA *** (NA)
$N_R$	1.3243 *** (0.02405)	1 *** NA	1 *** ()
$N_{CIP}$	1.3073 *** (0.02305)		
$N_{SA}$		-2.9998 *** NA	
$I_{2014+}$			
$I_{R_{NA}}$			
$I_{R_{EU}}$			

1-10 of 16 rows | 1-5 of 7 columns Previous **1** 2 Next

Table S3 shows the “distant” cross-domain combinations. When a cross-domain article is published in a relatively “far” category (for example a neuroscience research co-author publishes a paper in Engineering), it is considered “distant”. Data was filtered for distantSA, distantCIP, and distantSACIP. distantSA activity was found by finding if a research paper was in SA category 1 to 4 AND in SA 5 to 6. distantCIP was found by finding if a research paper was in any CIP category 1,3, or 5 and also in CIP 4 or 8. distantSACIP was found by checking if the article had both distantCIP and distantSA. Data was filtered for cross-domain activity by the years (Yp) from 1970 to 2018, the number of co-authors of a paper (Kp) ≥ 2, and the number Medical Subject Heading (MeSH) words ≥ 2 to distinguish mono-domain versus cross-domain activity.

The parameters contribute to the total information of predicting papers that have cross domain activity with numbers reported in odds ratios (what are the odds of cross-domain activity based on a high value of certain parameters). The columns are each separate models that show what parameters contribute to distantSA, distantCIP, or distantSACIP in the logistic regressions. Robust error estimations were also made which act as confidence intervals for the estimated contribution of the parameters. In distantSA and distantSACIP columns, the log(w) seemed to relate more to the logistic regression information compared to the other parameters (with a very low pseudo R^2 of 0.0375 without year interactions and 0.496 with year interactions) while in CIP and SACIP, the nRegp appears to also be strongly related and contributed the largest amount to the regression information (pseudo-R^2 without year interactions of about 0.149 and 0.147).

The pseudo R^2 value also showed that the interactions of years papers were published contributed more to the distantCIP and distantSACIP logistic regression models (adding about 0.025 and 0.45 to the odds of predicting the paper being cross-domain), suggesting that there was slight increases in distant cross-domain brain-related research published from 2014 to 2018.

	model1_full <chr>	model2_full <chr>	▸
ln k	0.415*** (0.001986)	0.4157*** (0.001984)	
ln w	0.03184*** (0.003262)	0.03745*** (0.003252)	
t	-0.01982 (0.01488)	-0.01935 (0.01487)	
$I_{XSA}$	0.04803*** (0.00269)		

	model1_full <chr>	model2_full <chr>	▸
I_{XCIP}	0.0691*** (0.002928)		
$I_{X_{Neighboring,SA}}$		0.0878*** (0.003114)	
$I_{X_{Neighboring,CIP}}$		0.06751*** (0.003156)	
$I_{X_{Distant,SA}}$			
$I_{X_{Distant,CIP}}$			
$I_{X_{SA\\&CIP}}$			

1-10 of 21 rows | 1-3 of 7 columns Previous **1** 2 3 Next

Table S4 shows the career-level analysis with individual researcher fixed effects. Data was filtered by the years (Yp) from 1970 to 2018, the number of co-authors of a paper (Kp) ≥ 2, the number of MeSH words ≥ 2, and researchers with number of articles published (Na) ≥ 10. Robust standard errors are shown in parenthesis below each estimate and Y indicates additional fixed effects in the regression model. Each column is comparing normalized citation measures (Zp) with log(Kp), log(w), and the difference between the year the paper was published and the main authors first publication year (τ). There are additional parameters for each column comparing the “broad” SA and CIP, “neighbor” SA and CIP, “distant” SA and CIP, “broad” SACIP, “neighbor” SACIP, and “distant” SACIP.

The log number of coauthors (Kp) appears to be relatively the strongest positive correlation to predicting the number of citations a paper will receive. This is prevalent in all columns of data showing higher regression coefficients than any of the additional parameters as well. The pseudo R^2 values are relatively low, showing a range of adjusted R^2 values of 0.09 to 0.13. The highest adjusted R^2 value of 0.13 is comparing zp with “broad” SACIP with also the lowest number of articles (N=358237) of the 6 columns. All of the F-statistics were >> 1 ranging from 193 to 262 and every parameter robust standard error was statistically significant.

	model1_full <chr>	model2_ <chr>
ln k	0.4382*** (0.002953)	0.4247***
ln w	0.02491*** (0.004838)	0.05446*
t	-0.01838* (0.009258)	-0.01392
I_{2014+}	0.04437 (0.04239)	0.02929
$I_{X_{SA\\&CIP}}$	0.1554*** (0.005306)	
$I_{X_{SA\\&CIP}}\\times I_{2014+}$	-0.08808*** (0.007352)	
$I_{X_{Neighboring,SA\\&CIP}}$		0.1822***
$I_{X_{Neighboring,SA\\&CIP}}\\times I_{2014+}$		-0.16*** (
$I_{X_{Distant,SA\\&CIP}}$		
$I_{X_{Distant,SA\\&CIP}}\\times I_{2014+} $		

1-10 of 19 rows | 1-3 of 4 columns Previous **1** 2 Next

Table S5 shows the Flagship Project Effect using career-level analysis with researcher fixed effects. Data was filtered by the years (Yp) from 1970 to 2018, Kp ≥ 2, w ≥ 2, and researchers with Na ≥ 10. Robust standard errors are shown in parenthesis below each estimate and Y indicates additional fixed effects in the regression model. Each of the 3 columns are comparing Zp with log(Kp), log(w), and the whether the papers were published from years 2014 to 2018 (I_year). There are additional parameters for each column comparing the “broad” SACIP and “broad” SACIP cross interaction with I_year, “neighbor” SACIP and “neighbor” SACIP cross interaction with I_year, along with “distant” SACIP and “distant” SACIP cross-interaction with I_year.

The log number of coauthors (Kp) appears to be relatively the strongest positive correlation to predicting the number of citations a paper will receive. This is prevalent in all columns of data showing higher regression coefficients than any of the additional parameters as well. The adjusted R^2 values are relatively low, showing a range of adjusted R^2 values of 0.09 to 0.13. The highest adjusted R^2 value of 0.13 is comparing zp with “broad” SACIP and the cross-interaction of “broad” SACIP ith I_year. It is also the lowest number of articles (N=358237) of the 3 columns. All of the F-statistics were >> 1 ranging from 191 to 229 and every parameter robust standard error except “distant” SACIP parameter in column 3 was statistically significant.

```
## Warning: package 'car' was built under R version 4.0.5
```

```
## Loading required package: carData
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

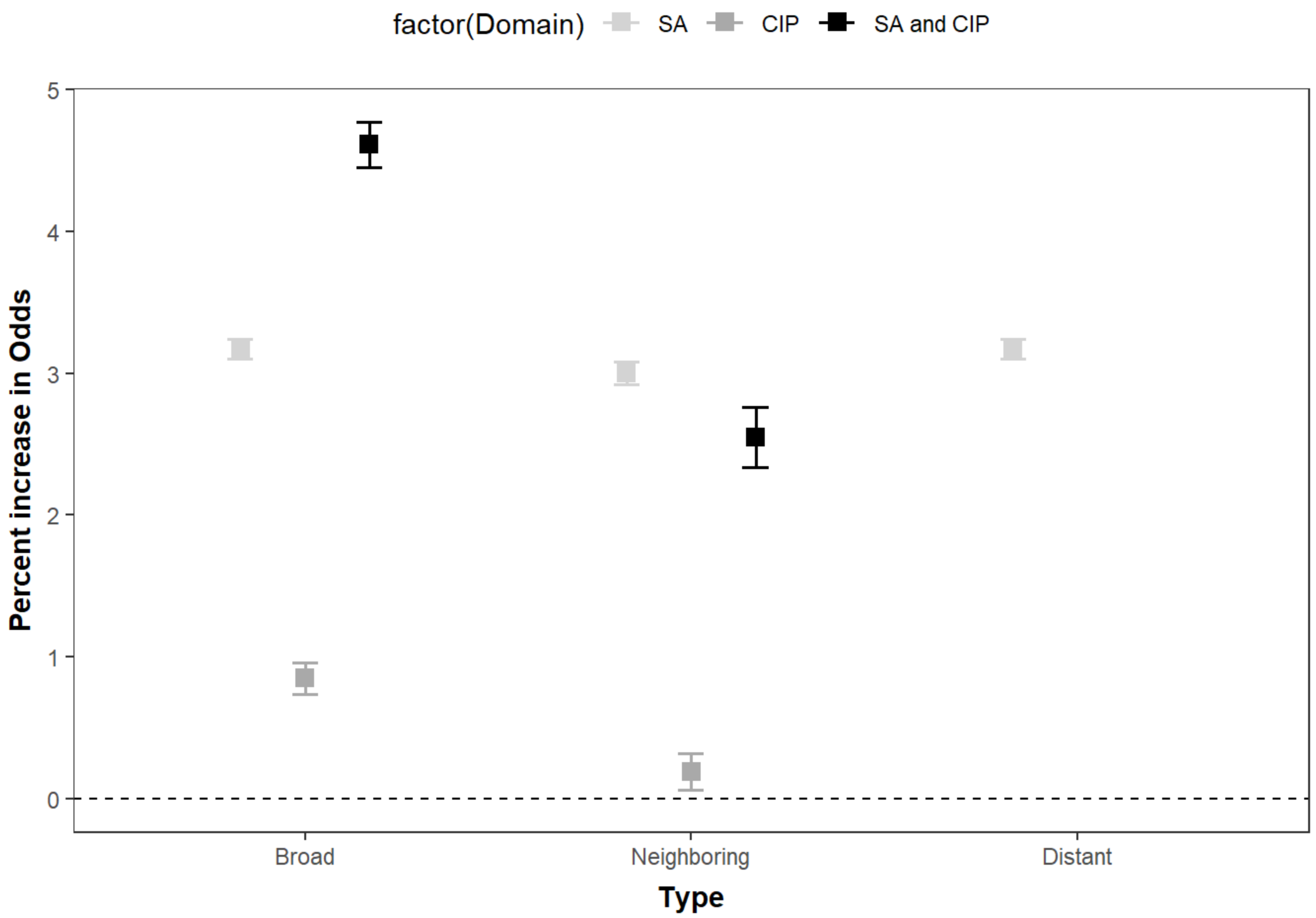
```
## v tibble  3.1.0     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v purrr   0.3.4     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter()     masks stats::filter()
## x dplyr::group_rows() masks kableExtra::group_rows()
## x dplyr::lag()        masks stats::lag()
## x dplyr::recode()     masks car::recode()
## x purrr::some()       masks car::some()
```

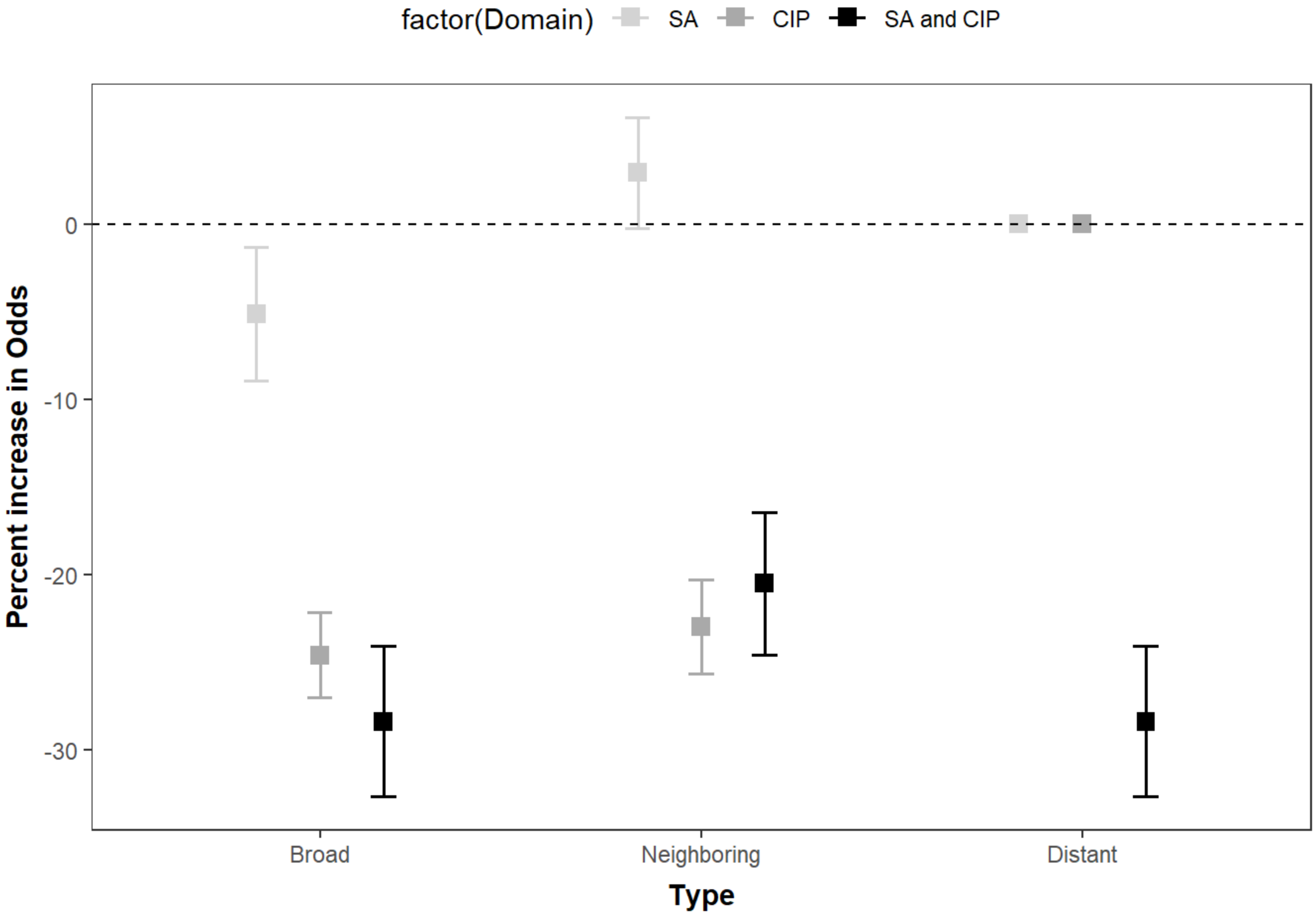
```
## Warning: NAs introduced by coercion
```

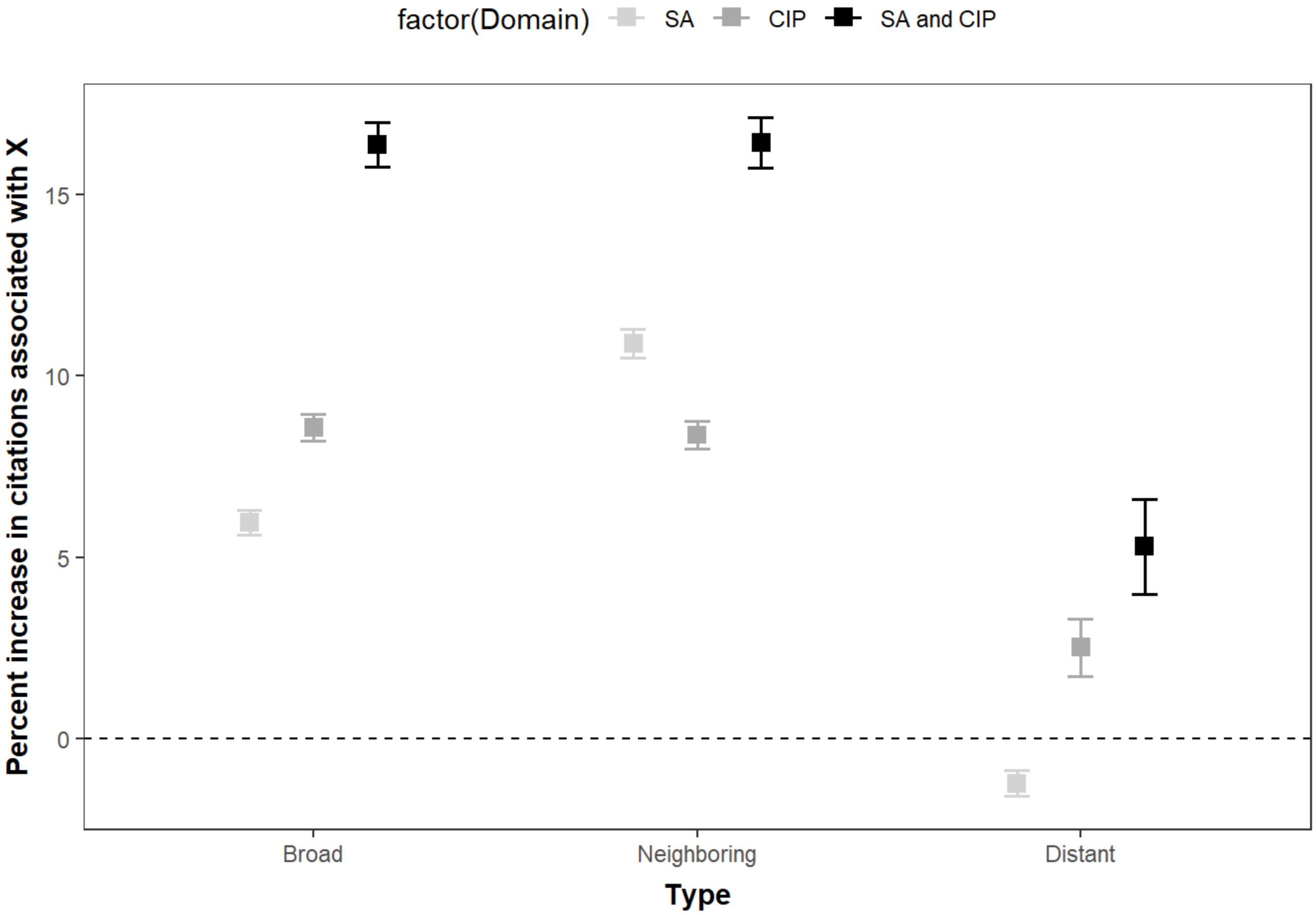
```
## Warning: 1 parsing failure.
## row col expected actual
##   9  -- a number   (NA)
```

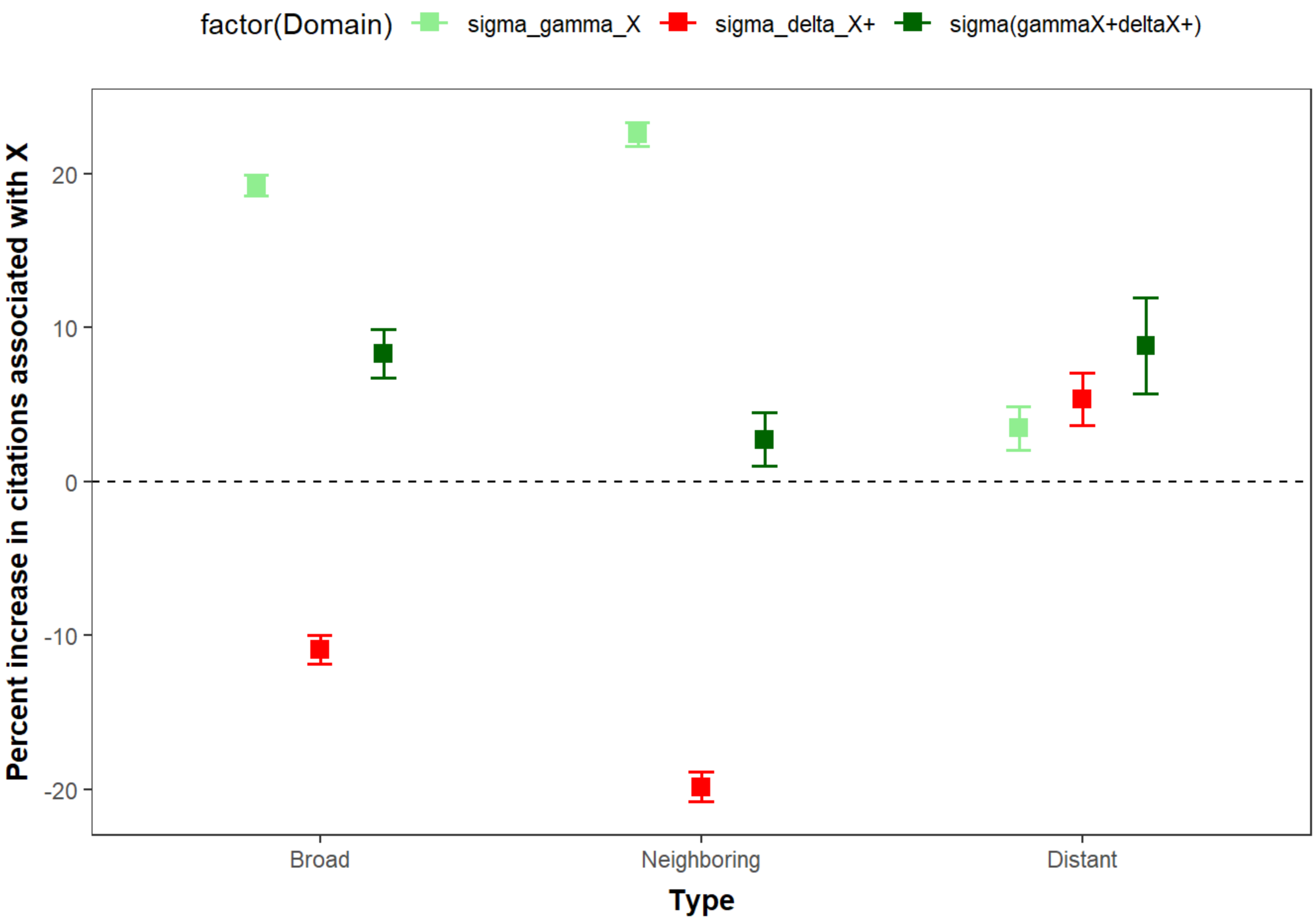
```
## Warning: Removed 2 rows containing missing values (geom_point).
```



```
## Warning: 1 parsing failure.
## row col expected actual
##   8  -- a number   (NA)
```


factor(Domain)
SA
CIP
SA and CIP
Percent increase in Odds
0
-10
-20
-30
Broad
Neighboring
Distant
Type


factor(Domain)
SA
CIP
SA and CIP
Percent increase in citations associated with X
15
10
5
0
Broad
Neighboring
Distant
Type


factor(Domain)
sigma_gamma_X
sigma_delta_X+
sigma(gammaX+deltaX+)
Percent increase in citations associated with X
20
10
0
-10
-20
Broad
Neighboring
Distant
Type