

# Compare Linking across linking runs

Christoph, Flavio, Mona

2023-07-23

## Linking Advisors and Graduates from ProQuest to MAG AuthorIds

To check whether the linking makes sense we compare the links across several independent runs of the labelling.

### Comparison for Graduates

```
res_combined <- reduce(res, inner_join, by = "variable")

res_combined %>%
  mutate(across(where(is.numeric), round, digits = 2)) %>% # this is an updated form of mutate_if()
  kable(format = "latex", digits = 2, booktabs = TRUE) %>%
  kable_styling(latex_options = c("striped", "scale_down"))

## Warning: There was 1 warning in `mutate()`.
## i In argument: `across(where(is.numeric), round, digits = 2)`.
## Caused by warning:
## ! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.
## Supply arguments directly to `.fns` through an anonymous function instead.
##
## # Previously
## across(a:b, mean, na.rm = TRUE)
##
## # Now
## across(a:b, \(x) mean(x, na.rm = TRUE))
```

### Comparison for Advisors

```
res_combined <- reduce(res, inner_join, by = "variable")

res_combined %>%
  mutate(across(where(is.numeric), round, digits = 2)) %>% # this is an updated form of mutate_if()
  kable(format = "latex", digits = 2, booktabs = TRUE) %>%
  kable_styling(latex_options = c("striped", "scale_down"))
```

variable	biology	business	computer science	economics	engineering	environmental science	geology	history	materials science	philosophy	physics	political science	psychology
same	0.45	0.24	0.52	0.20	0.68	0.43	0.29	0.42	0.38	0.11	0.06	0.20	0.24
only1	0.51	0.56	0.40	0.78	0.07	0.55	0.69	0.48	0.54	0.83	0.18	0.75	0.74
only2	0.03	0.19	0.08	0.01	0.25	0.01	0.00	0.09	0.07	0.06	0.39	0.04	0.01
diff	0.01	0.01	0.00	0.00	0.00	0.01	0.02	0.01	0.01	0.00	0.36	0.01	0.01
diff_rell	0.01	0.01	0.01	0.00	0.00	0.01	0.02	0.01	0.01	0.00	0.60	0.01	0.01
nlink1	79517.00	10683.00	18599.00	9367.00	19220.00	10843.00	7666.00	8277.00	16115.00	3301.00	20980.00	13506.00	61227.00
nlink2	40569.00	5872.00	12169.00	2048.00	23776.00	4944.00	2401.00	4744.00	7975.00	614.00	28191.00	3566.00	15897.00

variable	biology	business	computer science	economics	engineering	environmental science	geology	history	materials science	philosophy	physics	political science	psychology
same	0	0.69	0.64	0.65	0.15	0.66	0.28	0.67	0.12	0.92	0.49	0.94	0.89
only1	1	0.31	0.04	0.01	0.19	0.01	0.00	0.30	0.00	0.07	0.01	0.00	0.08
only2	0	0.00	0.20	0.30	0.11	0.32	0.35	0.02	0.14	0.00	0.50	0.04	0.02
diff	0	0.00	0.13	0.03	0.56	0.00	0.37	0.02	0.73	0.01	0.00	0.02	0.01
diff_rell	0	0.00	0.16	0.05	0.63	0.00	0.57	0.02	0.86	0.01	0.00	0.02	0.01
nlink1	29998	12213.00	17488.00	6024.00	21225.00	7205.00	3665.00	10681.00	12510.00	4830.00	6821.00	13160.00	32473.00
nlink2	1	8336.00	18989.00	8803.00	22962.00	10343.00	6236.00	7704.00	11244.00	4424.00	12611.00	13594.00	29902.00