

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.

Which fields and which linking runs to compare?

Comparison for Graduates

```
fields_to_process <-c( #"art", <- dropped
  "biology",
  "business",
  # "chemistry",
  "computer science" ,
  "economics",
  "engineering",
  "environmental science",
  #"geography",
  "geology" ,
  "history",
  "materials science",
  "mathematics",
  #"medicine", <- dropped
  "philosophy",
  "physics",
  "political science",
  "psychology" #,
  # "sociology"
)
```

```
linker1 = "mona_degree0"
linker2 = "christoph_degree0"
```

```
res_combined_graduates <- reduce(res, rbind)
```

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

```
#res_combined_graduates %>% mutate(across(where(is.numeric), \(x) round(x, digits = 2))) %>% View()
```

Comparison for Advisors

field	same	only1	only2	diff	diff_rell	nlink1	nlink2	namedist_pq_1	namedist_pq_2	namedist_diff	namedist_diff_pq_1	namedist_diff_pq_2	namedist_pq_only1	namedist_pq_only2
biology	0.45	0.03	0.51	0.01	0.02	40569	79517	0.07	0.08	0.05	0.07	0.08	0.07	0.01
business	0.24	0.19	0.56	0.01	0.01	5872	10683	0.08	0.08	0.27	0.08	0.08	0.08	0.03
computer science	0.52	0.08	0.40	0.00	0.01	12169	18599	0.07	0.05	0.17	0.07	0.05	0.07	0.01
economics	0.20	0.01	0.78	0.00	0.02	2048	9367	0.10	0.08	0.25	0.10	0.08	0.10	0.00
engineering	0.68	0.25	0.07	0.00	0.00	23776	19220	0.06	0.06	0.07	0.06	0.06	0.06	0.23
environmental science	0.43	0.01	0.55	0.01	0.01	4944	10843	0.09	0.09	0.06	0.09	0.09	0.09	0.00
geology	0.29	0.00	0.69	0.02	0.07	2401	7666	0.09	0.12	0.16	0.09	0.12	0.09	0.00
history	0.42	0.09	0.48	0.01	0.02	4744	8277	0.10	0.10	0.11	0.10	0.10	0.10	0.02
materials science	0.38	0.07	0.54	0.01	0.02	7975	16115	0.08	0.08	0.07	0.08	0.08	0.08	0.01
mathematics	0.20	0.00	0.79	0.00	0.00	3829	18520	0.06	0.06	0.17	0.06	0.06	0.06	0.00
philosophy	0.11	0.06	0.83	0.00	0.00	614	3301	0.08	0.08	NaN	NaN	NaN	0.08	0.01
physics	0.06	0.39	0.18	0.36	0.44	28191	20980	0.32	0.08	0.32	0.32	0.08	0.32	0.16
political science	0.20	0.04	0.75	0.01	0.06	3566	13506	0.12	0.08	0.25	0.12	0.08	0.12	0.00
psychology	0.24	0.01	0.74	0.01	0.05	15897	61227	0.08	0.08	0.10	0.08	0.08	0.08	0.00

field	same	only1	only2	diff	diff_rell	nlink1	nlink2	namedist_pq_1	namedist_pq_2	namedist_diff	namedist_diff_pq_1	namedist_diff_pq_2	namedist_pq_only1	namedist_pq_only2
biology	0.78	0.00	0.22	0.00	0.00	27516	35175	0.02	0.02	0.16	0.02	0.02	0.02	0.00
business	0.69	0.01	0.30	0.00	0.00	8060	11629	0.02	0.02	0.05	0.02	0.02	0.02	0.00
chemistry	0.93	0.00	0.02	0.05	0.05	11713	12138	0.02	0.02	0.02	0.02	0.02	0.02	0.00
computer science	0.89	0.08	0.02	0.01	0.01	17469	16498	0.03	0.03	0.15	0.03	0.03	0.03	0.11
economics	0.96	0.00	0.02	0.02	0.02	8446	8655	0.02	0.02	0.05	0.02	0.02	0.02	0.00
engineering	0.81	0.00	0.19	0.00	0.00	21162	26140	0.04	0.05	0.20	0.04	0.05	0.04	0.00
environmental science	0.73	0.23	0.03	0.01	0.01	9066	7312	0.03	0.03	0.07	0.03	0.03	0.03	0.22
geography	0.62	0.32	0.02	0.03	0.04	7107	5114	0.04	0.03	0.12	0.04	0.03	0.04	0.43
geology	0.84	0.00	0.14	0.01	0.01	4584	5436	0.03	0.03	0.01	0.03	0.03	0.03	0.00
history	0.85	0.01	0.12	0.02	0.03	10129	11817	0.02	0.03	0.12	0.02	0.03	0.02	0.00
materials science	0.74	0.01	0.13	0.12	0.14	14246	16079	0.07	0.05	0.13	0.07	0.05	0.07	0.00
mathematics	0.78	0.15	0.01	0.06	0.06	12759	10924	0.04	0.03	0.11	0.04	0.03	0.04	0.57
philosophy	0.93	0.00	0.06	0.00	0.00	4016	4381	0.02	0.02	0.03	0.02	0.02	0.02	0.00
physics	0.51	0.15	0.03	0.31	0.32	13955	12235	0.11	0.17	0.28	0.11	0.17	0.11	0.85
political science	0.87	0.00	0.08	0.04	0.05	13040	14002	0.02	0.05	0.27	0.02	0.05	0.02	0.00
psychology	0.91	0.00	0.08	0.00	0.00	28878	32740	0.02	0.03	0.09	0.02	0.03	0.02	0.00
sociology	0.46	0.01	0.51	0.02	0.03	8270	16343	0.03	0.02	0.21	0.03	0.02	0.03	0.00

```
fields_to_process <-c( #"art", <- dropped
  "biology",
  "business",
  "chemistry",
  "computer science" ,
  "economics",
  "engineering",
  "environmental science",
  "geography",
  "geology" ,
  "history",
  "materials science",
  "mathematics",
  #"medicine", <- dropped
  "philosophy",
  "physics",
  "political science",
  "psychology" ,
  "sociology")

linker1 = "flavio_with_protocol_cleaninst" # flavio_with_protocol_cleaninst
linker2 = "christoph_with_protocol_cleaninst" #christoph_degree0_with_protocol_updated
```

```
res_combined_advisors <- reduce(res, rbind)

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

```
# res_combined_advisors %>%  
#   mutate(across(where(is.numeric), \(x) round(x, digits = 2))) %>% View()
```

```
#  
# field="chemistry"  
# linktype="advisors"  
# years="19902015"  
#  
# links <- compare(field, linktype, linker1, linker2, years, inspect=TRUE)
```

Christoph and Flavio with Protocol and cleaned institutions