das-versprechen-der-vernetzung-datenpublikation

January 31, 2021

1 Prolegomenon

This is the publication of all data used analysing the *Nationale Forschungsdateninfrastruktur* (NFDI) with its different consortia. The source of the data are the binding Letters of Intent (LoI) of the consortia in which they name their collaboration partners.[1] The following analysis only includes documents that were submitted to the DFG as binding pre-applications for 2019 and 2020 (binding Letters of Intent). Consortia that did not submit a binding Letter of Intent in 2019 or 2020 were not included. We gathered all the data at the point when the LoI have been turned in. Consortia which have not turned in a binding Letter of Intent are not considered at all.

```
setwd(getwd())
library('igraph')
library('dplyr')
```

1.1 Presettings

Making sure that we have a common seed.

```
nfdi_seed <- function() {
  set.seed(1234)
}
```

We define colors we will use for grouping consortia.

```
nfdi_conference_colors <- function() {

  nfdi_color_code <<- c("#f5ac9f", # Medizin
                        "#e43516", # Lebenswissenschaften
                        "#f9b900", # Geisteswissenschaften
                        "#007aaf", # Ingenieurwissenschaften
                        "#6ca11d"  # Chemie und Physik
                      )
  nfdi_color_groups <<- nfdi_color_code[as.numeric(as.
 ↪factor(V(nfdi_network_year)$group))]
}
```

Some presettings for plotting the networks.

[1] Have a look at the GitHub repository of Dorothea Strecker (https://github.com/dorothearrr/NFDI_Netzwerk), where the data has been destilled from the various LoI.

```r
nfdi_plot_settings <- function(){

    nfdi_conference_colors()

    graph_attr(nfdi_network_year, "layout") <<- norm_coords(layout.
 ↪graphopt(nfdi_network_year),
                                                           ymin = -1, ymax =␣
 ↪1, xmin = -1, xmax = 1)*2

    vertex_attr(nfdi_network_year, "label.cex") <<- .7
    vertex_attr(nfdi_network_year, "label.color") <<- "black"
    vertex_attr(nfdi_network_year, "label.font") <<- 1
    vertex_attr(nfdi_network_year, "label.family") <<- "Helvetica"
    vertex_attr(nfdi_network_year, "color") <<- nfdi_color_groups
    vertex_attr(nfdi_network_year, "frame.color") <<- nfdi_color_groups
    vertex_attr(nfdi_network_year, "size") <<-␣
 ↪rescale_vertices(degree(nfdi_network_year,
                                  mode="total"))*20

    edge_attr(nfdi_network_year, "color") <- "#808080"
    edge_attr(nfdi_network_year, "curved") <<- 0.1
    edge_attr(nfdi_network_year, "arrow.size") <<- .5
    edge_attr(nfdi_network_year, "arrow.width") <<- .5

    rescale = F

}
```

We can define a function for rescaling the vertex size. This makes plots of the 2019 and 2020 networks visually comparable.

```r
rescale_vertices <- function(x){(x - 3)/(20 - 3)} # min-max normalization; min␣
 ↪= (min - 1) to omit zeroes
```

Next we are setting up a function to get the same presettings for all different data frames.

```r
nfdi_presettings <- function(nfdi_edges_year,nfdi_directed_graph){
  nfdi_seed()
  nfdi_edges <<- get(paste0("nfdi_edges_",nfdi_edges_year))
  nfdi_nodes <<- get(paste0("nfdi_nodes_",nfdi_edges_year))
  nfdi_network_year <<- graph_from_data_frame(
      d = nfdi_edges,
      vertices = nfdi_nodes,
      directed = nfdi_directed_graph)
  nfdi_plot_settings()
}
```

1.2 Data sets

The core of the this publication are the sets of edges i.e. the connections for collaborations between the consortia. So far there are only two data sets available, for 2019 and for 2020.

1.2.1 2019

Now follows the information regarding the consortia. Precisely the allocated group according to the NFDI-conference system (`group`), and since at the time of turning in the LoI none of the consortia had been funded the column (`funded`) has `0` as value for all the consortia.

```
nfdi_edges_2019 <- read.table(
  header=TRUE,
  sep=",",
  text="
from,to
Astro-NFDI,PAHN-PaN
Astro-NFDI,DAPHNE
Astro-NFDI,NFDI4Earth
BERD\\@NFDI,KonsortSWD
BERD\\@NFDI,ForumX
BERD\\@NFDI,Text+
DAPHNE,FAIRmat
DAPHNE,NFDI4Chem
DAPHNE,NFDI4Ing
DAPHNE,NFDI4MSE
DAPHNE,NFDI4Cat
DAPHNE,PAHN-PaN
DAPHNE,Astro-NFDI
DataPLANT,NFDI4BioDiversity
DataPLANT,NFDI4Agri
DataPLANT,NFDI4Chem
FAIRmat,DAPHNE
FAIRmat,MaRDI
FAIRmat,NFDI4Chem
FAIRmat,NFDI4Cat
FAIRmat,NFDI4Ing
FAIRmat,NFDI4MSE
ForumX,NFDI4Medicine
ForumX,KonsortSWD
ForumX,BERD\\@NFDI
ForumX,NFDI4Culture
ForumX,Text+
GHGA,NFDI4Medicine
GHGA,NFDI4Health
KonsortSWD,BERD\\@NFDI
KonsortSWD,NFDI4BioDiversity
KonsortSWD,NFDI4Earth
```

```
KonsortSWD,NFDI4Health
KonsortSWD,Text+
MaRDI,NFDI4MSE
MaRDI,PAHN-PaN
MaRDI,FAIRmat
MaRDI,NFDI4Ing
MaRDI,NFDI4Chem
MaRDI,NFDI4Culture
NFDI4Agri,NFDI4Health
NFDI4Agri,DataPLANT
NFDI4Agri,NFDI4BioDiversity
NFDI4Agri,KonsortSWD
NFDI4Agri,NFDI4Earth
NFDI4BioDiversity,NFDI4Earth
NFDI4BioDiversity,NFDI4Agri
NFDI4BioDiversity,NFDI4Chem
NFDI4BioDiversity,NFDI4Health
NFDI4BioDiversity,KonsortSWD
NFDI4BioDiversity,NFDI4Crime
NFDI4BioDiversity,DataPLANT
NFDI4BioDiversity,NFDI4Medicine
NFDI4Cat,FAIRmat
NFDI4Cat,NFDI4Chem
NFDI4Cat,NFDI4Ing
NFDI4Cat,DAPHNE
NFDI4Chem,FAIRmat
NFDI4Chem,NFDI4Ing
NFDI4Chem,NFDI4Cat
NFDI4Chem,DAPHNE
NFDI4Chem,PAHN-PaN
NFDI4Chem,NFDI4Health
NFDI4Chem,NFDI4BioDiversity
NFDI4Crime,NFDI4BioDiversity
NFDI4Crime,NFDI4Medicine
NFDI4Crime,Text+
NFDI4Culture,Text+
NFDI4Culture,MaRDI
NFDI4Culture,NFDI4Ing
NFDI4Earth,Astro-NFDI
NFDI4Earth,KonsortSWD
NFDI4Earth,NFDI4Agri
NFDI4Earth,NFDI4BioDiversity
NFDI4Earth,NFDI4Ing
NFDI4Health,NFDI4Medicine
NFDI4Health,GHGA
NFDI4Health,KonsortSWD
NFDI4Health,NFDI4Chem
```

```
NFDI4Health,NFDI4Agri
NFDI4Health,NFDI4Earth
NFDI4Health,NFDI4BioDiversity
NFDI4Ing,NFDI4MSE
NFDI4Ing,FAIRmat
NFDI4Ing,NFDI4MobilTech
NFDI4Ing,NFDI4Chem
NFDI4Ing,NFDI4Earth
NFDI4Ing,MaRDI
NFDI4Ing,NFDI4Medicine
NFDI4Ing,Text+
NFDI4Ing,NFDI4Culture
NFDI4Medicine,GHGA
NFDI4Medicine,NFDI4Health
NFDI4Medicine,NFDI4Ing
NFDI4Medicine,NFDI4Crime
NFDI4Medicine,ForumX
NFDI4Medicine,KonsortSWD
NFDI4Medicine,NFDI4Agri
NFDI4MobilTech,NFDI4Ing
NFDI4MobilTech,ForumX
NFDI4MobilTech,NFDI4Earth
NFDI4MSE,FAIRmat
NFDI4MSE,NFDI4Ing
NFDI4MSE,MaRDI
NFDI4MSE,NFDI4Chem
NFDI4MSE,DAPHNE
Text+,NFDI4Culture
Text+,KonsortSWD
Text+,NFDI4Ing
Text+,NFDI4Earth
Text+,NFDI4BioDiversity
")
```

```
nfdi_nodes_2019 <- read.table(
  header=TRUE,
  sep=",",
  text="
name,group,funded
Astro-NFDI,5,0
BERD\\@NFDI,3,0
DAPHNE,5,0
DataPLANT,2,0
FAIRmat,5,0
ForumX,3,0
GHGA,1,0
KonsortSWD,3,0
```

```
MaRDI,4,0
NFDI4Agri,2,0
NFDI4BioDiversity,2,0
NFDI4Cat,5,0
NFDI4Chem,5,0
NFDI4Crime,2,0
NFDI4Culture,3,0
NFDI4Earth,2,0
NFDI4Health,1,0
NFDI4Ing,4,0
NFDI4MSE,5,0
NFDI4Medicine,1,0
NFDI4MobilTech,4,0
PAHN-PaN,5,0
Text+,3,0
")
```

1.2.2 2020

What follows is the data for the year 2020 and the collaboration intentions between new and already funded consortia.

```
nfdi_edges_2020 <- read.table(
  header=TRUE,
  sep=",",
  text="
from,to
BERD\\@NFDI,KonsortSWD
BERD\\@NFDI,MaRDI
BERD\\@NFDI,NFDI4Memory
BERD\\@NFDI,Text+
DAPHNE4NFDI,FAIRmat
DAPHNE4NFDI,NFDI-MatWerk
DAPHNE4NFDI,NFDI4Cat
DAPHNE4NFDI,NFDI4Chem
DAPHNE4NFDI,NFDI4Health
DAPHNE4NFDI,NFDI4Ing
DAPHNE4NFDI,NFDI4Objects
DAPHNE4NFDI,PUNCH4NFDI
FAIRmat,DAPHNE4NFDI
FAIRmat,DataPLANT
FAIRmat,MaRDI
FAIRmat,NFDI-MatWerk
FAIRmat,NFDI4Cat
FAIRmat,NFDI4Chem
FAIRmat,NFDI4DataScience
FAIRmat,NFDI4Ing
```

```
FAIRmat,NFDIxCS
FAIRmat,PUNCH4NFDI
MaRDI,BERD\\@NFDI
MaRDI,FAIRmat
MaRDI,NFDI-MatWerk
MaRDI,NFDI-Neuro
MaRDI,NFDI4Cat
MaRDI,NFDI4Chem
MaRDI,NFDI4Ing
MaRDI,PUNCH4NFDI
NFDI-MatWerk,DAPHNE4NFDI
NFDI-MatWerk,DataPLANT
NFDI-MatWerk,FAIRmat
NFDI-MatWerk,MaRDI
NFDI-MatWerk,NFDI4Chem
NFDI-MatWerk,NFDI4DataScience
NFDI-MatWerk,NFDI4Ing
NFDI-MatWerk,NFDIxCS
NFDI-Neuro,DataPLANT
NFDI-Neuro,GHGA
NFDI-Neuro,NFDI4BioDiversity
NFDI-Neuro,NFDI4Culture
NFDI-Neuro,NFDI4Earth
NFDI-Neuro,NFDI4Health
NFDI-Neuro,NFDI4Ing
NFDI-Neuro,NFDI4Microbiota
NFDI4Agri,DataPLANT
NFDI4Agri,KonsortSWD
NFDI4Agri,NFDI4BioDiversity
NFDI4Agri,NFDI4Earth
NFDI4Agri,NFDI4Health
NFDI4Agri,NFDI4Immuno
NFDI4Agri,NFDI4Microbiota
NFDI4DataScience,KonsortSWD
NFDI4DataScience,MaRDI
NFDI4DataScience,NFDI-MatWerk
NFDI4DataScience,NFDI4BioDiversity
NFDI4DataScience,NFDI4Cat
NFDI4DataScience,NFDI4Chem
NFDI4DataScience,NFDI4Culture
NFDI4DataScience,NFDI4Health
NFDI4DataScience,NFDI4Ing
NFDI4DataScience,NFDI4Microbiota
NFDI4DataScience,NFDIxCS
NFDI4Earth,DataPLANT
NFDI4Earth,GHGA
NFDI4Earth,KonsortSWD
```

```
NFDI4Earth,NFDI4Agri
NFDI4Earth,NFDI4BioDiversity
NFDI4Earth,NFDI4Cat
NFDI4Earth,NFDI4Chem
NFDI4Earth,NFDI4Culture
NFDI4Earth,NFDI4Health
NFDI4Earth,NFDI4Ing
NFDI4Earth,NFDI4Objects
NFDI4Immuno,GHGA
NFDI4Immuno,NFDI4Agri
NFDI4Immuno,NFDI4Health
NFDI4Immuno,NFDI4Microbiota
NFDI4Memory,BERD\\@NFDI
NFDI4Memory,KonsortSWD
NFDI4Memory,MaRDI
NFDI4Memory,NFDI4Culture
NFDI4Memory,NFDI4Objects
NFDI4Memory,Text+
NFDI4Microbiota,DataPLANT
NFDI4Microbiota,GHGA
NFDI4Microbiota,NFDI4Agri
NFDI4Microbiota,NFDI4BioDiversity
NFDI4Microbiota,NFDI4Chem
NFDI4Microbiota,NFDI4DataScience
NFDI4Microbiota,NFDI4Health
NFDI4Microbiota,NFDI4Immuno
NFDI4Microbiota,NFDI4Ing
NFDI4Objects,KonsortSWD
NFDI4Objects,NFDI4Agri
NFDI4Objects,NFDI4BioDiversity
NFDI4Objects,NFDI4Culture
NFDI4Objects,NFDI4Earth
NFDI4Objects,NFDI4Memory
NFDI4Objects,Text+
NFDI4SD,NFDI4Culture
NFDI4SD,NFDI4DataScience
NFDI4SD,NFDI4Memory
NFDI4SD,NFDI4Objects
NFDIxCS,FAIRmat
NFDIxCS,MaRDI
NFDIxCS,NFDI4Chem
NFDIxCS,NFDI4DataScience
NFDIxCS,NFDI4Earth
NFDIxCS,NFDI4Ing
PUNCH4NFDI,DAPHNE4NFDI
PUNCH4NFDI,FAIRmat
PUNCH4NFDI,GHGA
```

```
PUNCH4NFDI,MaRDI
PUNCH4NFDI,NFDI4Earth
PUNCH4NFDI,NFDI4Ing
PUNCH4NFDI,NFDIxCS
Text+,KonsortSWD
Text+,NFDI4BioDiversity
Text+,NFDI4Culture
Text+,NFDI4Earth
Text+,NFDI4Ing
Text+,NFDI4Memory
Text+,NFDI4Objects
")
```

```
nfdi_nodes_2020 <- read.table(
  header=TRUE,
  sep=",",
  text="
name,group,funded
BERD\\@NFDI,3,0
DAPHNE4NFDI,5,0
DataPLANT,2,1
FAIRmat,5,0
GHGA,1,1
KonsortSWD,3,1
MaRDI,4,0
NFDI-MatWerk,4,0
NFDI-Neuro,1,0
NFDI4Agri,2,0
NFDI4BioDiversity,2,1
NFDI4Cat,5,1
NFDI4Chem,5,1
NFDI4Culture,3,1
NFDI4DataScience,4,0
NFDI4Earth,2,0
NFDI4Health,1,1
NFDI4Immuno,1,0
NFDI4Ing,4,1
NFDI4Memory,3,0
NFDI4Microbiota,2,0
NFDI4Objects,3,0
NFDI4SD,3,0
NFDIxCS,4,0
PUNCH4NFDI,5,0
Text+,3,0
")
```

1.3 Edges

Let us now display the edges for the data sets we have so far. We do this using a particular function.

```
nfdi_edges <- function(nfdi_edges_year) {
    df <- get(paste0("nfdi_edges_",nfdi_edges_year))

   # df
    print(tbl_df(df), n=Inf)
}
# nfdi_edges(2019)
# nfdi_edges(2020)
```

1.3.1 2019

```
nfdi_edges(2019)
```

```
# A tibble: 111 × 2
                 from               to
               <fctr>           <fctr>
1          Astro-NFDI         PAHN-PaN
2          Astro-NFDI           DAPHNE
3          Astro-NFDI       NFDI4Earth
4         BERD\\@NFDI       KonsortSWD
5         BERD\\@NFDI           ForumX
6         BERD\\@NFDI            Text+
7              DAPHNE          FAIRmat
8              DAPHNE        NFDI4Chem
9              DAPHNE         NFDI4Ing
10             DAPHNE         NFDI4MSE
11             DAPHNE         NFDI4Cat
12             DAPHNE         PAHN-PaN
13             DAPHNE       Astro-NFDI
14          DataPLANT NFDI4BioDiversity
15          DataPLANT        NFDI4Agri
16          DataPLANT        NFDI4Chem
17            FAIRmat           DAPHNE
18            FAIRmat            MaRDI
19            FAIRmat        NFDI4Chem
20            FAIRmat         NFDI4Cat
21            FAIRmat         NFDI4Ing
22            FAIRmat         NFDI4MSE
23             ForumX     NFDI4Medicine
24             ForumX       KonsortSWD
25             ForumX      BERD\\@NFDI
26             ForumX     NFDI4Culture
27             ForumX            Text+
28               GHGA     NFDI4Medicine
```

29	GHGA	NFDI4Health
30	KonsortSWD	BERD\\@NFDI
31	KonsortSWD	NFDI4BioDiversity
32	KonsortSWD	NFDI4Earth
33	KonsortSWD	NFDI4Health
34	KonsortSWD	Text+
35	MaRDI	NFDI4MSE
36	MaRDI	PAHN-PaN
37	MaRDI	FAIRmat
38	MaRDI	NFDI4Ing
39	MaRDI	NFDI4Chem
40	MaRDI	NFDI4Culture
41	NFDI4Agri	NFDI4Health
42	NFDI4Agri	DataPLANT
43	NFDI4Agri	NFDI4BioDiversity
44	NFDI4Agri	KonsortSWD
45	NFDI4Agri	NFDI4Earth
46	NFDI4BioDiversity	NFDI4Earth
47	NFDI4BioDiversity	NFDI4Agri
48	NFDI4BioDiversity	NFDI4Chem
49	NFDI4BioDiversity	NFDI4Health
50	NFDI4BioDiversity	KonsortSWD
51	NFDI4BioDiversity	NFDI4Crime
52	NFDI4BioDiversity	DataPLANT
53	NFDI4BioDiversity	NFDI4Medicine
54	NFDI4Cat	FAIRmat
55	NFDI4Cat	NFDI4Chem
56	NFDI4Cat	NFDI4Ing
57	NFDI4Cat	DAPHNE
58	NFDI4Chem	FAIRmat
59	NFDI4Chem	NFDI4Ing
60	NFDI4Chem	NFDI4Cat
61	NFDI4Chem	DAPHNE
62	NFDI4Chem	PAHN-PaN
63	NFDI4Chem	NFDI4Health
64	NFDI4Chem	NFDI4BioDiversity
65	NFDI4Crime	NFDI4BioDiversity
66	NFDI4Crime	NFDI4Medicine
67	NFDI4Crime	Text+
68	NFDI4Culture	Text+
69	NFDI4Culture	MaRDI
70	NFDI4Culture	NFDI4Ing
71	NFDI4Earth	Astro-NFDI
72	NFDI4Earth	KonsortSWD
73	NFDI4Earth	NFDI4Agri
74	NFDI4Earth	NFDI4BioDiversity
75	NFDI4Earth	NFDI4Ing
76	NFDI4Health	NFDI4Medicine

```
77         NFDI4Health             GHGA
78         NFDI4Health       KonsortSWD
79         NFDI4Health        NFDI4Chem
80         NFDI4Health        NFDI4Agri
81         NFDI4Health       NFDI4Earth
82         NFDI4Health NFDI4BioDiversity
83            NFDI4Ing         NFDI4MSE
84            NFDI4Ing          FAIRmat
85            NFDI4Ing    NFDI4MobilTech
86            NFDI4Ing        NFDI4Chem
87            NFDI4Ing       NFDI4Earth
88            NFDI4Ing            MaRDI
89            NFDI4Ing     NFDI4Medicine
90            NFDI4Ing            Text+
91            NFDI4Ing      NFDI4Culture
92       NFDI4Medicine             GHGA
93       NFDI4Medicine      NFDI4Health
94       NFDI4Medicine         NFDI4Ing
95       NFDI4Medicine       NFDI4Crime
96       NFDI4Medicine           ForumX
97       NFDI4Medicine       KonsortSWD
98       NFDI4Medicine        NFDI4Agri
99      NFDI4MobilTech         NFDI4Ing
100     NFDI4MobilTech           ForumX
101     NFDI4MobilTech       NFDI4Earth
102           NFDI4MSE          FAIRmat
103           NFDI4MSE         NFDI4Ing
104           NFDI4MSE            MaRDI
105           NFDI4MSE        NFDI4Chem
106           NFDI4MSE           DAPHNE
107              Text+      NFDI4Culture
108              Text+       KonsortSWD
109              Text+         NFDI4Ing
110              Text+       NFDI4Earth
111              Text+ NFDI4BioDiversity
```

1.3.2 2020

```r
nfdi_edges(2020)
```

```
# A tibble: 125 × 2
                from                to
              <fctr>            <fctr>
1         BERD\\@NFDI        KonsortSWD
2         BERD\\@NFDI             MaRDI
3         BERD\\@NFDI       NFDI4Memory
4         BERD\\@NFDI             Text+
5         DAPHNE4NFDI           FAIRmat
```

6	DAPHNE4NFDI	NFDI-MatWerk
7	DAPHNE4NFDI	NFDI4Cat
8	DAPHNE4NFDI	NFDI4Chem
9	DAPHNE4NFDI	NFDI4Health
10	DAPHNE4NFDI	NFDI4Ing
11	DAPHNE4NFDI	NFDI4Objects
12	DAPHNE4NFDI	PUNCH4NFDI
13	FAIRmat	DAPHNE4NFDI
14	FAIRmat	DataPLANT
15	FAIRmat	MaRDI
16	FAIRmat	NFDI-MatWerk
17	FAIRmat	NFDI4Cat
18	FAIRmat	NFDI4Chem
19	FAIRmat	NFDI4DataScience
20	FAIRmat	NFDI4Ing
21	FAIRmat	NFDIxCS
22	FAIRmat	PUNCH4NFDI
23	MaRDI	BERD\\@NFDI
24	MaRDI	FAIRmat
25	MaRDI	NFDI-MatWerk
26	MaRDI	NFDI-Neuro
27	MaRDI	NFDI4Cat
28	MaRDI	NFDI4Chem
29	MaRDI	NFDI4Ing
30	MaRDI	PUNCH4NFDI
31	NFDI-MatWerk	DAPHNE4NFDI
32	NFDI-MatWerk	DataPLANT
33	NFDI-MatWerk	FAIRmat
34	NFDI-MatWerk	MaRDI
35	NFDI-MatWerk	NFDI4Chem
36	NFDI-MatWerk	NFDI4DataScience
37	NFDI-MatWerk	NFDI4Ing
38	NFDI-MatWerk	NFDIxCS
39	NFDI-Neuro	DataPLANT
40	NFDI-Neuro	GHGA
41	NFDI-Neuro	NFDI4BioDiversity
42	NFDI-Neuro	NFDI4Culture
43	NFDI-Neuro	NFDI4Earth
44	NFDI-Neuro	NFDI4Health
45	NFDI-Neuro	NFDI4Ing
46	NFDI-Neuro	NFDI4Microbiota
47	NFDI4Agri	DataPLANT
48	NFDI4Agri	KonsortSWD
49	NFDI4Agri	NFDI4BioDiversity
50	NFDI4Agri	NFDI4Earth
51	NFDI4Agri	NFDI4Health
52	NFDI4Agri	NFDI4Immuno
53	NFDI4Agri	NFDI4Microbiota

54	NFDI4DataScience	KonsortSWD
55	NFDI4DataScience	MaRDI
56	NFDI4DataScience	NFDI-MatWerk
57	NFDI4DataScience	NFDI4BioDiversity
58	NFDI4DataScience	NFDI4Cat
59	NFDI4DataScience	NFDI4Chem
60	NFDI4DataScience	NFDI4Culture
61	NFDI4DataScience	NFDI4Health
62	NFDI4DataScience	NFDI4Ing
63	NFDI4DataScience	NFDI4Microbiota
64	NFDI4DataScience	NFDIxCS
65	NFDI4Earth	DataPLANT
66	NFDI4Earth	GHGA
67	NFDI4Earth	KonsortSWD
68	NFDI4Earth	NFDI4Agri
69	NFDI4Earth	NFDI4BioDiversity
70	NFDI4Earth	NFDI4Cat
71	NFDI4Earth	NFDI4Chem
72	NFDI4Earth	NFDI4Culture
73	NFDI4Earth	NFDI4Health
74	NFDI4Earth	NFDI4Ing
75	NFDI4Earth	NFDI4Objects
76	NFDI4Immuno	GHGA
77	NFDI4Immuno	NFDI4Agri
78	NFDI4Immuno	NFDI4Health
79	NFDI4Immuno	NFDI4Microbiota
80	NFDI4Memory	BERD\\@NFDI
81	NFDI4Memory	KonsortSWD
82	NFDI4Memory	MaRDI
83	NFDI4Memory	NFDI4Culture
84	NFDI4Memory	NFDI4Objects
85	NFDI4Memory	Text+
86	NFDI4Microbiota	DataPLANT
87	NFDI4Microbiota	GHGA
88	NFDI4Microbiota	NFDI4Agri
89	NFDI4Microbiota	NFDI4BioDiversity
90	NFDI4Microbiota	NFDI4Chem
91	NFDI4Microbiota	NFDI4DataScience
92	NFDI4Microbiota	NFDI4Health
93	NFDI4Microbiota	NFDI4Immuno
94	NFDI4Microbiota	NFDI4Ing
95	NFDI4Objects	KonsortSWD
96	NFDI4Objects	NFDI4Agri
97	NFDI4Objects	NFDI4BioDiversity
98	NFDI4Objects	NFDI4Culture
99	NFDI4Objects	NFDI4Earth
100	NFDI4Objects	NFDI4Memory
101	NFDI4Objects	Text+

```
102          NFDI4SD      NFDI4Culture
103          NFDI4SD  NFDI4DataScience
104          NFDI4SD       NFDI4Memory
105          NFDI4SD      NFDI4Objects
106          NFDIxCS           FAIRmat
107          NFDIxCS             MaRDI
108          NFDIxCS         NFDI4Chem
109          NFDIxCS  NFDI4DataScience
110          NFDIxCS        NFDI4Earth
111          NFDIxCS          NFDI4Ing
112       PUNCH4NFDI       DAPHNE4NFDI
113       PUNCH4NFDI           FAIRmat
114       PUNCH4NFDI              GHGA
115       PUNCH4NFDI             MaRDI
116       PUNCH4NFDI        NFDI4Earth
117       PUNCH4NFDI          NFDI4Ing
118       PUNCH4NFDI           NFDIxCS
119            Text+        KonsortSWD
120            Text+ NFDI4BioDiversity
121            Text+      NFDI4Culture
122            Text+        NFDI4Earth
123            Text+          NFDI4Ing
124            Text+       NFDI4Memory
125            Text+      NFDI4Objects
```

2 Data vectors

Following we present and explain the data and its origin.

2.1 Funded

The values whether a consortia is funded or not cannot be calculated. Therefore we already passed the information as a column to the consortia above.

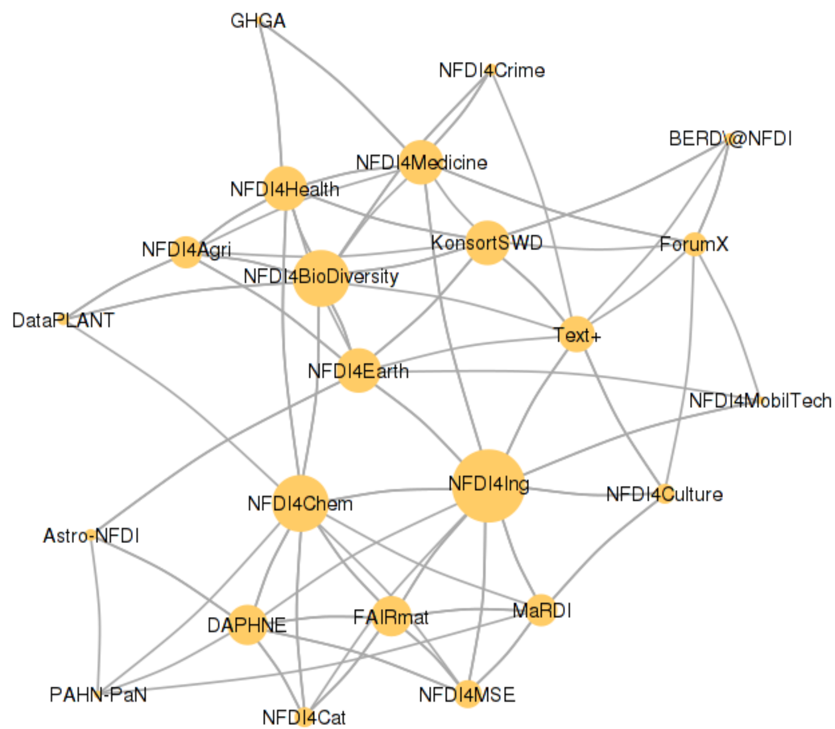
The column `funded` has either the value `0` (consortium has *not* been funded) or `1` (consortium has been funded). Since we look at the consortia at the time of turning in binding Letters of Intent, for 2019 there is no value at all.

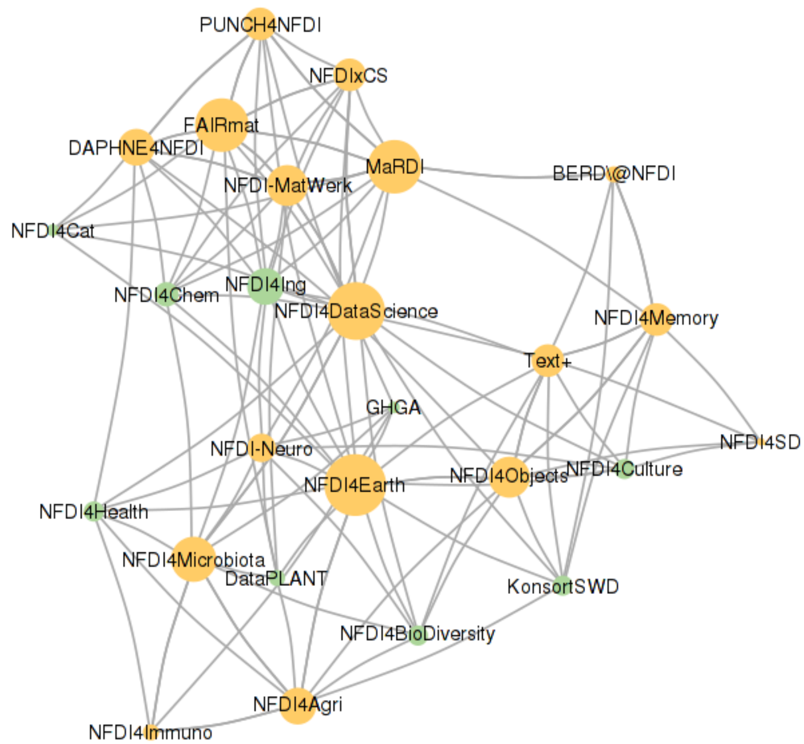
```
nfdi_funded <- function(nfdi_edges_year) {

  df <- data.frame(get(paste0("nfdi_nodes_",nfdi_edges_year)))

  names(df)[1] <- "Name"
  rownames(df) <- 1:nrow(df)

  df %>% select(1,3)

}
```

```
# nfdi_funded(2019)
# nfdi_funded(2020)
```

Now we can plot the networks and for 2020 we will highlight the funded consortia.

```
nfdi_plot<- function(nfdi_edges_year) {

  nfdi_presettings(nfdi_edges_year,FALSE)

    colrs <- c("#ffcc66","#acd69a")

    nfdi_local_network <- function() {
        plot(nfdi_network_year,
        vertex.color        = colrs[V(nfdi_network_year)$funded + 1 ],
        vertex.frame.color  = colrs[V(nfdi_network_year)$funded + 1 ]
        )
    }
    pdf(paste0("network_",nfdi_edges_year,".pdf"))
     nfdi_local_network()
    dev.off()

    nfdi_local_network()

}

nfdi_plot(2019)
nfdi_plot(2020)
```


GHGA
NFDI4Crime
BERD\@NFDI
NFDI4Medicine
NFDI4Health
KonsortSWD
ForumX
NFDI4Agri
NFDI4BioDiversity
DataPLANT
Text+
NFDI4Earth
NFDI4MobilTech
NFDI4Ing
NFDI4Culture
NFDI4Chem
Astro-NFDI
MaRDI
FAIRmat
DAPHNE
PAHN-PaN
NFDI4MSE
NFDI4Cat



2.2 NFDI conference system (`group`)

Each consortium has been allocated to a group which is known as the NFDI conference system.

In the first column you see the consortia's name in the second column (`group`) the number of the allocated group. There are five groups and each has a special color code.

No.	Conference group	Color code
(1)	Medizin	`#f5ac9f`
(2)	Lebens- und Geowissenschaften	`#e43516`
(3)	Geistes- und Sozialwissenschaften	`#f9b900`
(4)	Ingenieurwissenschaften und Mathemaik	`#007aaf`

No.	Conference group	Color code
(5)	Chemie und Physik	`#6ca11d`

```
nfdi_group <- function(nfdi_edges_year) {

  df <- data.frame(get(paste0("nfdi_nodes_",nfdi_edges_year)))

  names(df)[1] <- "Name"
  rownames(df) <- 1:nrow(df)

  df %>% select(1,2)
}

# nfdi_group(2019)
# nfdi_group(2020)
```

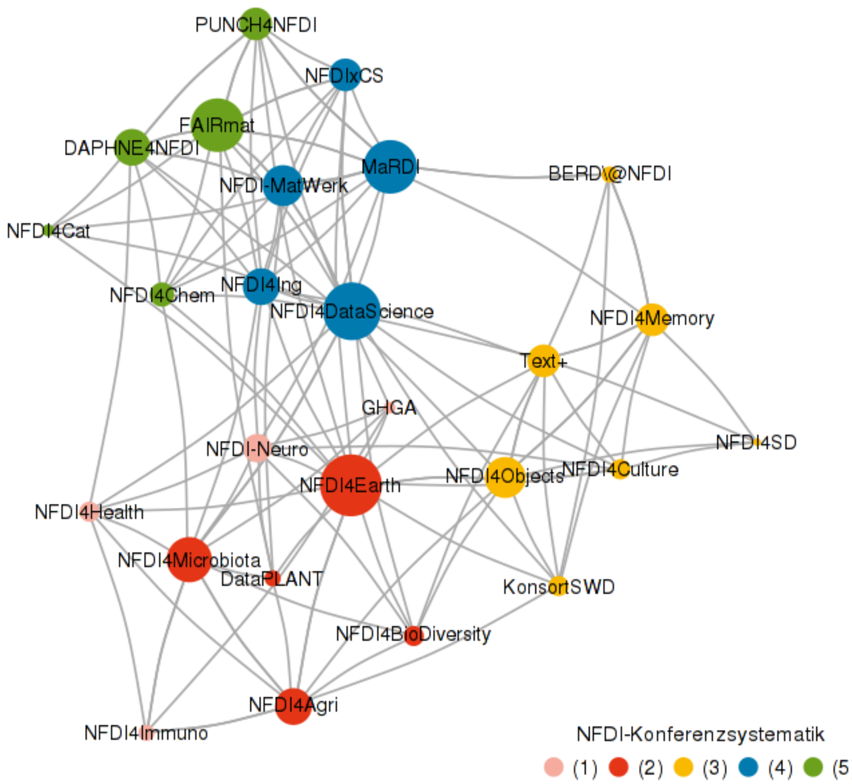
2.2.1 Plotting grouped network

First we are defining a common legend for the networks, since we want to be able to differentiate the various groups the consortia are colored in.

```
nfdi_plot_legend <- function(){

       legend("bottomright",   # x-position
       title  = "NFDI-Konferenzsystematik", # title
       legend = c(
           "(1)",
           "(2)",
           "(3)",
           "(4)",
           "(5)"
       ),  # the text of the legend
       col    = nfdi_color_code,  # colors of lines and points beside the␣
 ↪legend text
       pch    = 20,      # the plotting symbols appearing in the legend
       bty    = "n",     # no frame, the type of box to be drawn around the␣
 ↪legend (n=no frame)
       cex    = .75,     # character expansion factor relative to current␣
 ↪par("cex").
       pt.cex = 2,        # expansion factor(s) for the points
             ncol = 5,
)
}
```

Now we can go on and plot the network with colored consortia nodes.

```r
nfdi_plot_group <- function(nfdi_edges_year) {

  nfdi_presettings(nfdi_edges_year,FALSE)


  pdf(paste0("network_group_",nfdi_edges_year,".pdf"))
  plot(nfdi_network_year)
  nfdi_plot_legend()
  dev.off()


  plot(nfdi_network_year)
  nfdi_plot_legend()

}

# nfdi_plot_group(2019)
nfdi_plot_group(2020)
```



2.3 Cluster

The value of the column `cluster` has been calculated by the function `cluster_optimal`. Each consortia is allocated to a cluster.

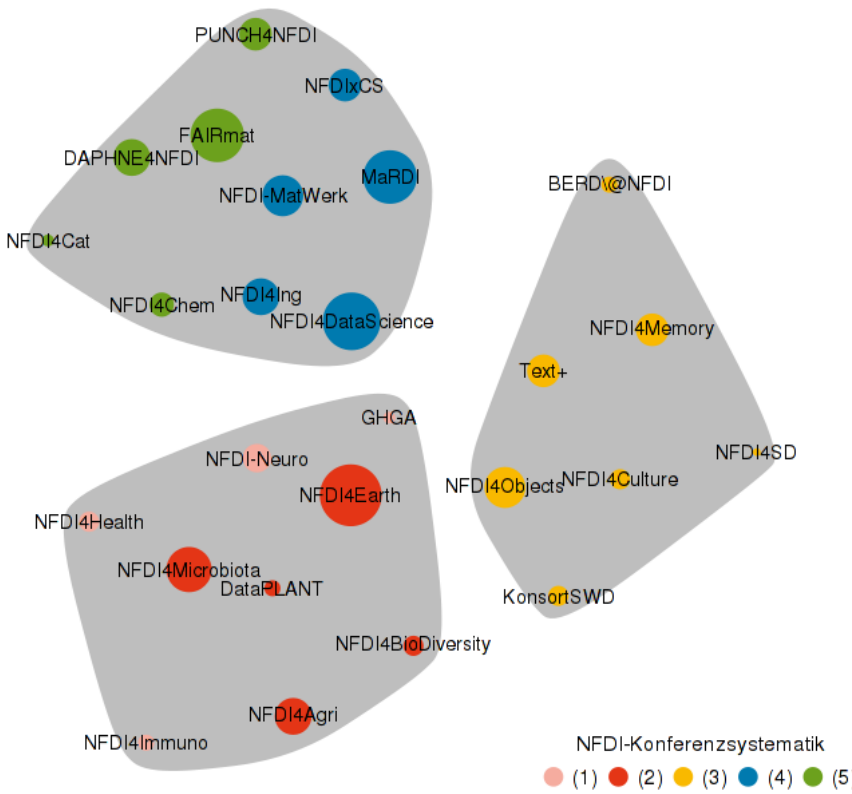
```
nfdi_list_cluster <- function(nfdi_edges_year) {
  nfdi_presettings(nfdi_edges_year,TRUE)

  liste <- list(y=c(membership(cluster_optimal(nfdi_network_year))))
   df = with(data=liste,
             expr=data.frame(y))
```

```
   df <- cbind(Name = rownames(df), df)
   rownames(df) <- 1:nrow(df)
   names(df) <- c("Name","cluster")


  df

}

# nfdi_list_cluster(2019)
# nfdi_list_cluster(2020)
```

2.3.1 Plotting clustered network

With the algorithm `cluster_optimal` we can do a community detection.

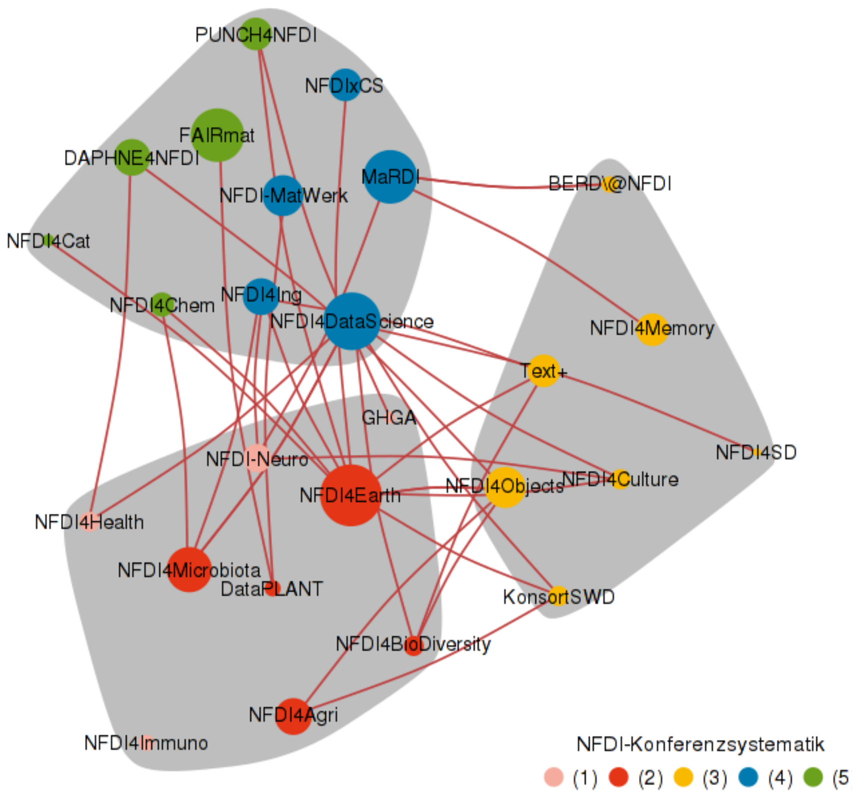
```
nfdi_plot_cluster <- function(nfdi_edges_year) {

  nfdi_presettings(nfdi_edges_year,FALSE)

  nfdi_network_year_cluster <- cluster_optimal(nfdi_network_year)

  nfdi_local_network <- function() {
    plot(nfdi_network_year_cluster,
       nfdi_network_year,
       edge.color          = NA,
       col                 = nfdi_color_groups, #color of nodes
       mark.col            = "grey",      # color groups
       mark.border         = NA,          # no border color
     )

    nfdi_plot_legend()
  }

    pdf(paste0("network_cluster_",nfdi_edges_year,".pdf"))
     nfdi_local_network()
    dev.off()

     nfdi_local_network()
}

# nfdi_plot_cluster(2019)
nfdi_plot_cluster(2020)
```



Now we are hightlighting only the connections of consortia between different clusters.

```
nfdi_plot_cluster_trans <- function(nfdi_edges_year) {

  nfdi_presettings(nfdi_edges_year,FALSE)

  nfdi_network_year_cluster <- cluster_optimal(nfdi_network_year)

    nfdi_local_network <- function() {
     plot(nfdi_network_year_cluster,
       nfdi_network_year,             # loading data frame
     col                 = nfdi_color_groups, #color of nodes
     mark.col            = "grey",       # color groups
```

```
     mark.border        = NA,           # no border color
     edge.color = c(NA, "#bf4040")[crossing(nfdi_network_year_cluster,␣
↪nfdi_network_year) + 1], # color of edges
     )
    nfdi_plot_legend()
    }


    pdf(paste0("network_cluster_trans_",nfdi_edges_year,".pdf"))
      nfdi_local_network()
    dev.off()

   nfdi_local_network()

}


# nfdi_plot_cluster_trans(2019)
nfdi_plot_cluster_trans(2020)
```



2.4 Amount of edges

There are three different ways of counting edges to a node: all, incoming ones and outgoing ones.

All edges (`degree.total`)

We get all the edges with the function

`degree(<GRAPH-OF-DATA-FRAME>, mode="total")`

and receive the following table.

```
nfdi_degree_total_filtered <- function(nfdi_edges_year) {

  nfdi_presettings(nfdi_edges_year,TRUE)
```

```
    df <- data.frame(degree(nfdi_network_year,
        mode = "total"))
    names(df)[1] <- "degree.total"
    df <- cbind(Name = rownames(df), df)
    rownames(df) <- 1:nrow(df)

    df

}

#nfdi_degree_total_filtered(2019)
#nfdi_degree_total_filtered(2020)
```

Incoming edges (`degree.in`)

For counting incomgin edges a directed network is necessary. Then the function `degree` with a different value for `mode` is applied.

`degree(<GRAPH-OF-DATA-FRAME>, mode="in")`

```
nfdi_degree_in <- function(nfdi_edges_year) {
    nfdi_presettings(nfdi_edges_year,TRUE)

    df <- data.frame(degree(nfdi_network_year,
                            mode = "in"))
    names(df)[1] <- "degree.in"
    df <- cbind(Name = rownames(df), df)
    rownames(df) <- 1:nrow(df)

    df

}

# nfdi_degree_in(2019)
# nfdi_degree_in(2020)
```

Outgoing edges (`degree.out`)

As before the function `degree` with a different value for `mode` is applied.

`degree(<GRAPH-OF-DATA-FRAME>, mode="out")`

```
nfdi_degree_out <- function(nfdi_edges_year) {
    nfdi_presettings(nfdi_edges_year,TRUE)

    df <- data.frame(degree(nfdi_network_year,
                            mode = "out"))
    names(df)[1] <- "degree.out"
    df <- cbind(Name = rownames(df), df)
```

```r
    rownames(df) <- 1:nrow(df)
    df

}

#nfdi_degree_out(2019)
#nfdi_degree_out(2020)
```

2.5 Vertex and edge betweenness centrality (`betweenness`)

The vertex betweenness has been calculated by the function `betweenness()`[2].

```r
nfdi_betweenness <- function(nfdi_edges_year) {
    nfdi_presettings(nfdi_edges_year,TRUE)

    df <- data.frame(betweenness(nfdi_network_year,
                                 v=V(nfdi_network_year)))

    names(df)[1] <- "betweenness"
    df <- cbind(Name = rownames(df), df)
    rownames(df) <- 1:nrow(df)
  df

}

#nfdi_betweenness(2019)
#nfdi_betweenness(2020)
```

2.6 Closeness centrality

> Clo[se]ness centrality measures how many steps is required to access every other vertex from a given vertex.[3]

Thereby we can differentiate between three different ways of closeness centrality.

`closeness.total`

`closeness(<GRAPH-OF-DATA-FRAME>, mode="total")`

```r
nfdi_closeness_total <- function(nfdi_edges_year) {
    nfdi_presettings(nfdi_edges_year,TRUE)

    df <- data.frame(closeness(nfdi_network_year,
                               mode="total"))

    names(df)[1] <- "closeness.total"
    df <- cbind(Name = rownames(df), df)
```

[2] https://igraph.org/r/doc/betweenness.html
[3] https://igraph.org/r/doc/closeness.html

```r
   rownames(df) <- 1:nrow(df)
  df

}

#nfdi_closeness_total(2019)
#nfdi_closeness_total(2020)
```

closeness.in

closeness(<GRAPH-OF-DATA-FRAME>, mode="in")

```r
nfdi_closeness_in <- function(nfdi_edges_year) {
   nfdi_presettings(nfdi_edges_year,TRUE)

   df <- data.frame(closeness(nfdi_network_year,
                              mode="in"))

   names(df)[1] <- "closeness.in"
   df <- cbind(Name = rownames(df), df)
   rownames(df) <- 1:nrow(df)
  df

}

#nfdi_closeness_in(2019)
#nfdi_closeness_in(2020)
```

closeness.out

closeness(<GRAPH-OF-DATA-FRAME>, mode="out")

```r
nfdi_closeness_out <- function(nfdi_edges_year) {
   nfdi_presettings(nfdi_edges_year,TRUE)

   df <- data.frame(closeness(nfdi_network_year,
                              mode="out"))

   names(df)[1] <- "closeness.out"
   df <- cbind(Name = rownames(df), df)
   rownames(df) <- 1:nrow(df)
  df

}

#nfdi_closeness_out(2019)
#nfdi_closeness_out(2020)
```

3 Data Collection

Here is an overview of all the data we have gathered so far.

```
nfdi_data_set <- function(nfdi_edges_year) {
  d1 <- nfdi_funded(nfdi_edges_year)
  d2 <- nfdi_group(nfdi_edges_year)
  d3 <- nfdi_list_cluster(nfdi_edges_year)
  d4 <- nfdi_degree_in(nfdi_edges_year)
  d5 <- nfdi_degree_out(nfdi_edges_year)
  d6 <- nfdi_degree_total_filtered(nfdi_edges_year)
  d7 <- nfdi_betweenness(nfdi_edges_year)
  d8 <- nfdi_closeness_in(nfdi_edges_year)
  d9 <- nfdi_closeness_out(nfdi_edges_year)
  d10 <-nfdi_closeness_total(nfdi_edges_year)

  df <- d1 %>% inner_join(d2, by = "Name") %>%
             full_join(d3, by = "Name") %>%
             full_join(d4, by = "Name") %>%
             full_join(d5, by = "Name") %>%
             full_join(d6, by = "Name") %>%
             full_join(d7, by = "Name") %>%
             full_join(d8, by = "Name") %>%
             full_join(d9, by = "Name") %>%
             full_join(d10, by = "Name") %>%
             arrange(Name) %>%
             replace(., is.na(.), "")

  write.csv(df, paste0("nfdi_data_set",nfdi_edges_year,".csv"))
   df
}
```

3.1 2019

```
nfdi_data_set(2019)
```

Name	funded	group	cluster	degree.in	degree.out	degree.total	betweenness	closeness.i
Astro-NFDI	0	5	1	2	3	5	5.975000	0.01408451
BERD\@NFDI	0	3	2	2	3	5	2.000000	0.0125000
DAPHNE	0	5	1	5	7	12	16.019048	0.0138888
DataPLANT	0	2	3	2	3	5	1.416667	0.0133333
FAIRmat	0	5	1	6	6	12	10.063889	0.0153846
ForumX	0	3	2	3	5	8	11.720635	0.0133333
GHGA	0	1	3	2	2	4	0.000000	0.0131578
KonsortSWD	0	3	3	8	5	13	33.279762	0.0161290
MaRDI	0	4	1	4	6	10	9.225000	0.0142857
NFDI4Agri	0	2	3	5	5	10	8.885714	0.0149253
NFDI4BioDiversity	0	2	3	8	8	16	57.990476	0.0175438
NFDI4Cat	0	5	1	3	4	7	0.000000	0.0133333
NFDI4Chem	0	5	1	9	7	16	75.065079	0.0172413
NFDI4Crime	0	2	3	2	3	5	1.616667	0.0138888
NFDI4Culture	0	3	2	4	3	7	7.630556	0.0149253
NFDI4Earth	0	2	3	8	5	13	48.296429	0.0175438
NFDI4Health	0	1	3	6	7	13	32.397222	0.0161290
NFDI4Ing	0	4	1	11	9	20	123.969841	0.0185185
NFDI4Medicine	0	1	3	6	7	13	54.310317	0.0166666
NFDI4MobilTech	0	4	2	1	3	4	5.461905	0.0135135
NFDI4MSE	0	5	1	4	5	9	2.880556	0.0142857
PAHN-PaN	0	5	1	4	0	4	0.000000	0.0200000
Text+	0	3	2	6	5	11	32.795238	0.0161290

3.2 2020

```
nfdi_data_set(2020)
```

Name	funded	group	cluster	degree.in	degree.out	degree.total	betweenness	closeness.in
BERD\@NFDI	0	3	1	2	4	6	9.9523810	0.00366300
DAPHNE4NFDI	0	5	2	3	8	11	9.7119048	0.00357142
DataPLANT	1	2	3	6	0	6	0.0000000	0.00418410
FAIRmat	0	5	2	5	10	15	14.0928571	0.00366300
GHGA	1	1	3	5	0	5	0.0000000	0.00414937
KonsortSWD	1	3	1	7	0	7	0.0000000	0.00425531
MaRDI	0	4	2	7	8	15	77.0333333	0.00380228
NFDI4Agri	0	2	3	4	7	11	28.5000000	0.00375939
NFDI4BioDiversity	1	2	3	7	0	7	0.0000000	0.00425531
NFDI4Cat	1	5	2	5	0	5	0.0000000	0.00420168
NFDI4Chem	1	5	2	8	0	8	0.0000000	0.00427350
NFDI4Culture	1	3	1	7	0	7	0.0000000	0.00423728
NFDI4DataScience	0	4	2	5	11	16	66.3333333	0.00374531
NFDI4Earth	0	2	3	6	11	17	51.9023810	0.00384615
NFDI4Health	1	1	3	7	0	7	0.0000000	0.00421940
NFDI4Immuno	0	1	3	2	4	6	0.3333333	0.00363636
NFDI4Ing	1	4	2	11	0	11	0.0000000	0.00432900
NFDI4Memory	0	3	1	4	6	10	29.0595238	0.00366300
NFDI4Microbiota	0	2	3	4	9	13	46.4166667	0.00374531
NFDI4Objects	0	3	1	5	7	12	41.5833333	0.00377358
NFDI4SD	0	3	1	0	4	4	0.0000000	0.00153846
NFDI-MatWerk	0	4	2	4	8	12	18.1428571	0.00371747
NFDI-Neuro	0	1	3	1	8	9	11.0595238	0.00362318
NFDIxCS	0	4	2	4	6	10	6.6166667	0.00366300
PUNCH4NFDI	0	5	2	3	7	10	10.7023810	0.00363636
Text+	0	3	1	3	7	10	8.5595238	0.00364963

4 Further stats

4.1 Wilcoxon rank sum test

```
nfdi_count_edges_out <- function(nfdi_edges_year) {

   nfdi_presettings(nfdi_edges_year,TRUE)
    df <- as.data.frame(table(nfdi_edges$from))
    df <- as.data.frame(df)
     names(df)[1] <- "Name"
     names(df)[2] <- "count"
    rownames(df) <- 1:nrow(df)
    df
}
#nfdi_count_edges_out(2019)
#nfdi_count_edges_out(2020)
```

```
wilcox.test(nfdi_count_edges_out(2019)$count,
                  nfdi_count_edges_out(2020)$count,
```

```
                  alternative="two.sided",
                  exact=F,
                  correct=T
     )
```

```
	Wilcoxon rank sum test with continuity correction

data:  nfdi_count_edges_out(2019)$count and nfdi_count_edges_out(2020)$count
W = 80, p-value = 0.002314
alternative hypothesis: true location shift is not equal to 0
```

4.2 Amount of Letters of Intent with collaborations mentioned

```
nfdi_count_LoI <- function(nfdi_edges_year) {

   nfdi_presettings(nfdi_edges_year,TRUE)

   sum(table(nfdi_count_edges_out(nfdi_edges_year)))
}

# nfdi_count_LoI(2019)
# nfdi_count_LoI(2020)

print(paste0(2019,": ",nfdi_count_LoI(2019)))
print(paste0(2020,": ",nfdi_count_LoI(2020)))
```

```
[1] "2019: 22"
[1] "2020: 17"
```

4.3 Nodes and edges in the network

Nodes

We can easily count all the nodes in a network of a particular year by using the function `gorder`.[4]

```
nfdi_count_nodes <- function(nfdi_edges_year) {

    nfdi_presettings(nfdi_edges_year,TRUE)

    gorder(nfdi_network_year)

}

#nfdi_count_nodes(2019)
#nfdi_count_nodes(2020)
```

[4] https://igraph.org/r/doc/gorder.html

```
print(paste0(2019,": ",nfdi_count_nodes(2019)))
print(paste0(2020,": ",nfdi_count_nodes(2020)))
```

```
[1] "2019: 23"
[1] "2020: 26"
```

Edges

We can easily count all the nodes in a network of a particular year by using the function `gsize`.[5]

```
nfdi_count_edges <- function(nfdi_edges_year) {

    nfdi_presettings(nfdi_edges_year,TRUE)

    gsize(nfdi_network_year)

}

#nfdi_count_edges(2019)
#nfdi_count_edges(2020)

print(paste0(2019,": ",nfdi_count_edges(2019)))
print(paste0(2020,": ",nfdi_count_edges(2020)))
```

```
[1] "2019: 111"
[1] "2020: 125"
```

[5] https://igraph.org/r/doc/gsize.html