Supplementary Material Note 3:
Quantitative analysis of Knowledge Communities

Understanding emergence of the Sustainable Development Goals Research.

3.1. knowledge communities description.

The community network of Utrecht University is visualised in Figure. Each node represents a community or research group, consisting of a group of similar publications in terms of their bibliography. The communities coloured in green are the those defined as SDG communities, whereas the grey nodes are the other communities. The network shows that the SDG communities are located throughout the network and interact with many other SDG communities and other communities. This means that SDG research at Utrecht University is well embedded in the scientific landscape and can be found in different research areas and disciplines.

Figure S7: The Utrecht University network, with the SDG communities in green. Each node is a community, and each line represents a connection between two communities (based on the shared bibliography of publications within the community). Node size is the betweenness centrality score.

Overview of the main clusters and research areas

The seven main clusters of the SDG communities are represented in Figure S8. Most communities are in cluster 6, the medicine and public health cluster. This cluster is mostly concerned with Good Health and Wellbeing (SDG 3), Gender Equality (SDG 5) and Quality

Education (SDG 4). In the network we see that this cluster is mostly located at the lower half and left of the network, stretching out to the centre. This means that research on medicine and public health, and SDG 3, 5 and 4, is also important for other communities, and interacts with communities through a large part of the network. At the top of the network two clusters related to climate change, both ecological and physical, can be seen These clusters are mostly about Climate Action (SDG 13), Life Below Water (SDG 14) and Clean Water and Sanitation (SDG 6). They are located at the edge of the network, which indicates that the research concerned with climate change shows less interaction with other research areas and communities in the network. It is a niche of communities related to climate action. There is some interaction with non-SDG communities, geophysics and geography (see FigureS9, N3 – the orange cluster).

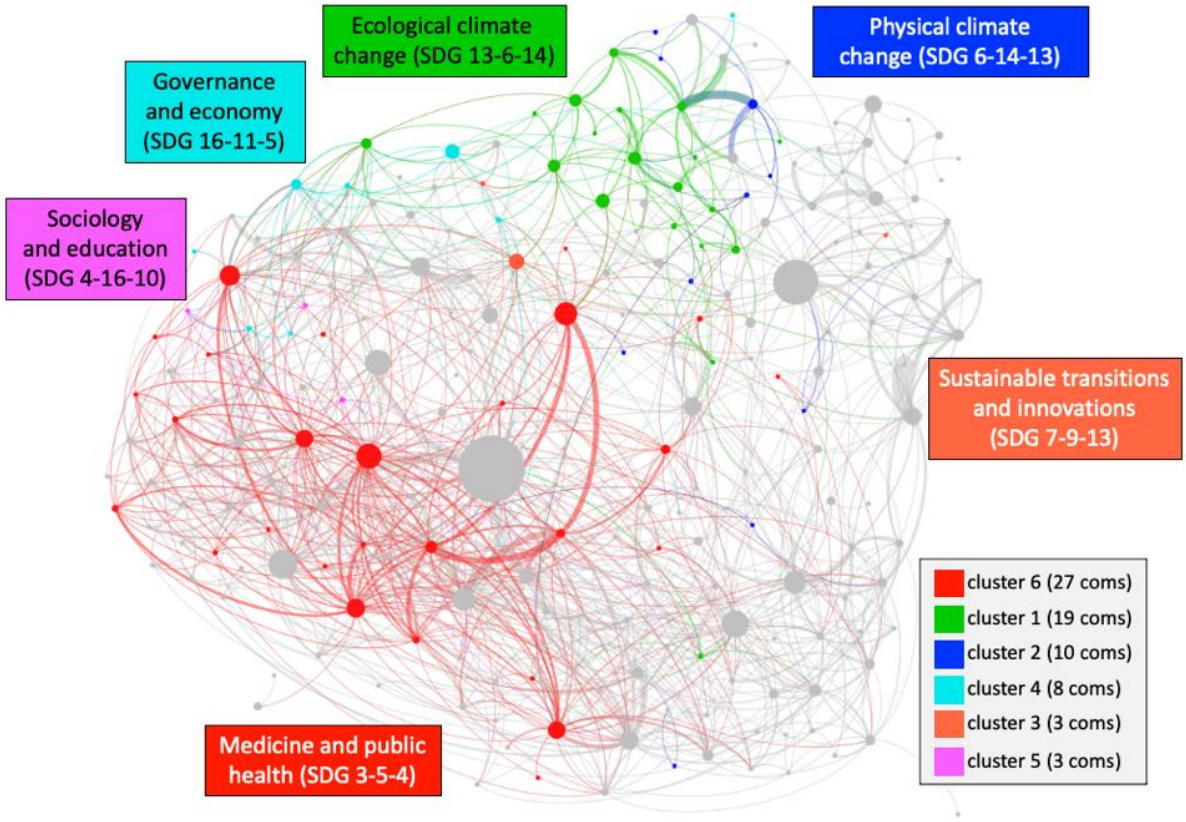

Figure S8: Network with the seven main clusters of the SDG communities, their main research area(s) and most common SDGs. Node size is the betweenness centrality score.

The seven main clusters of the other communities are represented in FigureS9. Cluster N4, biochemistry, medicine and veterinary, is the largest with 68 communities. It is located in the right bottom corner of the map, stretching up to the centre of the graph. On comparing the location of cluster N4 with Figure S8 there is an overlap with cluster 6 (medicine and public health). We also see that cluster N6, psychology and neurosciences, shows overlap with cluster 6. That means that these SDG communities and non-SDG communities interact with one another and knowledge circulates between these communities. We see the same for cluster N5 and cluster 4.

In Figure S8 there is a notable grey area at the right side of the network, where no SDG communities are located. In Figure we see that there are two clusters located in this grey area, cluster N1, astronomy and astrophysics, and cluster N2, physics, chemistry and material sciences. The research topics of these two clusters already indicate why there is scare SDG research in this area of the network,: the SDGs are not directly concerned with these topics.

The SDGs are more concerned with applied research, whereas these clusters conduct more fundamental research which forms the basis of other research, including other SDG research. What is interesting however, is how this fundamental research contributes to the SDGs, and how this knowledge circulates through the network to be used in SDG related research: i.e., how do these communities support and enable SDG research?

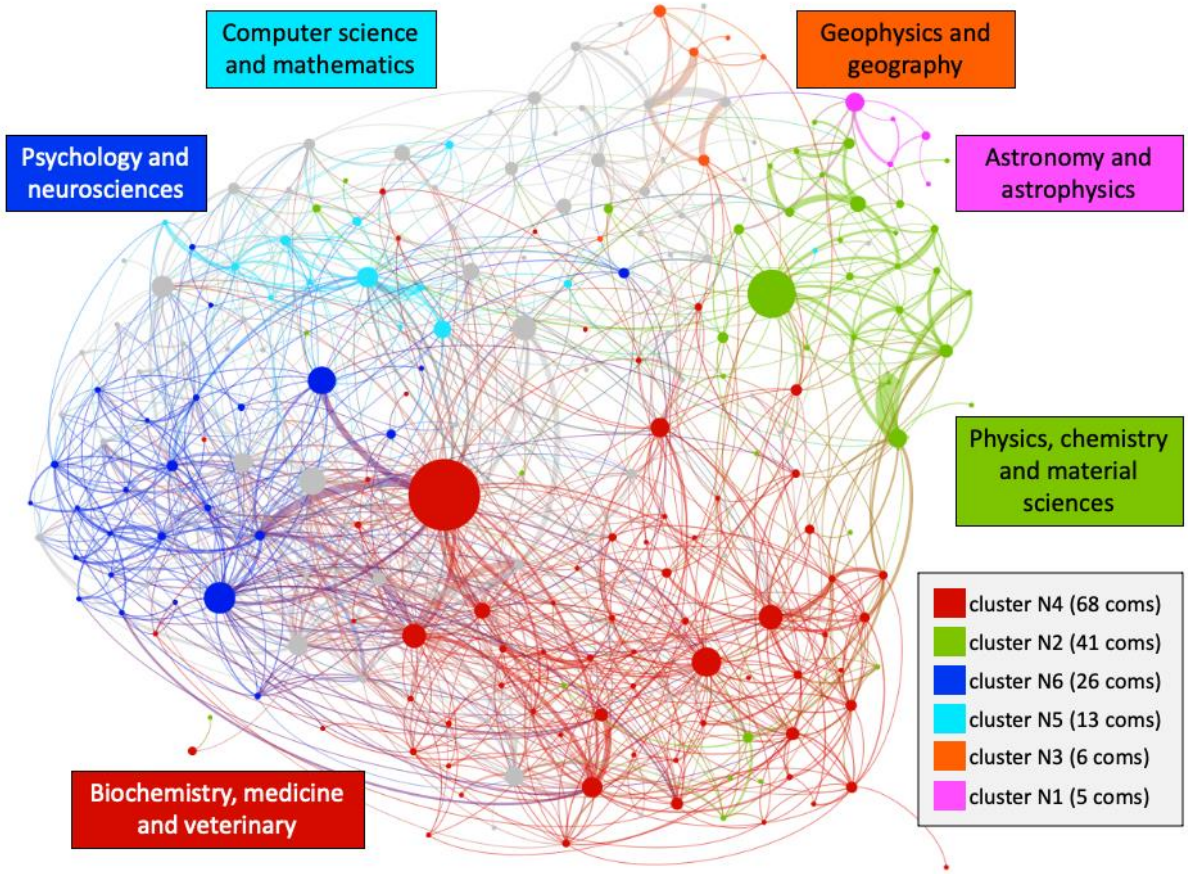

Figure S9: Network with the seven main clusters of the other communities and their main research area(s). Node size is the betweenness centrality score.

3.2. Description of the knowledge communities selected.

3.2.1 H.1: Community 176

Community 176 has a total of 290 publications, of which 144 are SDG-related publications. It has a betweenness centrality score of 0.027 (7th highest of the SDG communities). There are 26 different triads combinations in the community.

	SDGs		Triads		Research areas	
1	7	(39.8%)	7-7-7 (ST-ST-ST)	30.3%	Energy & fuels	(71x)
2	13	(18.8%)	11-7-7 (ST-ST-ST)	15.6%	Engineering	(41x)
3	4	(6.3%)	17-7-7 (FC-ST-ST)	9.0%	Physics	(21x)

Table S3: Main SDGs, triads and research areas of community 176

When we dive deeper into the characteristics of community 176 (Table S3) and analyse the most frequent SDGs in the publications of the community, we see that the three most prominent SDGs are SDG 7, 13 and 4, which are about affordable and Clean Energy, Climate Action, and

Quality Education. The most frequent triads in community 176 are 7-7-7, 11-7-7 and 17-7-7, of which the first two fall in the ST-ST-ST group and the last one in the FC-ST-ST group. The last one shows an interesting connection between the sociotechnical systems and framework conditions.
The most frequent research areas of the publications in this community are energy & fuels, engineering, and physics, which are in line with SDG 7 and 13. They seem somewhat less related to SDG 4 – quality education, but the publications could be about physics education.

The most frequent keywords of the publications in community 176 are shown in the wordcloud in Figure. Most are related to green and sustainable energy topics, such as 'self-consumption' and 'photovoltaic(s)'. The keywords 'demand response' and 'demand side management' indicate a business and economics side, which could be more related to SDG 4, while 'smart grids' is more related to engineering.

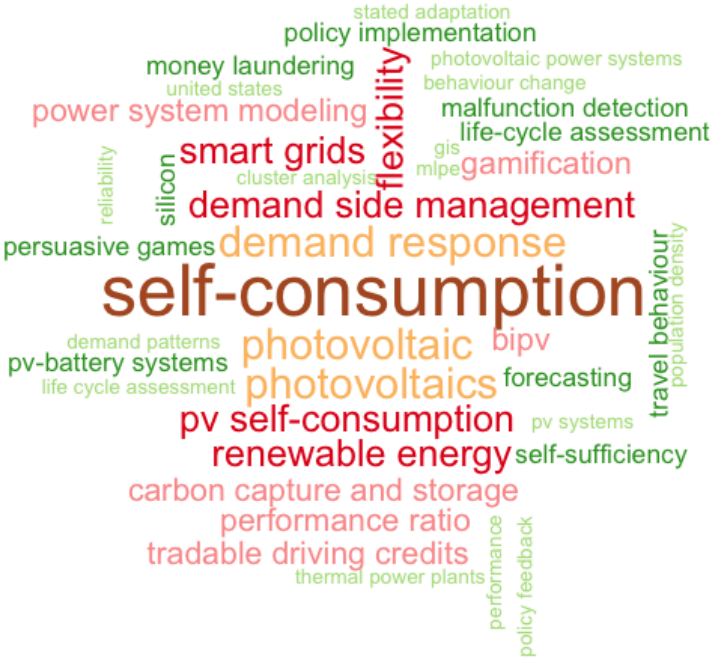

Figure S10: Wordcloud of the author keywords of the publications in community 176. The size of the keyword indicates the frequency of the keyword in the publications, with a maximum of 40 words.

To analyse the connections that community 176 has with other communities and its location in the network, we look at the egocentric network. The ego network in Figure S11 shows the communities with which community 176 has a direct link. Community 176 is connected to many SDG communities related to a diverse set of SDGs, such as SDG 3, 11 and 16 (health care, sustainable cities, and peace, justice and strong institutions). The thickness of the line indicates how strong the connection is (when the publications in a community have more shared bibliographies with the publicities in another community, there is a stronger connection between the two communities). Communities 176 shows a strong connection with community 226, which is about SDG 16, and with community 180, which is about SDG 3.

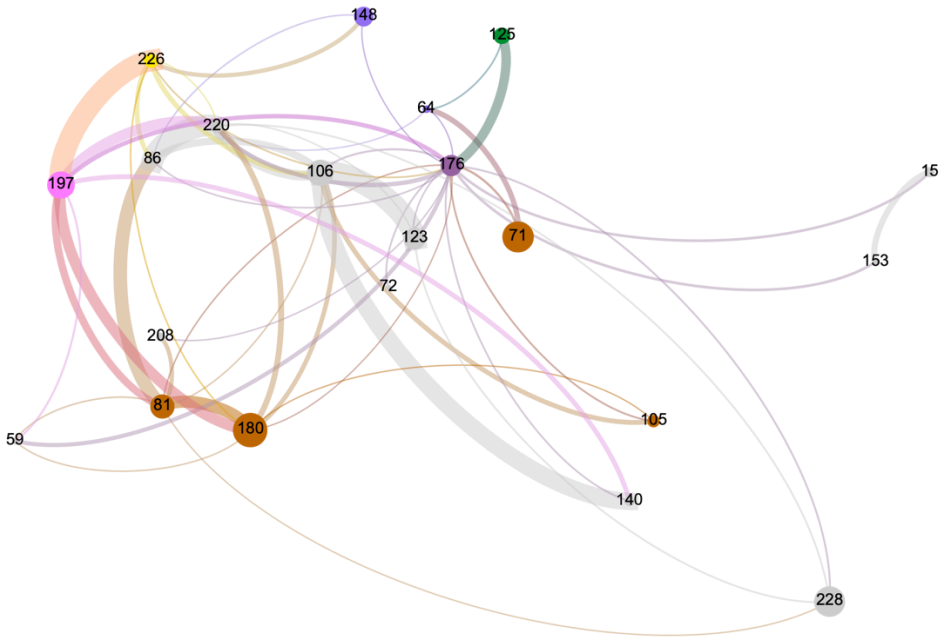

Figure S11: Ego network of community 176. Node size is betweenness centrality.

Temporal analysis

The temporal analysis shows how community 176 developed over the years. This community was formed from six communities in T6 (2015-2017). Three of these six communities emerged in T6, meaning they did not exist in T5 (2012-2014) and were formed from new publications that were published after 2014. These three communities, combined, are accountable for 25% of the publications of community 176. From the other three communities in T6 of which community 176 was formed, only a part of their publications went to community 176. These communities were split in T6, where a small part of the publications went to community 176, but most publications went to other communities. These latter three communities show a longer existence over the years, whereas each of them can be traced back to T2 (2003-2005). These three communities are accountable for 35% of the publications in community 176. The other 40% of the publications in community 176 are published after 2017 (after T6) and were included in community 176 in T7. Figure 2 shows the timeline of the communities over the past 17 years, up to T2. Each rectangle represents a community, and each column represents a timeframe. In the column T7 is the community of interest, in this case community 176. The figure shows from which communities a community is formed, and how many publications of the community in the previous timeframe (T-1) went to the community of interest (indicated by the large grey arrow). The small grey arrow indicates into how many communities a community split. For example, community 101 in T6 split into 3 communities, where 22% of its publications went to community 176 (and the other 78% of the publications went to 2 other communities in T7). The numbers in T2 indicate from how many communities the community in T3 is formed, i.e., community 35 in T3 came from one community, whereas community 116 comes from two merged communities.

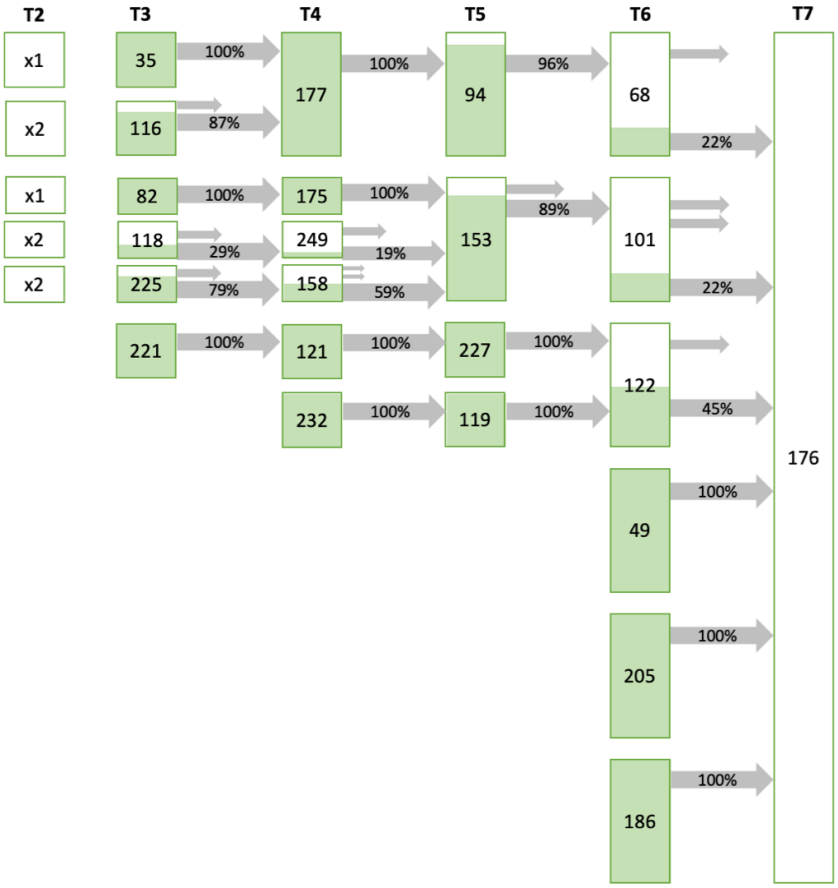

Figure S12: The emergence and development of community 176

To see from which research areas community 176 comes, we first look into the six communities in T6, from which community 176 is formed. These have the following most frequent research area(s):

- Physics, and material science (community 68)
- Transportation, and business & economics (community 101)
- Environmental sciences & ecology, and government & law (community 122)
- Linguistics and psychology (community 49)
- Education & educational research (community 205)
- Energy & fuels (community 186)

Each community has different research areas, which indicates the diversity of community 176. The community is formed from six different communities with many different research areas. It is most likely that from the first three communities (community 68, 101 and 122), which were split from T6>T7, the research publications with similar topics went to community 176, and the more unrelated to other communities. However, from the last three communities, all of the publications went to community 176 (100% of the publications in the community in T6 went to community 176 in T7). This means that community 176 contains the publications on each of these research areas and is a combination of these research areas, combined with the publications from the first three communities. Furthermore, we look at the main research areas of each of the communities over the past 20 years. The diversity of the research areas in community 176 is visualised in Figure 11 (main manuscript), which shows a streamgraph of the research areas over the years. Each colour represents a research area, and the thickness of the line represents the number of publications within that research area in each year. For each year, the top three most frequent research areas for each community are used. The figure shows that community 176 has been formed from many different research areas over the past 20 years.

Concluding, community 176 is a diverse community which is formed from different communities and research areas over time.

3.2.3: Community 154

Community 154 has a total of 317 publications, of which 141 are SDG-related publications. It has a betweenness centrality score of 0.00069. There are 24 different triads combinations in the community.

	SDGs	**Triads**	**Research areas**
1	4 (37.1%)	16-16-4 (FC-FC-ST) 12.5%	Psychology (80x)
2	10 (24.3%)	16-4-4 (FC-ST-ST) 11.7%	Sociology (30x)
3	-	10-16-4 (FC-ST-TD) 10.0%	Education & educational research (25x)

Table S4: Main SDGs, triads and research areas of community 154

In community 154 the most prominent SDGs are SDG 4 and 10, which are about quality education and reduced inequalities (Table S4). The most frequent triads in community 154 are 16-16-4, 16-4-4 and 10-16-4, of which the first two combine framework conditions and sociotechnical systems. The last one combines all three SDG categories and is of interest for the transformative potential of this research.

The most frequent research areas are psychology, sociology and education & educational research. Both psychology and sociology could be related to SDG 10 but are both very broad. Education & educational research is very closely linked to SDG 4. The most frequent keywords of the publications in community 154 are shown in the wordcloud in Figure 1. There is a strong focus on immigrants and ethnicity, combined with religion and discrimination. There are also multiple keywords related to education, such as ‘teacher-student relationship’, and ‘classroom management’.

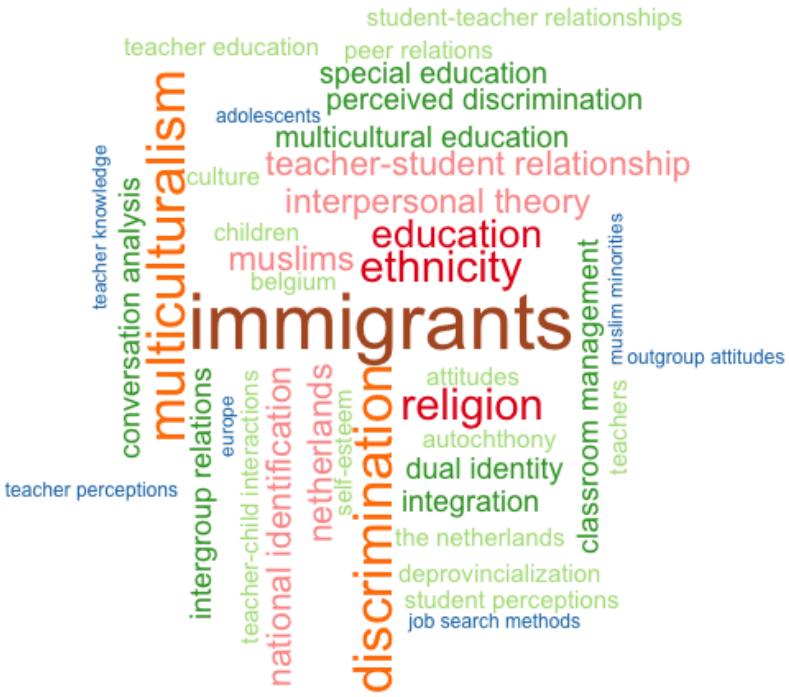

Figure 1: Wordcloud of the author keywords of the publications in community 154. The size of the keyword indicates the frequency of the keyword in the publications, with a maximum of 40 words.

The ego network of community 154 shows that this community has less connections as compared to community 176 (Figure 2). This is also shown in its much lower betweenness centrality score. It does show a strong connection with community 83, which is about SDG 11 (sustainable cities) mostly. It also shows a connection with community 226, which has as main SDG 16 (peace, justice and strong institutions).

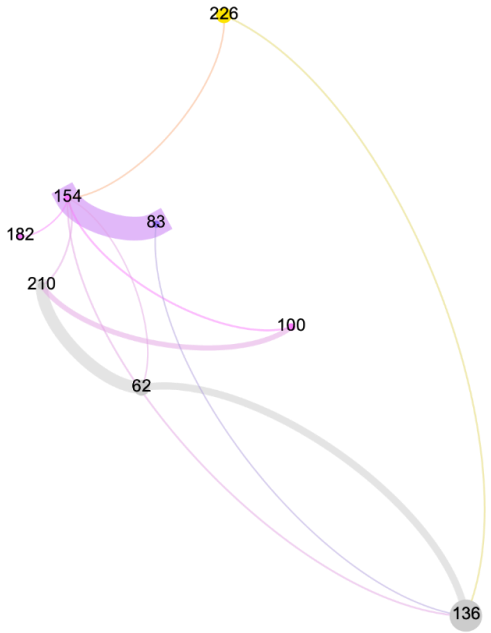

Figure 2: Ego network of community 154. Node size is betweenness centrality.

Temporal analysis

The temporal analysis shows how community 154 developed over the years (Figure S15). This community already existed in T6 (2015-2017) as one community, in T7 only publications were added to the community. Looking at Figure S15 we see that community 154 comes from three different communities that split and merged over the years. None of the communities existed before T3. Moreover, the development of community 154 is fairly consistent over the years. Only once was a community split (community 251 in T5), but mostly it is comprised of communities that are merged together.

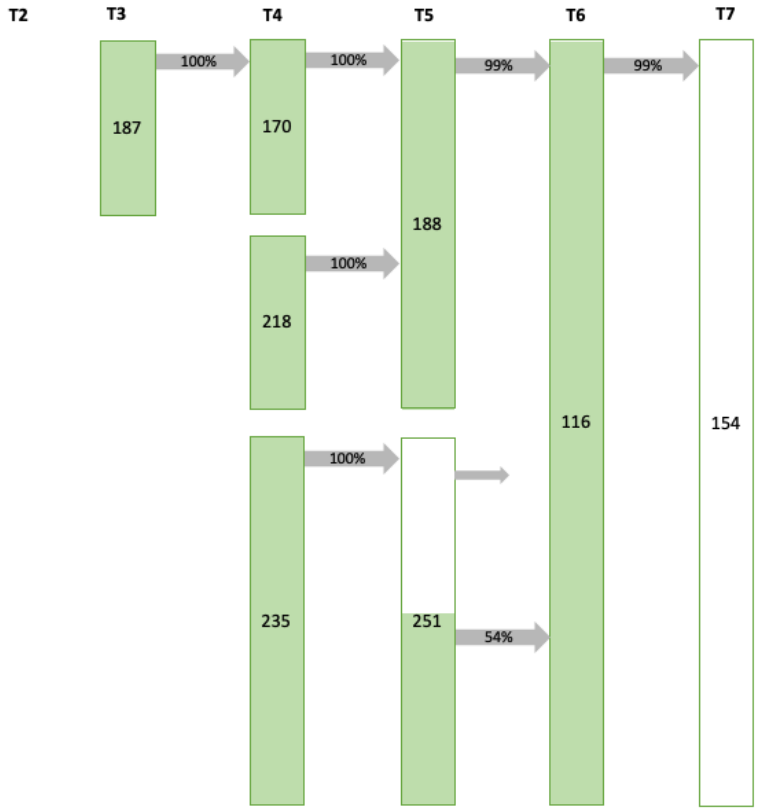

Figure S15: The emergence and development of community 154

Looking at the main research areas over the years, the streamgraph of community 154 in Figure 3 shows that there have only been seven different research areas in the communities from which community 154 has been formed. Largely the research areas stayed the same over the years, whereas sociology (pink) and psychology (light blue) have been the most important research areas over the past 20 years. Education & educational research was introduced in 2008 and is still an important research area in the community. This shows how stable the research areas of community 154 are over the years. Psychology, sociology and education & educational research are the main research areas of community 154 and have been the same throughout the years. The only deviant research area is demography but, looking back at the wordcloud in Figure S13 this is also closely related to the topics of community 154.

Community 154 is a very stable community throughout the time, with a strong research focus that stayed the same over the past 20 years. If we compare this to community 197 we see that community 154 has a more specific focus which stayed similar over the years, whereas community y 197 has been build up from many diverse research areas and communities.

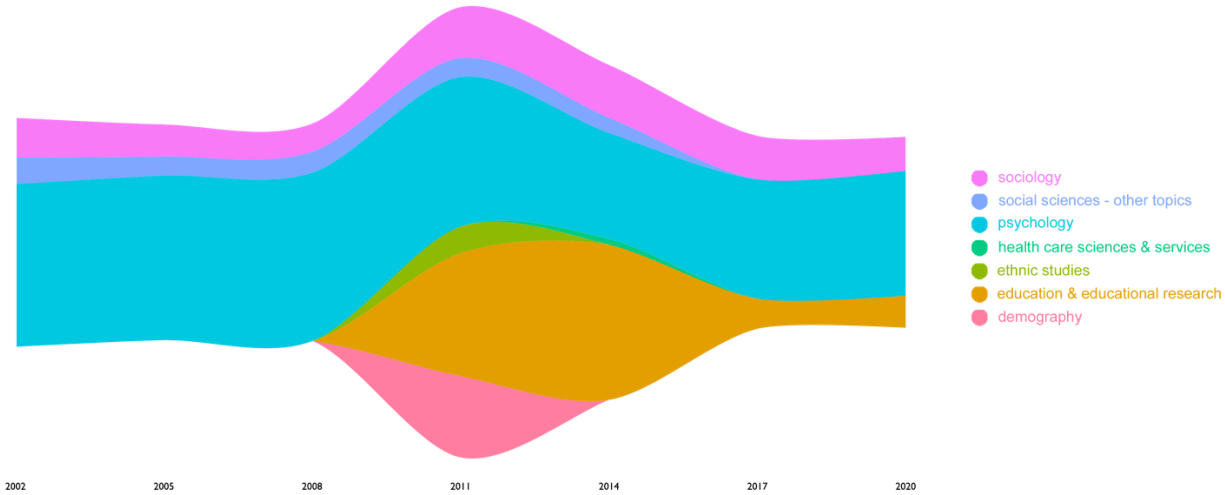

Figure 3: Streamgraph of the man research areas of community 154 over the past 20 years (2000-2020)

Triad analysis

Community 154 was selected because it has a relative high frequency of FC-ST-TD triads. The most frequent triad that combines SDGs in these categories is 10-16-4, there are 12 of these triads in community 154. A couple of these triads are analysed in more depth, to see which publications that combine the three SDG categories are related to each other. Table S5 shows the titles of these publications.

Publication 1	**Publication 2**	**Publication 3**
WOS:000214608400018-1 Social psychology and multiculturalism (SDG 10)	WOS:000250050400007-1 Tolerance of Muslim beliefs and practices: age related differences and context effects (SDG 4)	WOS:000255944700002-1 Muslim and non-Muslim adolescents' reasoning about freedom of speech and minority rights (SDG 16)
WOS:000222212600001-1 Global and ethnic self-esteem in school context: minority and majority groups in the Netherlands (SDG 4)	WOS:000169604000006-1 Peer victimisation and self-esteem of ethnic minority group children (SDG 16)	WOS:000181918500004-1 Positive and negative self-esteem among ethnic minority early adolescents: social and cultural sources and threats (SDG 10)
WOS:000220517900006-1 Psychological disidentification with the academic domain among ethnic minority adolescents in the Netherlands (SDG 4)	WOS:000169604000006-1 Peer victimisation and self-esteem of ethnic minority group children (SDG 16)	WOS:000180296900001-2 Perceptions of ethnic discrimination by minority and majority early adolescents in the Netherlands (SDG 10)

Table S5: Titles of publications in the FC-ST-TD triads in community 154

We see that even though these are triads with SDGs in all three categories, the publications themselves are very similar to each other and are actually about similar subjects (like

employment of immigrants) but analysed from a slightly different point of view (leading to the different SDG attributed to the publication).

3.2.3: Community 71

Community 71 has a total of 366 publications, of which 223 are SDG publications. It has a betweenness centrality score of 0.045 (2nd highest of the SDG communities). There are 98 different triads combinations in the community.

	SDGs	Triads	Research areas
1	3 (57.2%)	11-3-3 (ST-ST-ST) 21.8%	Environmental sciences & ecology (141x)
2	11 (23.2%)	3-3-3 (ST-ST-ST) 19.0%	Public, environmental & occupational health (94x)
3	-	11-11-3 (ST-ST-ST) 11.6%	Toxicology (44x)

Table S6: Main SDGs, triads and research areas of community 71

In community 71 the most prominent SDGs are SDG 3 and 11, which are about health care and sustainable cities (Table S6). The most frequent triads are 11-3-3, 3-3-3 and 11-11-3, which all combine SDGs that fall within the sociotechnical systems.

The most frequent research areas are environmental sciences & ecology, public, environmental & occupational health and toxicology. The first one is closely related to SDG 11, sustainable cities, whereas the latter two are connected to healthcare, SDG 3. The most frequent keywords of the publications in community 71 are shown in the wordcloud in Figure 4. There is a strong focus on air pollution and particulates, as well as traffic. This relates to a combination of SDG 3 and SDG 11, for example research on air pollution (SDG 3 – health care) due to traffic in cities (SDG 11 – sustainable cities).

Figure 4: Wordcloud of the author keywords of the publications in community 71. The size of the keyword indicates the frequency of the keyword in the publications, with a maximum of 40 words.

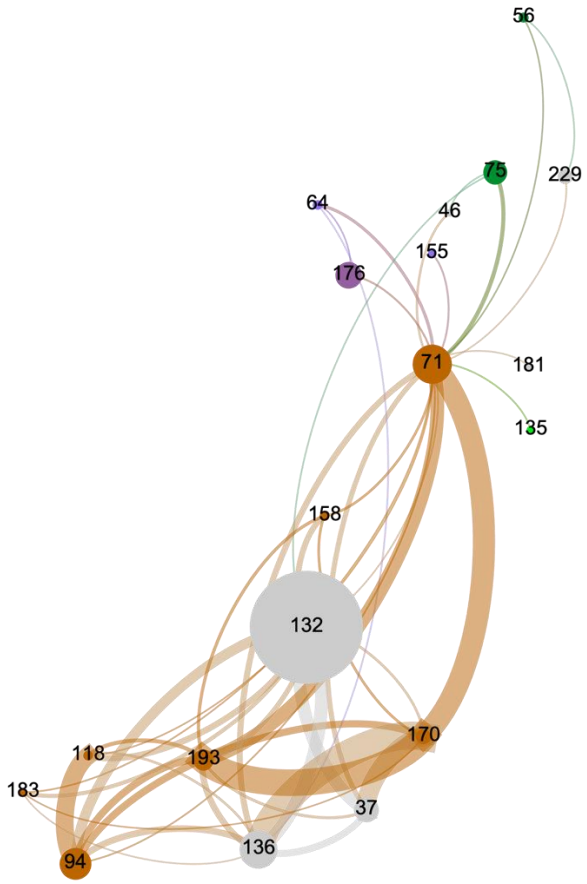

Figure 5: Ego network of community 71. Node size is betweenness centrality.

The ego network of community 71 in Figure 5 shows multiple strong connections to communities related to SDG 3 (especially community 170 and 139). On the other side, it shows connections to communities about a diverse range of SDGs, such as community 176 about SDG 7, community 155 about SDG 11 and community 75 about SDG 13.

Temporal analysis

Looking at the development of community 71 over the years we see an even more stable development than we saw for community 154 (Figure 6). Over the years, the community hasn't split or been merged from different communities. Community 71 has been existing since T1 as one community. The research areas from community 71 over the years have also been very stable. If we look at the streamgraph in Figure 7 we see that environmental sciences & ecology is the main research area of community 71 and has been since 2000. The research areas general & internal medicine and meteorology & atmospheric sciences disappeared from the community over the years, and public, environmental & occupational health and toxicology were introduced over the years.

Concluding, community 71 is a well-established community with a clear research focus. It has been very stable over the years and has a central location in the network. If we compare this to the previous two communities, we see that community 71 is even more stable over the years than community 154. This could be explained by the very specific research topic of community 71 and the clear research focus, which stayed important over the years.

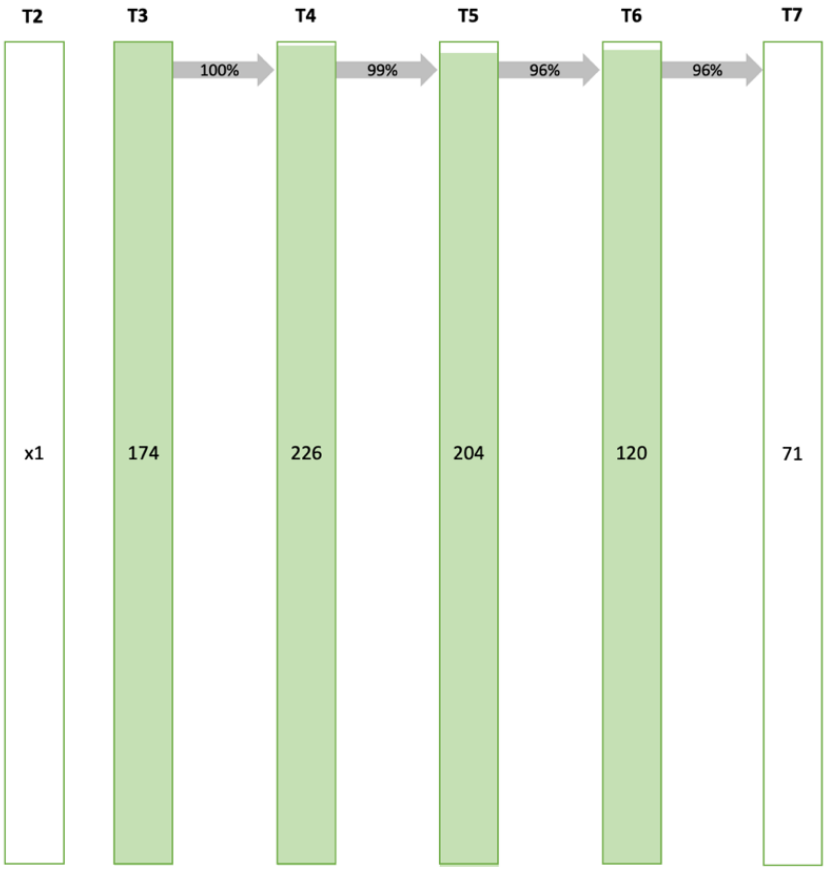

Figure 6: The emergence and development of community 71

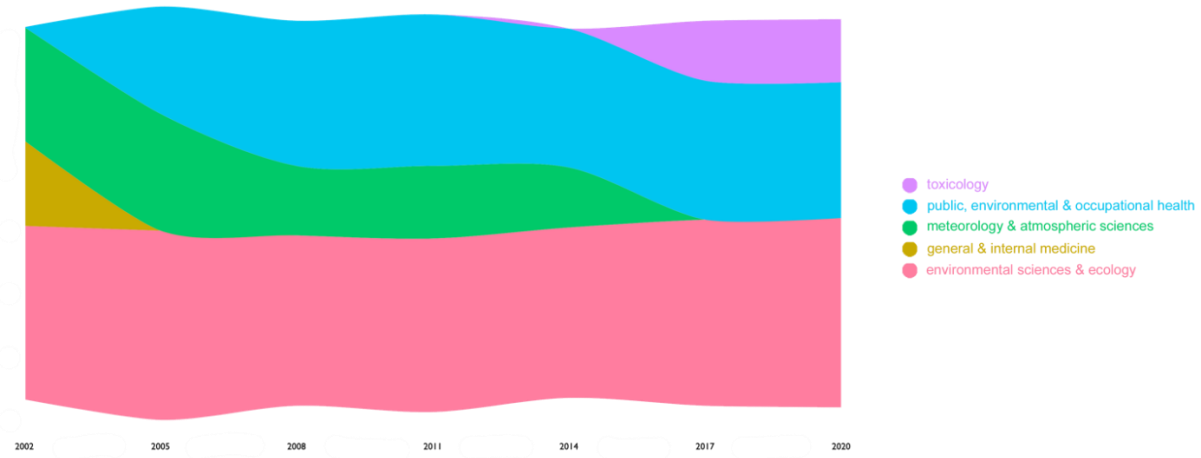

Figure 7: Streamgraph of the man research areas of community 71 over the past 20 years (2000-2020)

3.2.4: Community 197

Community 197 has a total of 366 publications, of which 223 are SDG publications. It has a betweenness centrality score of 0.038 (3th highest of the SDG communities) There are 37 different triad combinations in the community.

	SDGs		Triads			Research areas	
1	4	(30.7%)	4-4-4	(ST-ST-ST)	17.9%	Education & educational research	(47x)
2	16	(23.3%)	12-4-4	(ST-ST-TD)	9.7%	Health care sciences & services	(35x)
3	3	(21.3%)	16-5-5	(FC-TD-TD)	8.2%	Public, environmental & occupational health	(28x)

Table S7: Main SDGs, triads and research areas of community 197

In community 197 the most prominent SDGs are SDG 4, 16 and 3, which are about quality education, peace, justice & strong institutions and healthcare (Table S7). The most frequent triads are 4-4-4, 12-4-4 and 16-5-5. The latter two show an interaction between different SDG categories, whereas 12-4-4 combines sociotechnical systems with transversal directionalities and 16-5-5 combines framework conditions with transversal directionalities. Interesting in this community is that even though its second most frequent SDG is SDG 16, it has only *one* triad that combines SDG 16 (framework conditions) with sociotechnical systems and transversal directionalities (this triad is 12-16-4). So even though they have research on SDGs in all three categories, they are not able to combine this research.

The most frequent research areas are education & educational research, health care sciences & services and public, environmental & occupational health. The first one is related to SDG 4, whereas the latter two are both strongly related to SDG 3. The public health and health care services indicates SDG 16. The most frequent keywords of the publications in community 197 are shown in the wordcloud in Figure 8. There is a large focus on education, human rights and gender. This is mostly in line with SDG 4 and SDG 16. There are also multiple keywords indicating SDG 3, such as ‘vaccination’ and ‘health behaviour’.

Figure 8: Wordcloud of the author keywords of the publications in community 197. The size of the keyword indicates the frequency of the keyword in the publications, with a maximum of 40 words.

The ego network of community 197 shows many relations to communities related to SDG 3, such as community 94, 180 and 81 (Figure 9). It also shows interaction with communities on

SDG 13 (community 150 and 124), on SDG 16 (community 226) and SDG 5 (community 209). It is connected to 17 non-SDG communities and 16 SDG communities.

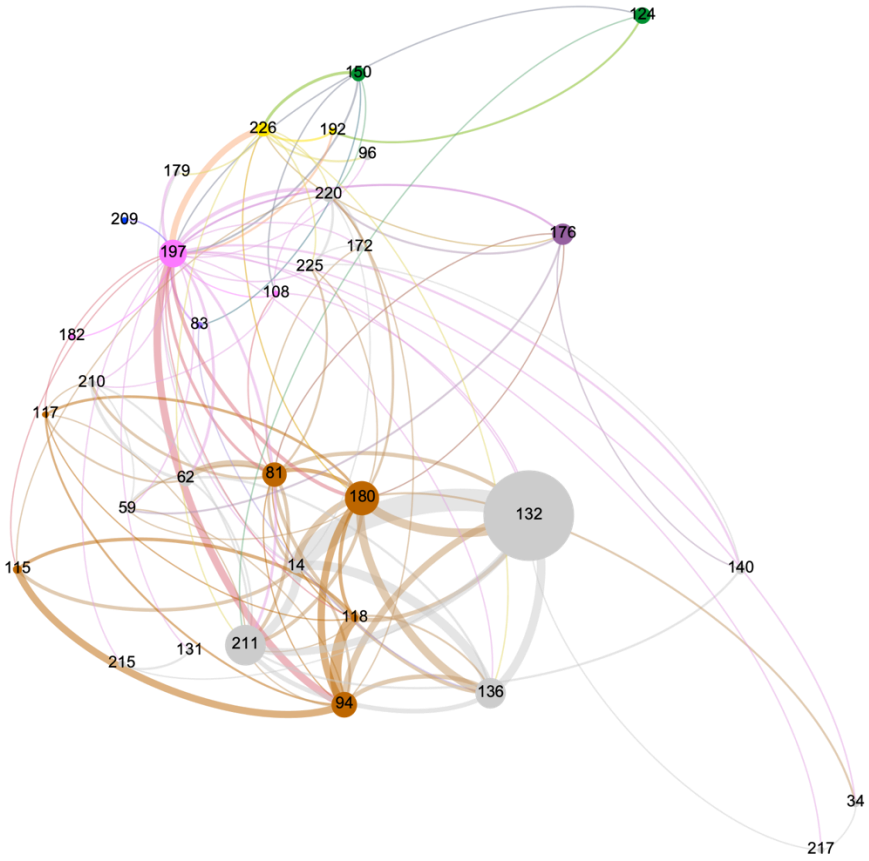

Figure 9: Ego network of community 71. Node size is betweenness centrality.

Temporal analysis

Analysing at the development of community 197 we see that this community has its roots in many different communities (Figure 10). Community 197 in T7 is formed from three communities in T6, of which one was fully attributed to community 197 (153), and the other two were split and a part of the publications went to community 197. Going further back, we see that the origin of community 197 goes back to T2 and comes from 8 communities in T2.

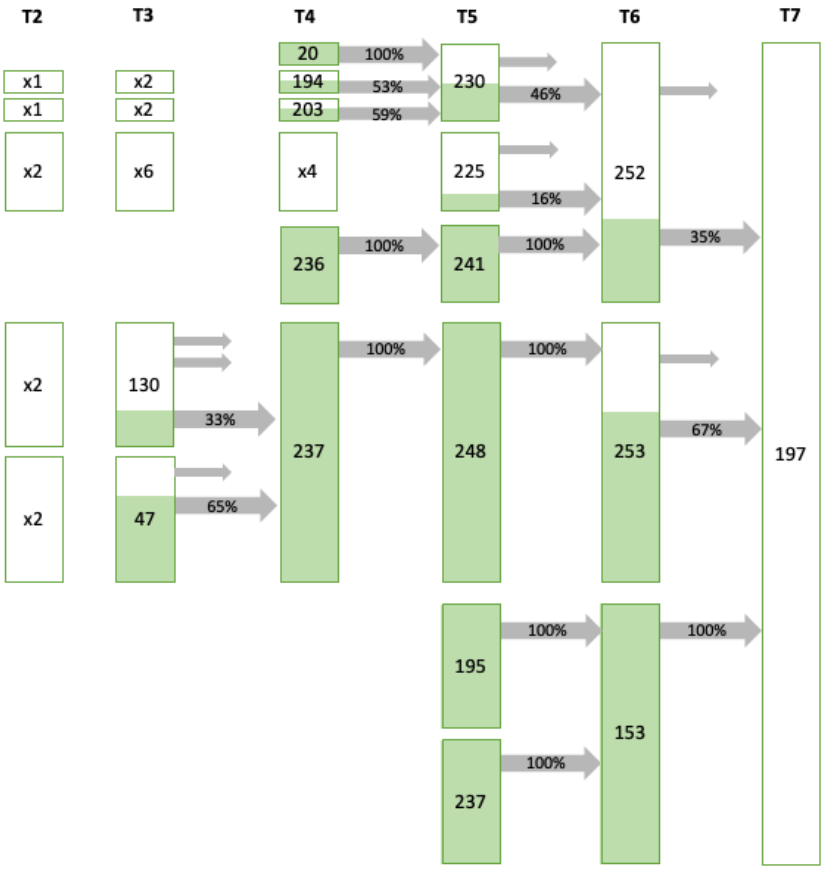

Figure 10: The emergence and development of community 197

The research areas of the many communities that 197 is formed upon are diverse. Similar to community 176, it is based on a diverse range of research areas. This is also because many communities were split, whereby the least relevant part of the research went to another community. If we look at the streamgraph of the research areas over the past years (Figure 11) we see the research areas are diverse, from zoology to oncology and cell biology. This diverse set of main research areas of the communities shows the diversity of community 197. It could also explain its high betweenness centrality, whereas over the years many communities split partly into community 197 and partly into different communities, meaning that community 197 has a relation with many other communities, that are closely related to it.

Concluding, community 197 is, similar to community 176, a very diverse community which is formed from different communities and build upon different research areas over time. It integrates different knowledge topics and is therefore an important community in the network.

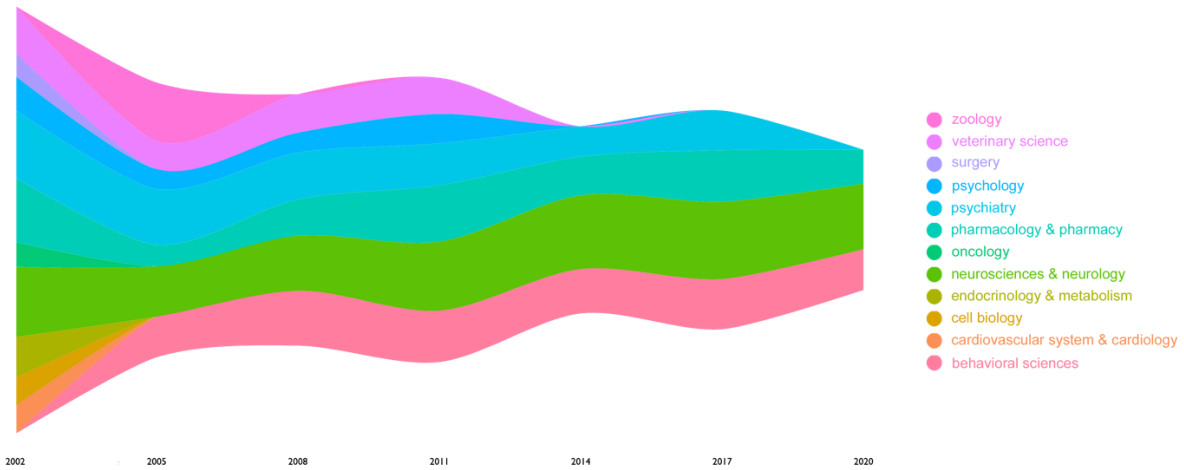

Figure 11: Streamgraph of the man research areas of community 176 over the past 20 years (2000-2020)