A Q-analysis of the Patras Student Hypernetwork Data

Jeffrey Johnson, Patras PhD School & The Open University, 24th July 2011

[1] Introduction

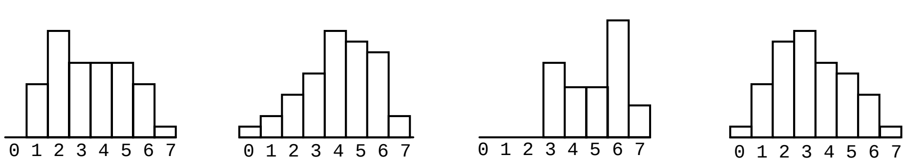
On 21st July 2011 a group of fifty two students and teachers at the Patras PhD School were asked to rate the first forty six of the following things according to how much they liked them, 0 meaning not at all and 6 or 7 meaning very much. The last eight were contributed by member of the group and also rated.

The list is rather heterogeneous and the data rather poor, since some people used a 0-6 scale while a few used a 0-7 scale. Also the last eight items were not all rated by all participants. Furthermore some participants complained that rating some items would be ambiguous with the possibility, for example, of liking very much to listen to one kind of music and hating to listen to another. This illustrates a common problem that complex systems data are often *incomplete and inconsistent.*

1 food	19 snow	37 psychology
2 wine	20 sunshine	38 sociology
3 smoking	21 wind	39 political_science;
4 reading_science	22 holidays	40 economics
5 writing_science	23 computer_games	41 history
6 talking_science	24 facebook-networking	42 geography
7 parties	25 email	43 literature
8 football	26 bankers	44 art
9 sport	27 children	45 statistics
10 gardening	28 pets	46 mathematics
11 watch_TV	29 chat_friends;	47 sleeping
12 listen_music	30 be_with_family	48 travelling
13 play_music-sing	31 teamwork	49 complexity_science;
14 read_novels	32 administration	50 deaming-daydreaming
15 cooking	33 organise_events	51 give_presentations
16 walk-hike	34 physics	52 philosophy
17 swimming	35 chemistry	53 democracy
18 rain	36 biology	54 meditation

Table 1. The things people rated according to their likes and dislikes.

What can be done with messy data like this? At the level of the whole population, one possibility is to plot the responses as shown in Figure 1. Unremarkably, from this it can be concluded that for this group of people, the majority like sunshine and snow and the majority dislike wind and rain. This could be done at the population level for everything listed. This would show that no-one disliked liked food, 48% liked wine while 26% disliked it, 88% disliked smoking while 12% liked it, and so on.



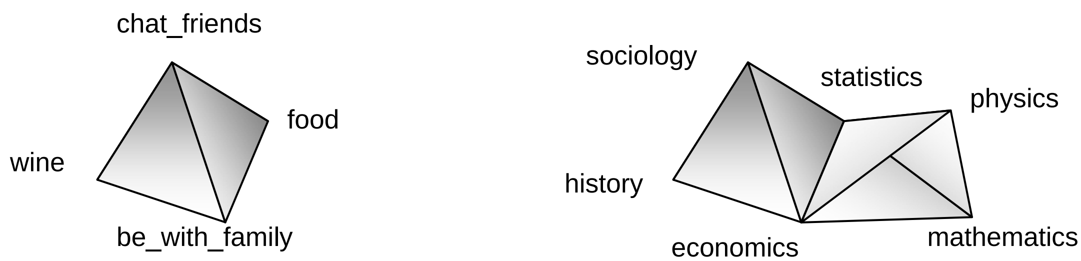
(a) rain (b) snow (c) sunshine (d) wind

Figure 1. Show and sunshine are preferred to rain and wind.

[2] A Q-analysis of the Student Hypernetwork

Recall that a *relational simplex* is a set of vertices combined by an *n*-ary relation. For example, a person p_i might have a simplex $\sigma\langle$food, wine, chat_friends, be_with_family; $R_{questionnaire+}\rangle$ where these four vertices are combined by them being positive answers to the questionnaire, as shown in Figure 2(a). A person with such a simplex is likely to enjoy dinner parties with their family and friends. In its simplest form, a *hypernetwork* is any set of relational simplices.

Two simplices are defined to be *q-near* if they share a *q*-dimensional face. For example the simplices in Figure 2(b) share the 1-dimensional face $\langle$economics, statistics$\rangle$ and they are 1-near.

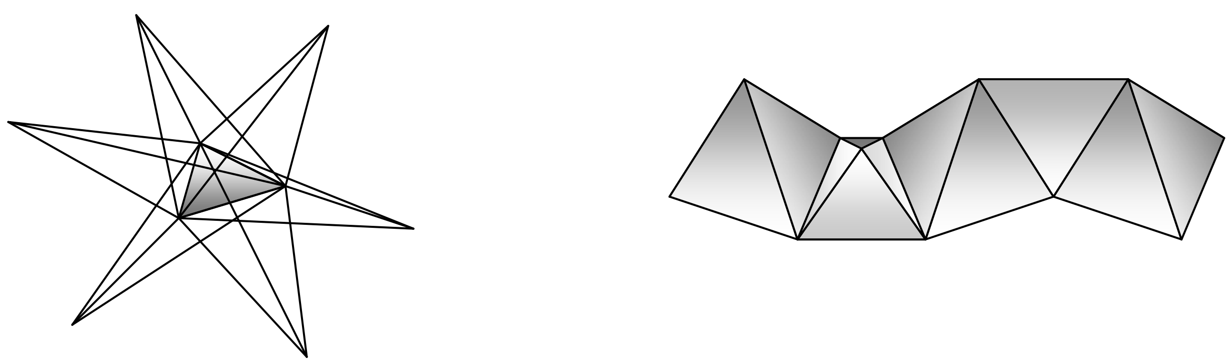


(a) A relational simplex (b) connected relational simplices

Figure 2. Relational simplices (a) predisposed to social meals and (b) to discussing economic trends

The q-nearness relation is reflexive and symmetric and its transitive closure is an equivalence relation that partitions a hypernetwork into *q-connected components*. The *hub* of a component is defined to be the intersection of all its simplices. There are two major types of component, those that have an intersection and form star-like configurations (Figure 3(a)) and those that make chains which have no non-trivial hub (Figure 3(b)).

A *Q-analysis* of a hypernetwork is a listing of all its q-connected components for all values of q. The Q-analysis for the student's like relation is given in Table 2.



(a) simplices connected through a 2-hub (b) a hub-free chain of 1-connected simplices

Figure 3. Two major classes of *q*-connected components

```
Q-analysis:
patras_student_Qdata-D_Q_analysis_A

 qq = 39
Component 39-1  1 simplex  : |S02
Component 39-2  1 simplex  : |S41

 qq = 38
Component 38-1  1 simplex  : |S02
Component 38-2  1 simplex  : |S41

 qq = 37
Component 37-1  1 simplex  : |S02
Component 37-2  1 simplex  : |S41

 qq = 36
Component 36-1  1 simplex  : |S02
Component 36-2  1 simplex  : |S33
Component 36-3  1 simplex  : |S41

 qq = 35
Component 35-1  1 simplex  : |S02
Component 35-2  1 simplex  : |S33
Component 35-3  1 simplex  : |S41

 qq = 34
Component 34-1  1 simplex  : |S02
Component 34-2  1 simplex  : |S26
Component 34-3  1 simplex  : |S33
Component 34-4  1 simplex  : |S41

 qq = 33
Component 33-1  1 simplex  : |S02
Component 33-2  1 simplex  : |S14
Component 33-3  1 simplex  : |S26
Component 33-4  1 simplex  : |S33
Component 33-5  1 simplex  : |S41

 qq = 32
Component 32-1  1 simplex  : |S02
Component 32-2  1 simplex  : |S08
Component 32-3  1 simplex  : |S14
Component 32-4  1 simplex  : |S26
Component 32-5  1 simplex  : |S33
Component 32-6  1 simplex  : |S41

 qq = 31
Component 31-1  1 simplex  : |S02
Component 31-2  1 simplex  : |S08
Component 31-3  1 simplex  : |S14
Component 31-4  1 simplex  : |S26
Component 31-5  1 simplex  : |S33
Component 31-6  1 simplex  : |S40
Component 31-7  1 simplex  : |S41

 qq = 30
Component 30-1  2 simplices: |S02 S26
Component 30-2  1 simplex  : |S05
Component 30-3  1 simplex  : |S08
Component 30-4  1 simplex  : |S14
Component 30-5  1 simplex  : |S21
Component 30-6  1 simplex  : |S33
Component 30-7  1 simplex  : |S40
Component 30-8  1 simplex  : |S41

 qq = 29 & 28
Component 29-1  4 simplices: |S02 S26
                              S33 S41
Component 29-2  1 simplex  : |S05
Component 29-3  1 simplex  : |S08
Component 29-4  1 simplex  : |S14
Component 29-5  1 simplex  : |S21
Component 29-6  1 simplex  : |S40
Component 29-7  1 simplex  : |S42
```

Table 2: Q-analysis

```
  qq = 27
Component 27-1  6 simplices: |S02 S14 S26 S33 S41 S40
Component 27-2  1 simplex  : |S05
Component 27-3  1 simplex  : |S08
Component 27-4  1 simplex  : |S21
Component 27-5  1 simplex  : |S29
Component 27-6  1 simplex  : |S42

 qq = 26
Component 26-1  8 simplices: |S02 S08 S14 S26 S33 S40 S41 S05
Component 26-2  1 simplex  : |S04
Component 26-3  1 simplex  : |S21
Component 26-4  1 simplex  : |S29
Component 26-5  1 simplex  : |S42

 qq = 25
Component 25-1  10 simplices:|S02 S08 S14 S26 S33 S40 S41
                              S42 S21 S05
Component 25-2  1 simplex  : |S04
Component 25-3  1 simplex  : |S17
Component 25-4  1 simplex  : |S29
Component 25-5  1 simplex  : |S36

 qq = 24
Component 24-1  1 simplex  : |S01
Component 24-2  11 simplices |S02 S08 S14 S26 S33 S40 S41 S42
                              S05 S21 S04
Component 24-3  1 simplex  : |S15
Component 24-4  1 simplex  : |S17
Component 24-5  1 simplex  : |S18
Component 24-6  1 simplex  : |S29
Component 24-7  1 simplex  : |S30
Component 24-8  1 simplex  : |S36
Component 24-9  1 simplex  : |S39

 qq = 23
Component 23-1  1 simplex  : |S01
Component 23-2  13 simplices:|S02 S08 S14 S26 S29 S33 S40 S41
                              S42 S05 S21 S04 S18
Component 23-3  1 simplex  : |S07
Component 23-4  1 simplex  : |S15
Component 23-5  1 simplex  : |S17
Component 23-6  1 simplex  : |S30
Component 23-7  1 simplex  : |S36
Component 23-8  1 simplex  : |S39

 qq = 22
Component 22-1  1 simplex  : |S01
Component 22-2  16 simplices:|S02 S08 S14 S26 S29 S33 S40 S41
                              S42 S05 S17 S21 S04 S18 S36 S39
Component 22-3  1 simplex  : |S07
Component 22-4  1 simplex  : |S09
Component 22-5  1 simplex  : |S15
Component 22-6  1 simplex  : |S25
Component 22-7  1 simplex  : |S30
Component 22-8  1 simplex  : |S34

 qq = 21
Component 21-1  1 simplex  : |S01
Component 21-2  18 simplices:|S02 S04 S08 S14 S26 S29 S33 S40
                              S41 S42 S05 S17 S21 S30 S36 S18
                              S07 S39
Component 21-3  1 simplex  : |S09
Component 21-4  1 simplex  : |S15
Component 21-5  1 simplex  : |S25
Component 21-6  1 simplex  : |S34

 qq = 20
Component 20-1  1 simplex  : |S01
Component 20-2  20 simplices:|S02 S04 S08 S14 S26 S29 S33 S40
                              S41 S42 S05 S17 S25 S21 S30 S36
                              S39 S18 S15 S07
Component 20-3  1 simplex  : |S03
Component 20-4  1 simplex  : |S09
Component 20-5  1 simplex  : |S11
Component 20-6  1 simplex  : |S20
Component 20-7  1 simplex  : |S27
Component 20-8  1 simplex  : |S34
```

```
 qq = 19
Component 19-1  1 simplex  : |S01
Component 19-2  23 simplices:|S02 S04 S08 S14 S18 S26 S29 S33 S36 S40 S41 S42 S05 S09 S15 S17 S25 S07 S21 S30
                              S39 S34 S03
Component 19-3  1 simplex  : |S11
Component 19-4  1 simplex  : |S20
Component 19-5  1 simplex  : |S24
Component 19-6  1 simplex  : |S27

 qq = 18
Component 18-1  25 simplices:|S01 S41 S02 S04 S08 S14 S18 S26 S29 S33 S36 S40 S42 S05 S09 S15 S17 S20 S25 S07
                              S21 S30 S34 S39 S03
Component 18-2  1 simplex  : |S11
Component 18-3  1 simplex  : |S24
Component 18-4  1 simplex  : |S27

 qq = 17
Component 17-1  25 simplices:|S01 S14 S41 S02 S04 S08 S18 S26 S29 S33 S36 S40 S42 S05 S15 S17 S20 S07 S21 S30
                              S39 S09 S25 S34 S03
Component 17-2  1 simplex  : |S06
Component 17-3  1 simplex  : |S11
Component 17-4  1 simplex  : |S16
Component 17-5  1 simplex  : |S24
Component 17-6  1 simplex  : |S27

 qq = 16
Component 16-1  29 simplices:|S01 S02 S14 S36 S41 S42 S04 S08 S18 S26 S29 S33 S40 S05 S09 S15 S17 S20 S25 S27
                              S07 S21 S30 S34 S39 S24 S06 S16 S03
Component 16-2  1 simplex  : |S11
Component 16-3  1 simplex  : |S22

 qq = 15
Component 15-1  30 simplices:|S01 S33 S02 S08 S09 S14 S18 S36 S41 S42 S04 S26 S29 S40 S05 S07 S17 S20 S21 S25
                              S30 S34 S39 S15 S27 S11 S24 S06 S16 S03
Component 15-2  1 simplex  : |S19
Component 15-3  1 simplex  : |S22
Component 15-4  1 simplex  : |S28

 qq = 14
Component 14-1  30 simplices:|S01 S33 S02 S08 S09 S14 S15 S18 S21 S36 S40 S41 S42 S04 S26 S29 S27 S05 S07 S11
                              S17 S20 S25 S30 S34 S39 S24 S16 S03 S06
Component 14-2  1 simplex  : |S12
Component 14-3  1 simplex  : |S19
Component 14-4  1 simplex  : |S22
Component 14-5  1 simplex  : |S28
Component 14-6  1 simplex  : |S35

 qq = 13
Component 13-1  33 simplices:|S01 S33 S02 S08 S09 S14 S15 S18 S21 S26 S30 S36 S40 S41 S42 S04 S29 S27 S05 S07
                              S11 S17 S20 S25 S34 S39 S22 S24 S16 S19 S03 S06 S28
Component 13-2  1 simplex  : |S12
Component 13-3  1 simplex  : |S32
Component 13-4  1 simplex  : |S35
Component 13-5  1 simplex  : |S37

 qq = 12
Component 12-1  37 simplices:|S01 S33 S02 S04 S05 S08 S09 S14 S15 S18 S21 S25 S26 S29 S30 S34 S36 S39 S40 S41
                              S42 S03 S27 S07 S11 S16 S17 S20 S12 S22 S24 S06 S19 S28 S37 S32 S35
Component 12-2  1 simplex  : |S13
Component 12-3  1 simplex  : |S31

 qq = 11
Component 11-1  39 simplices:|S01 S33 S02 S04 S05 S06 S07 S08 S09 S11 S14 S15 S18 S20 S21 S25 S26 S29 S30 S34
                              S36 S39 S40 S41 S42 S03 S22 S27 S16 S17 S12 S24 S37 S19 S28 S31 S35 S32 S13
Component 11-2  1 simplex  : |S38

 qq = 10
Component 10-1  40 simplices:|S01 S33 S02 S04 S05 S06 S07 S08 S09 S11 S14 S15 S17 S18 S19 S20 S21 S24 S25 S26
                              S29 S30 S34 S36 S39 S40 S41 S42 S03 S12 S22 S35 S27 S16 S37 S28 S31 S32 S13 S38

 qq = 9
Component 9-1  40 simplices:|S01 S27 S33 S02 S04 S05 S06 S07 S08 S09 S11 S14 S15 S16 S17 S18 S19 S20 S21 S24
                                S25 S26 S28 S29 S30 S34 S36 S39 S40 S41 S42 S03 S12 S22 S35 S31 S37 S13 S38
                                S32
Component 9-2  1 simplex  : |S10

 qq = 8
Component 8-1  41 simplex  :|S01 S22 S27 S33 S02 S04 S05 S06 S07 S08 S09 S11 S14 S15 S16 S17 S18 S19 S20 S21
                             S24 S25 S26 S28 S29 S30 S34 S36 S37 S39 S40 S41 S42 S03 S12 S13 S35 S31 S38 S32
                             S10
qq = 7 to 0
Component 7-1  42 simplices: |S01 S22 S27 S33 S02 S04 S05 S06 S07 S08 S09 S11 S12 S14 S15 S16 S17 S18 S19 S20
                              S21 S24 S25 S26 S28 S29 S30 S31 S34 S36 S37 S39 S40 S41 S42 S03 S13 S10 S35 S38
                              S23 S32
```

Number of Students	Item liked	Number of Students	Item liked
37	reading_science	16	email
33	talking_science	16	history
33	physics	16	literature
33	mathematics	15	pets
30	sunshine	15	read_novels
29	travelling	15	statistics
28	listen_music	14	cooking
28	democracy	14	deaming-daydreaming
27	complexity_science	13	teamwork
26	holidays	13	rain
26	be_with_family	12	sociology
26	writing_science	12	give_presentations
26	walk-hike	12	meditation
25	chat_friends	11	wind
25	children	11	football
25	sleeping	11	organise_events
24	sport	10	chemistry
23	art	10	geography
22	parties	9	gardening
21	psychology	9	computer_games
21	play_music-sing	8	political_science
20	wine	7	facebook-networking
19	snow	6	economics
19	biology	4	watch_TV
18	food	4	smoking
17	swimming	2	bankers
17	philosophy		

Table 3. Number of items liked by students

Table 3 shows the numbers of items liked by two or more students. Since the largest of these is 37 students reading science, five students do not like reading science, and so no vertex is related to all the students and the components from q = 7 to q = 0 are hub-free.

This hypernetwork is characterised by a dominant component that emerges at q = 30 with two students S02 and S26 sharing 31 items. This component grows steadily as q decreases, as shown in Figure 4.

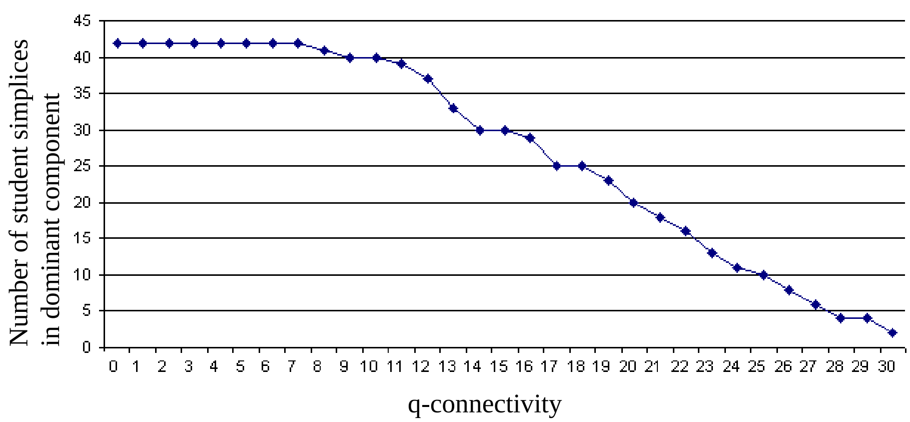


Figure 4. The emergence of the dominant component in the students' likes structure.

Table 4 shows the dimensions of the student simplices, the dimension of their largest shared face, q-shared, and their *eccentrity* which is defined to be (dimension – q-shared)/(dimension+1).

Student	Dimension	q-shared	eccentricty
S41	40	30	0.24
S02	40	31	0.22
S33	37	30	0.18
S26	35	31	0.11
S14	34	28	0.17
S08	33	27	0.18
S40	32	28	0.12
S21	31	26	0.16
S05	31	27	0.13
S42	30	26	0.13
S29	28	24	0.14
S04	27	25	0.07
S17	26	23	0.11
S36	26	23	0.11
S01	25	19	0.23
S15	25	21	0.15
S30	25	22	0.12
S39	25	23	0.08
S18	25	24	0.04
S07	24	22	0.08
S09	23	20	0.13
S34	23	20	0.13
S25	23	21	0.08
S11	21	16	0.23
S27	21	17	0.18
S20	21	19	0.09
S03	21	20	0.05
S24	20	17	0.14
S06	18	17	0.05
S16	18	17	0.05
S22	17	14	0.17
S19	16	14	0.12
S28	16	14	0.12
S12	15	13	0.13
S35	15	13	0.13
S32	14	13	0.07
S37	14	13	0.07
S13	13	12	0.07
S31	13	12	0.07
S38	12	11	0.08
S10	10	9	0.09
S23	8	8	0

(a) sorted by dimension

Student	Dimension	q-shared	eccentricty
S41	40	30	0.24
S01	25	19	0.23
S11	21	16	0.23
S02	40	31	0.22
S33	37	30	0.18
S08	33	27	0.18
S27	21	17	0.18
S14	34	28	0.17
S22	17	14	0.17
S21	31	26	0.16
S15	25	21	0.15
S29	28	24	0.14
S24	20	17	0.14
S05	31	27	0.13
S42	30	26	0.13
S09	23	20	0.13
S34	23	20	0.13
S12	15	13	0.13
S35	15	13	0.13
S40	32	28	0.12
S30	25	22	0.12
S19	16	14	0.12
S28	16	14	0.12
S26	35	31	0.11
S17	26	23	0.11
S36	26	23	0.11
S20	21	19	0.09
S10	10	9	0.09
S39	25	23	0.08
S07	24	22	0.08
S25	23	21	0.08
S38	12	11	0.08
S04	27	25	0.07
S32	14	13	0.07
S37	14	13	0.07
S13	13	12	0.07
S31	13	12	0.07
S03	21	20	0.05
S06	18	17	0.05
S16	18	17	0.05
S18	25	24	0.04
S23	8	8	0

(b) sorted by eccentricity

Table 4. Student dimensions, largest value of q-shared and eccentricities

Students S41 and S02 have the highest dimension of 40, meaning they like 41 of the 54 listed things. S41 is 30-near S33 and S02 is 31-near S26, meaning that they share 31 and 32 liked items respectively. The Q-analysis shows that these students form a q-connected component at q = 29.

The pattern of the Q-analysis is other students joining this component which gets larger until it includes all students, as shown in Figure 4.

The largest eccentricity for the students is 0.24, which is relatively low. It means that every student shares a lot of the things they like with other students. This was to be expected since this group of students is relatively homogeneous, with many having very good skills in physics and mathematics, all being interested in science, and most being young and in their twenties.

What else might have been observed in the Q-analysis student likes data?

Q-connected components can be viewed a *clusters* of similar things. One possible outcome from the Q-analysis could have been the emergence of distinct clusters of students. Such clusters depend on the selection of items which in this case was rather arbitrary and included things that most people like.

The scores from the students were interpreted as 0 and 1 meaning 'dislike' and 5, 6 and 7 mean 'like'. It is noteable that the 15 students with the lowest dimension all used a scale of 0 to 6 while some of those with highest dimensions used a scale of 0 to 7. This suggests that the ambiguity in the scoring has introduced some bias. A better interpretation could have been 5 and 6 meaning 'like' for those students who did not use 7, and 6 and 7 meaning 'like' for those students who did use 7. In fact when this was implemented it made little difference to the overall structure.

[3] Filtering Out Low Information Items

As Table 3 shows, many items are liked by the majority of the students. For example 42 students 37 like reading science, 33 like talking about science, 33 like physics and 33 like mathematics. This is not very surprising for a PhD school on mathematics and complexity! Nor is it surprising that 30 students like sunshine and 29 like travelling.

As an experiment these low information items were filtered out from the data ,with only those liked by 21 or less students included in the Q-analysis. It would have been possible for the students to cluster differently but the pattern of high connectivity again emerged. The following Galois pairs were detected in the components of this Q-analysis.

qq = 18
<S02 S33> < play_music-sing, snow, computer_games, biology, deaming-daydreaming, food, cooking, swimming, email, pets, teamwork, chemistry, psychology, sociology, political_science, economics, give_presentations, philosophy, meditation>
qq = 14
<S02 S33 S41> <play_music-sing, snow, biology, food, cooking, swimming, email, pets, teamwork, chemistry, psychology, sociology, philosophy>
qq = 12
<S02 S33 S41 S08> <play_music-sing, biology, food, cooking, swimming, pets, psychology, philosophy>
qq = 11
<S02 S08 S14 S26 S33 S41 S42> <food 7, psychology 7 >

Again it seems that students S02 and S33 are dominating the structure. As a final experiment these students were removed from the analysis to see if the other students would form separate clusters in their absence. The result was that that the students still formed a single cluster, reinforcing the observation that this group of students is a relatively homogeneous social group.

[4] Discussion

Although it is concluded that the student group is relatively homogeneous, does this mean that there is no structural differentiation between the students and no obvious clustering?

It should be realised that any combination of descriptors will identify a subset of students. For example, 23 students are related to the simplex < writing_science, talking_science>.

The structure analysed depends on the questions asked. For example, the students could have been asked which country they come from or what is their preferred language. At this school the simplices σ(Greek), σ(Russian), σ(German) defined significant clusters of students, and without doubt the amount of student interaction depends on this simple structure. On the other hand the simplices σ(male) and σ(female) partition the students into two groups. Is there evidence that this substructure is relevant to people's interactions?

One of the values of Q-analysis is that one can remain very close to the data. When two people are q-near, the reason can be examined immediately by inspection of the vertices they share. Sometimes the q-nearness will be exaggerated by the inclusion of irrelevant vertices, or noise, and sometimes the q-nearness will be under-estimated by the omission of relevant vertices.

[5] Conclusions

This paper illustrates the use of hypernetworks and Q-analysis to investigate the structure of a social group for the case of students at the Patras PhD School. The questionnaire used to collect the date had typical imperfections and the data collected were messy and imperfect. Despite this the analysis suggests that the students at the school are highly connected in terms of the things they like, and this is consistent with them being selected to attend the school.