Étoile

An Introduction to Hypernetworks

Lesson 2. Simplicial Complexes and Q-analysis

An Étoile Course in association with the

4th Ph.D. Summer School - Conference on

Mathematical Modeling of Complex Systems

Cultural Foundation “Kritiki Estia”, Athens

14th - 25th July 2014

The Open University

topdrim
Topology driven methods for complex systems

NESS

SEVENTH FRAMEWORK PROGRAMME

1 From sets to simplices

Although hypergraphs provide a method of representing relationships between more than two things they are not rich enough to make some basic distinctions, *e.g.* in Fig. 1 the arches a_1 and a_2 are represented by the same set of blocks, $\{x_1, x_2, x_3\}$, but they are different structures.

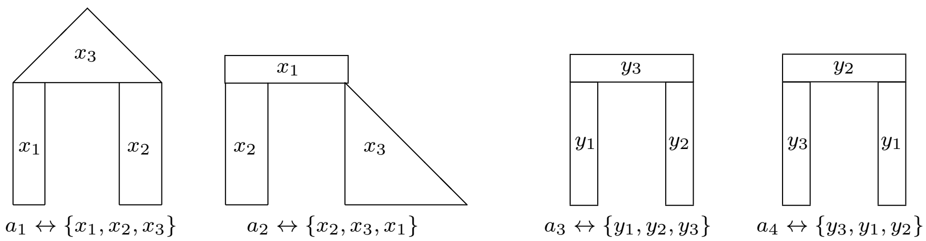


Figure 1: Limitations on the representational power of hypergraph edges

Let the rule for forming the arch be "(i) take a set of three blocks, (ii) take an element from the set and put it on the left; (iii) take another element from the set and put it on the right; (iv) take another element from the set and put it on top of the others". Selecting elements from a set is similar to pulling their elements out of a bag with your eyes closed. As far as the set is concerned, all the elements are *equivalent*, and the order in which they appear is not relevant.

Suppose one wanted the arch a_1 and not a_2. Then the elements have to be selected in the right order. Let the construction be modified as "(i) order the elements as x_1, x_2, and x_3. (ii) take element x_1 and put it on the left; (iii) take x_2 and put it on the right; (iv) take x_3 and put on top of x_1 and x_2". This gives the arch a_1 as desired. It is associated with an *ordered* set of vertices, which can be written as $\langle x_1, x_2, x_3\rangle$. This is different to $\langle x_2, x_3, x_1\rangle$ which represents a_2.

Simplices

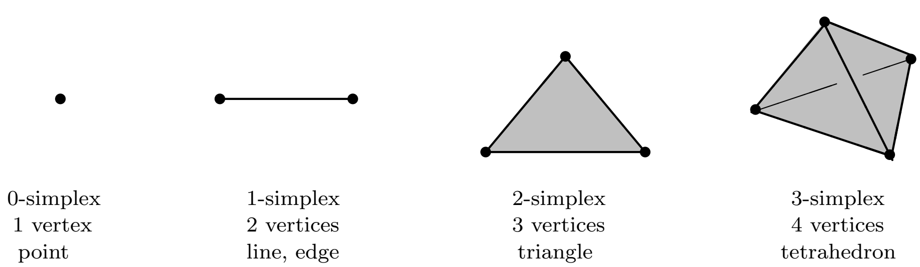
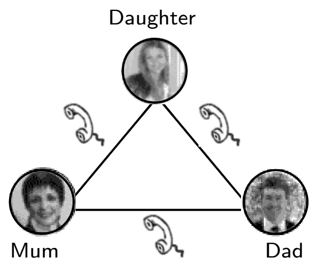


Figure 2: An n-dimensional simplex has $n + 1$ vertices

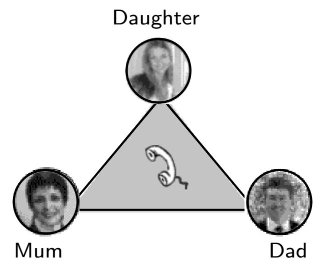
Let V be a set whose element are called *vertices*. Any subset of V, $\{v_0, v_1, ..., v_p\}$ determines an object called an *abstract p-simplex*, written $\sigma = \langle v_0, v_1, ..., v_p\rangle$. A p-simplex can be represented by a p-dimensional *polyhedron* in $(p + k)$-dimensional space, where $k \geq 0$.

Although they can be considered to be abstract objects determined by their vertices, simplices have a *geometric representation* as polyhedra in multidimensional space, *e.g.* a simplex with three vertices is a triangle in 2-dimensional space and a simplex with four vertices is a tetrahedron in 3-dimensional space. Let the notation $|\sigma|$ mean the number of vertices of a simplex σ. The *dimension* of simplex σ, $dim(\sigma)$, is defined as the number of vertices of σ minus one (Figure 2), $dim(\sigma) = |\sigma| - 1$.

Examples of simplicies



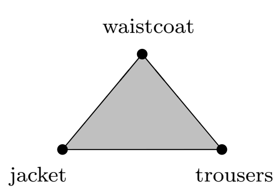
(a) three pairwise phone calls



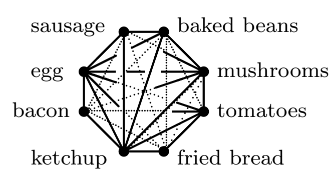
(b) a three-way phone call

Figure 3: Three pairwise phone calls $\neq$ one three-way phone call

In Fig. 3(a) an unsuspecting father gets a phone call from his daughter via the simplex ⟨Daughter, Dad⟩. “I’m in a shop and Mum said you would pay for my new dress”, to which he replies: “OK, it will be a pleasure”. Then Mum gets the message on the ⟨Daughter, Mum⟩ simplex that “Dad says he will pay for my new dress”. Then he gets a call via the simplex ⟨Mum, Dad⟩ “Are you crazy! Why didn’t you ask me first?” Poor Dad - if only there had been a three-way phone call as shown in Fig. 3(b) then none of this would have happened.



(a) The Full Monty



(b) English breakfast

Figure 4: Examples of simplices: the Full Monty

The phrase “The Full Monty” has come to mean “complete” or “‘the whole thing”. It is said to come from the Montague Burton tailoring chain which hired three-piece suits, ⟨jacket, trousers, waistcoat⟩, to men getting married. (Fig. 4(a)). Another, less likely, explanation is that it comes from the full English breakfasts General Montgomery recommended for his troops ⟨egg, bacon, sausage, fried bread, baked beans, mushrooms, tomatoes, ketchup⟩ ((Fig. 4(b)).

Vertex parts and polyhedral wholes

In his book on Gestalt psychology [Katz, 1951] rejects the equation

$$\text{Vanilla Ice Cream} = \text{Cold} + \text{Sweet} + \text{Vanilla Aroma} + \text{Softness} + \text{Yellow}$$

which suggests that each attribute can be sensed separately and put together in a linear way. In our terms, Vanilla Ice Cream is a polyhedron with five vertices bound together by an *indivisible* 5-ary relation. This can be written as

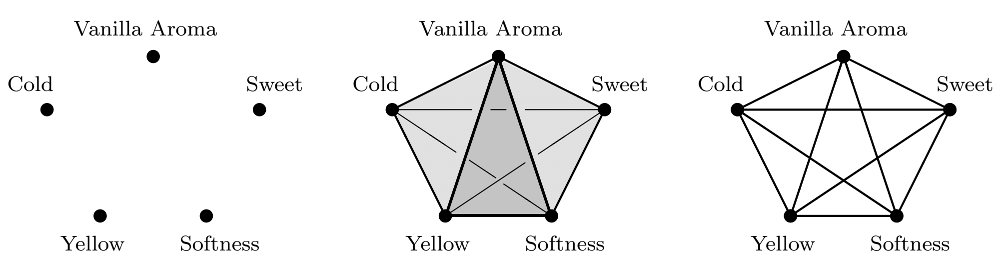
$$\begin{aligned}\text{Vanilla Ice Cream} &= \langle \text{Cold, Sweet, Vanilla, Softness, Yellow} \rangle \\ &\neq \langle \text{ Cold } \rangle + \langle \text{Sweet} \rangle + \langle \text{Vanilla} \rangle + \langle \text{ Softness} \rangle + \langle \text{Yellow} \rangle\end{aligned}$$

with the "Gestalt" construct of *Vanilla Ice Cream* represented by a polyhedron with five vertices. Figure 5 illustrates the distinction between an unrelated set of vertices and the "Gestalt" polyhedron. It also illustrates the difference between a polyhedron with five vertices embedded in a 4-dimensional space and a network-theoretic *clique* embedded in 2-dimensional space in which every vertex is connected to every other by a 1-dimensional link. The clique is the worst representation, since ice-cream is experienced as a whole, not as combinations of pairs of senses.

$$\begin{aligned}\text{Vanilla Ice Cream} &= \langle \text{Cold, Sweet, Vanilla, Softness, Yellow} \rangle & \underline{\textit{Polyhedron} \neq} \\ &\neq \{\langle \text{Cold} \rangle, \langle \text{Sweet} \rangle, \langle \text{Vanilla} \rangle, \langle \text{Soft} \rangle, \langle \text{Yellow} \rangle\} & \underline{\textit{Set of Vertices}} \\ &\neq \{\langle \text{Cold, Sweet} \rangle, \langle \text{Cold, Vanilla} \rangle, \langle \text{Cold, Soft} \rangle, & \underline{\neq \textit{Set of Lines}} \\ &\quad \langle \text{Cold, Yellow} \rangle, \langle \text{Sweet, Vanilla} \rangle, \langle \text{Sweet, Soft} \rangle, \\ &\quad \langle \text{Sweet, Yellow} \rangle, \langle \text{Vanilla, Soft} \rangle, \langle \text{Soft, Yellow} \rangle, \\ &\quad \langle \text{Vanilla, Yellow} \rangle\}\end{aligned}$$

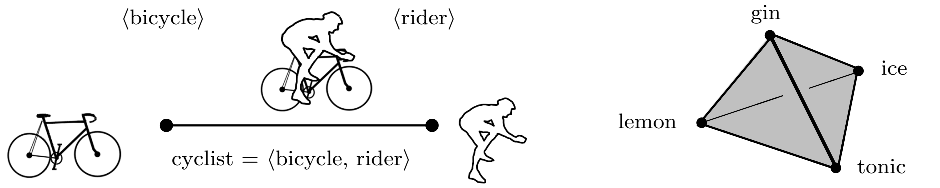
The polyhedron ⟨Cold, Sweet, Vanilla, Soft, Yellow⟩ here expresses the concept of *whole* which is clearly more than the sum of its parts:

$$\langle \text{Cold, Sweet, Vanilla, Soft, Yellow} \rangle \neq \langle \text{Cold} \rangle + \langle \text{Sweet} \rangle + \langle \text{Vanilla} \rangle + \langle \text{Soft} \rangle + \langle \text{Yellow} \rangle$$



(a) a set of vertices (b) a polyhedron (c) a clique

Figure 5: Set of vertices ≠ polyhedron ≠ clique



(a) Remove a vertex and the cyclist simplex ceases to exist

(b) Remove a vertex and the perfect gin and tonic ceases to exist

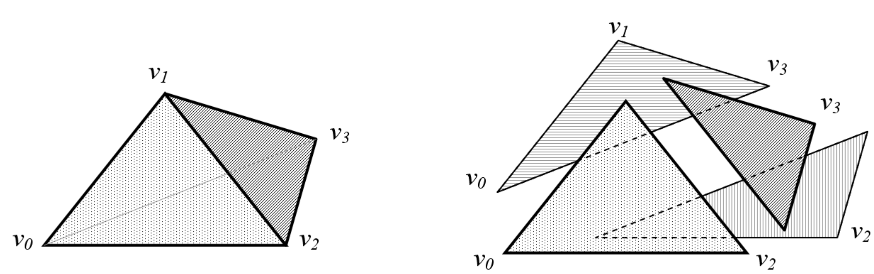
Figure 6: Remove a vertex and the simplex ceases to exist.

The essential feature of a polyhedron is that it ceases to exist if any of the vertices are removed. For example, consider a cyclist represented as the simplex ⟨rider, bicycle⟩. Remove either the man or the bicycle and what is left ceases to be a cyclist. Removing a vertex is like sticking a pin in a balloon, causing the structure to collapse and whatever is left is not the whole simplex. Remove any vertex from ⟨gin, tonic, ice, lemon⟩ and it ceases to be the perfect gin and tonic. Generalising edges to polyhedra allows a distinction to be made between the *parts* of things represented by vertices, and *wholes* represented by polyhedra.

2 Simplices, polyhedra and their faces

Connectivity is a one of the most powerful concepts for analysing complex systems as illustrated by the widespread use of networks. The vertices of networks are 0-dimensional simplices, $\langle v \rangle$ and the edges are 1-dimensional simplices, $\langle v, v' \rangle$. Two edges are "connected" if they share a vertex, and paths can be defined as chains of connected edges.

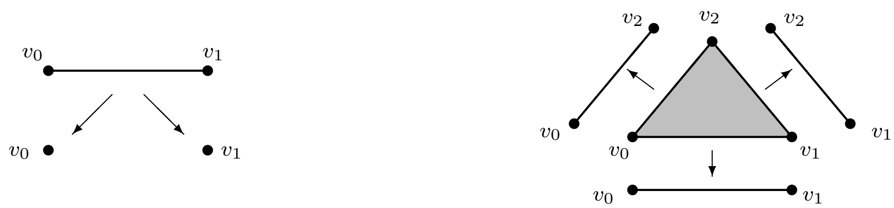
Simplices allow a natural multidimensional generalisation of this well-established concept of connectivity. For example, Figure 7 shows the four faces of a tetrahedron (3-simplex). This common use of the term "face" generalises. The 2-dimensional faces of a 3-dimensional tetrahedron are 2-dimensional triangles, the 1-dimensional faces of a 2-dimensional triangle are its 1-dimensional edges,



(a) a solid 3-dimensional tetrahedron

(b) the 2-dimensional faces of the tetrahedron

Figure 7: The 2-dimensional triangular faces of a 3-dimensional tetrahedron



(a) vertices are 0-dimensional faces of 1-dimensional edges

(a) edges are 1-dimensional faces of 2-dimensional triangles

Figure 8: The faces of edges and triangles

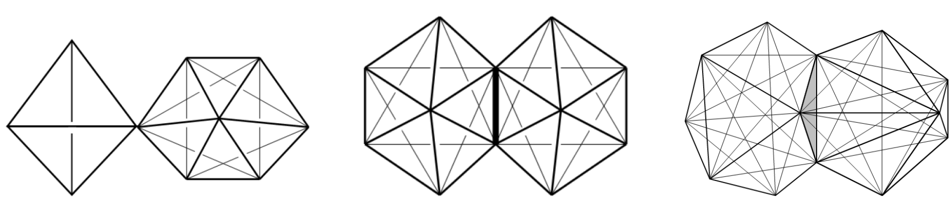
and the the 0-dimensional faces of a 1-dimensional edge are its 0-dimensional vertices (Fig. 8).

The simplex $\sigma = \langle v_0', v_1', ..., v_q' \rangle$ is defined to be a *q-dimensional face* of the simplex $\sigma' = \langle v_0, v_1, ..., v_p \rangle$ if $\{v_0', v_1', ..., v_q'\} \subseteq \{v_0, v_1, ..., v_p\}$. This is written as $\sigma \lesssim \sigma'$. For example, $\sigma = \langle v_0, v_2, v_3 \rangle$ is a 2-dimensional triangular face of the 3-dimensional tetrahedron $\sigma' = \langle v_0, v_1, v_2, v_3 \rangle$.

3 The intersection of simplices

In networks, links and arrows are connected by vertices. For multidimensional polyhedra, connectivities can have higher dimension than than the zero-dimensions of a vertex. Two simplices are *q-near* if they share a q-dimensional face. The *intersection* of two simplices σ and σ' is defined to be their highest dimensional shared face, σ''. We write $\sigma \cap \sigma' = \sigma''$.

In Fig. 9(a) the simplices share a vertex, which is a 0-dimensional face so they are 0-near. In Fig. 9(b) the simplices share an edge, which is a 1-dimensional face so they are 1-near. In Fig. 9(c) the simplices share a triangle, which is a 2-dimensional face so they are 2-near.



(a) 0-near (b) 1-near (c) 2-near

Figure 9: q-near simplices

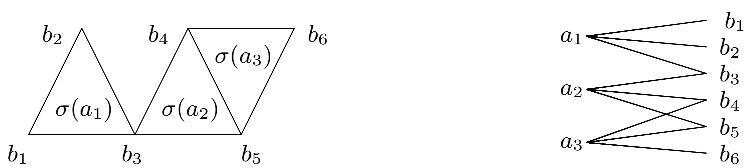
4 Simplicial families and simplicial complexes

Any set of simplices is defined to be a *simplicial family.*

A set of simplices with all their faces forms a *simplicial complex*, *i.e.* a set of simplices F is defined to be a simplicial complex if $\sigma \in K$ implies $\sigma' \in K$ for all $\sigma' \lesssim \sigma$.

Every simplicial family determines a simplicial complex, namely the simplices with all their faces.

Simplicial systems and bipartite relations



(a) The family $\{\sigma(a_1), \sigma(a_2), \sigma(a_3)\}$ (b) The bipartite relation $A \overset{R}{\longleftrightarrow} B$

Figure 10: Simplicial families and bipartite relations

Let F be a simplicial family with simplices A and vertices B. Then a bipartite relation can be defined between A and B with $a\,R\,b$ if b is a vertex of a. This is illustrated in Figure 10.

Alternatively, every bipartite relation $A \overset{R}{\longleftrightarrow} B$ defines two simplicial families. For each a in A let $\sigma(a)$ be the simplex with vertex set $\{b \,|\, a\ R\ b\}$ and for each b in B let $\sigma(b)$ be the simplex with vertex set $\{a \,|\, a\ R\ b\}$.

The *conjugate families* of $A \overset{R}{\longleftrightarrow} B$ are

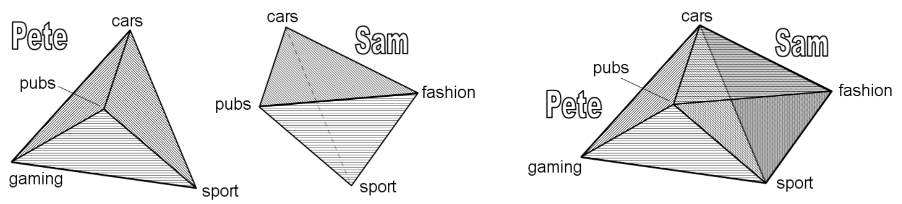
$$F_A(B,\, R\ ) = \{\sigma(a) \,|\, \text{for all } a \in A\} \text{ and}$$
$$F_B(A,\, R\ ) = \{\sigma(b) \,|\, \text{for all } b \in B\}.$$

Let $K_A(B,\, R) = \{\sigma \,|\, \sigma \lesssim \sigma(a) \text{ for all } a \in A\}$ be the simplices in $F_A(B,\, R) = \{\sigma(a) \,|\, \text{for all } a \in A\}$ together with all their faces, and let $K_B(A,\, R) = \{\sigma \,|\, \sigma \lesssim \sigma(b) \text{ for all } b \in B\}$ be the simplices in $F_B(A,\, R) = \{\sigma(b) | \text{ for all } b \in B\}$ with all their faces.

The *conjugate simplicial complexes* of $A \overset{R}{\longleftrightarrow} B$ are

$$K_A(B,\, R\ ) = \{\sigma \,|\, \sigma \lesssim \sigma(a) \text{ for any } a \in A\} \text{ and}$$
$$K_B(A,\, R\ ) = \{\sigma \,|\, \sigma \lesssim \sigma(b) \text{ for any } b \in B\}.$$

5 Multidimensional connectivity



(a) Pete and Sam as tetrahedra (b) Pete and Sam share a triangular face

Figure 11: People connected through their interests

Figure 11(a) shows two simplices representing the interests of two friends. Pete's tetrahedron (3-simplex) is $\sigma(\text{Pete}) = \langle$gaming, pubs, sport, cars$\rangle$ while Sam's simplex is $\sigma(\text{Sam}) = \langle$pubs, sport, cars, fashion$\rangle$. These friends share the triangular face $\langle$pubs, cars, sport$\rangle$ and are 2-near. Imagine them in a pub. Pete tells Sam about his successful poker game last night. Sam listens politely, before telling Pete about a new style of shoes in a magazine. Not interested in fashion, Pete might mention the car driven by his favourite soccer star, sparking Sam's interest in both cars and sport and lead to a more intense discussion.

In Fig. 12 Sue has the simplex $\langle$fashion, history, painting, literature$\rangle$. She shares just the vertex $\langle$fashion$\rangle$ with Sam, but has more in common with Jane, being 1-near through the face $\langle$history, literature$\rangle$.

The set of connected simplices in Figure 12 is a structure that supports different kinds of interaction. Whereas Pete and Sam can enjoy conversations in pubs about fast cars and their favourite team, Sue and Jane are more likely to have conversations combining history and literature such as the accuracy of Shakespeare's historical plays. In contrast Jane's conversations with Tim are likely to combine gardening with cooking, possibly discussing the seasonable implications of herbs and vegetables for the dishes they like to make.

In this micro-society, Pete and Sam are the closest sharing three interests. They form a relatively disconnected substructure from the rest, and they can

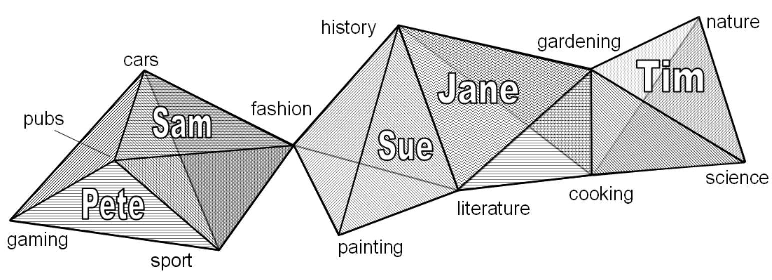


Figure 12: A simplicial family of people and their interests, $F_{\text{People}}(\text{Interests})$

be imagined chatting easily at a party. Tim is also rather peripheral, being connected only to Jane. In comparison, Sue and Jane are the most integrated, each being connected to two other people. They seem to be the most central people in this system.

The simplices σ and σ' are defined to be *q-connected* in a simplicial family F if there is a chain of simplices $\sigma_1, \sigma_2, ..., \sigma_\ell$ with $\sigma = \sigma_1$, $\sigma' = \sigma_\ell$, and σ_i being at least q-near σ_{i+1} for $i = 1, ..., \ell - 1$. $\sigma_1, \sigma_2, ..., \sigma_\ell$ is called a *chain of connection* between σ and σ'. The simplices σ and σ' are said to be *q-connected*. By this definition, if σ and σ' are q-connected then they are p-connected for all $p \leq q$.

Simplicial families and complexes extend the idea of connectivity in networks to higher dimensions. For example, Pete is 2-near Sam, Sam is 0-near Sue, Sue is 1-near Jane, Jane is 1-near Tim, so Sue is 1-connected to Tim through Jane:

$$\sigma(\text{Pete}) \overset{\text{2-near}}{\longrightarrow} \sigma(\text{Sam}) \overset{\text{0-near}}{\longrightarrow} \sigma(\text{Sue}) \overset{\text{1-near}}{\longrightarrow} \sigma(\text{Jane}) \overset{\text{1-near}}{\longrightarrow} \sigma(\text{Tim})$$

Thus two simplices can be be q-connected, even though they have no vertices in common. For example, Sue is 1-near Tim, even though $\sigma(\text{Sue}) \cap \sigma(\text{Tim}) = \emptyset$. Similarly, Pete and Tim are 0-connected, even though $\sigma(\text{Pete}) \cap \sigma(\text{Tim}) = \emptyset$.

6 Q-analysis

In general, being q-connected is an equivalence relation on a set of simplices and partitions them into *q-connected components*. A listing of the components for each dimensional q-value is called a *Q-analysis*, *e.g* the Q-analysis for $F_{\text{People}}(\text{Interests})$ in Fig. 12 is

q = 3: $\{\sigma(\text{Pete})\}, \{\sigma(\text{Sam})\}, \{\sigma(\text{Sue})\}, \{\sigma(\text{Jane})\}, \{\sigma(\text{Tim})\}$

q = 2: $\{\sigma(\text{Pete}), \sigma(\text{Sam})\}, \{\sigma(\text{Sue})\}, \{\sigma(\text{Jane})\}, \{\sigma(\text{Tim})\}$

q = 1: $\{\sigma(\text{Pete}), \sigma(\text{Sam})\}, \{\sigma(\text{Sue}), \sigma(\text{Jane}), \sigma(\text{Tim})\}$

q = 0: $\{\sigma(\text{Pete}), \sigma(\text{Sam}), \sigma(\text{Sue}), \sigma(\text{Jane}), \sigma(\text{Tim})\}$

For a small system, Q-analysis can be presented as a *skyscraper diagram* as shown in Fig. 13, as suggested in [Atkin, 1977].

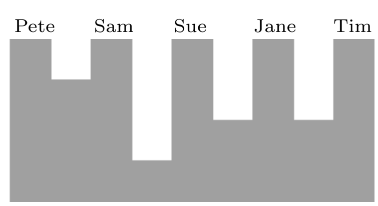


Figure 13: A Q-analysis skyscraper diagram

7 Structure Vectors

In a Q-analysis things cluster together through their shared vertices, and the pattern of components gives an insight into the connectivity of a simplicial family. The *structure vector* of a Q-analysis is a list of the number of components, Q_q at each dimension q. For example, the structure vector for the simplicial family $F_{\text{People}}(\text{Interests})$ in Fig. 12 is $(\overset{0}{1}, \overset{1}{2}, \overset{2}{4}, \overset{3}{5})$ where the dimension appears above the number of components.

For large data sets listing the number of components is impractical and it can be more useful to display the structure vectors as a graph. For example, the Observatorium project at the University of Lisbon is storing online newspapers from various countries. The web pages they are archiving have a lot of subtle structure, and there are many hundreds of thousands of them going back a year or more. As an experiment we analysed the *Australian* online newspaper articles over a period of three days. These 104 web pages used 8816 words, and there were 81,825 occurrences of these words in the 104 articles.

This structure vector illustrates a common feature in Q-analysis. At the higher dimensions there are relatively few simplices. As q decreases the number of simplices increases causing Q_q to increase, but simplices begin to become q-connected causing Q_q to decrease. Initially Q_q increases until it reaches a maximum, here denoted max-Q_q, and then decreases to Q_0, which is usually 1.

In this context we define the *q-percolation* value, P_q of the complex to be the highest value of q for which $P_q = Q_0$, *i.e.* the largest value at which all the simplices form one q-connected component when $Q_0 = 1$, or the number of disconnected components.

As shown in Fig. 14 in this case max-Q_q = 57 at $q = 221$, while P_q = 110. Thus the 104 articles form a maximum of 57 components at $q = 221$ and these all become connected at $q = 110$. Thus the percolation from maximum to minimum number of components occurs relatively rapidly between $q = 221$ and $q = 110$, which is about one sixth of the dimension range.

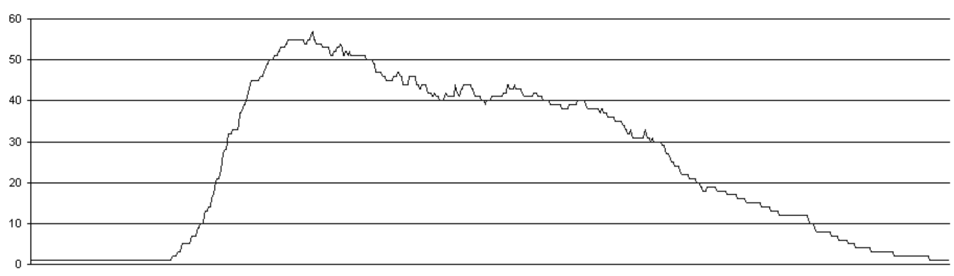


Figure 14: The structure vector for the article-word Q-analysis

8 Eccentricity

$\sigma(g_1) = \langle$ThinStem, Small, TulipShape, Narrow, Curved$\rangle$, ecc$(\sigma(g_1)|\sigma(g_3)) = 0.20$

$\sigma(g_2) = \langle$ThinStem, Tall, CupShape, Wide, Curved, Logo$\rangle$, ecc$(\sigma(g_2)|\sigma(g_3)) = 0.50$

$\sigma(g_3) = \langle$ThinStem, Tall, TulipShape, Narrow, Curved$\rangle$, ecc$(\sigma(g_3)|\sigma(g_1)) = 0.20$

$\sigma(g_4) = \langle$FatStem, Tall, VeeShape, Narrow, Straight$\rangle$, ecc$(\sigma(g_4)|\sigma(g_5)) = 0.20$

$\sigma(g_5) = \langle$FatStem, Small, VeeShape, Narrow, Straight$\rangle$, ecc$(\sigma(g_5)|\sigma(g_4)) = 0.20$

$\sigma(g_6) = \langle$ThinStem, Small, TubeShape, Narrow, Straight$\rangle$, ecc$(\sigma(g_6)|\sigma(g_1)) = 0.40$

Figure 15: A set of wine glasses, their descriptive simplices, and eccentricities

Some simplices are highly connected to other simplices while some simplices are relatively disconnected. Those simplices that do not share many of their vertices with other simplices are relatively eccentric. This is not always clear from the Q-analysis. For example, Fig. 15 shows descriptive simplices for six wine glasses.

Let $F = \{\sigma(g_1), \sigma(g_2), \sigma(g_3), \sigma(g_4), \sigma(g_5), \sigma(g_6)\}$. The Q-analysis is:

$Q = 5$: $\{\sigma(g_2)\}$

$Q = 4$: $\{\sigma(g_1)\}$ $\{\sigma(g_2)\}$ $\{\sigma(g_3)\}$ $\{\sigma(g_4)\}$ $\{\sigma(g_5)\}$ $\{\sigma(g_6)\}$

$Q = 3$: $\{\sigma(g_1), \sigma(g_3)\}$ $\{\sigma(g_2)\}$ $\{\sigma(g_4), \sigma(g_5)\}$ $\{\sigma(g_6)\}$

$Q = 2$: $\{\sigma(g_1), \sigma(g_2), \sigma(g_3), \sigma(g_4), \sigma(g_5)\}, \sigma(g_6)\}$

Let the *difference* between the simplices σ and σ', σ minus σ' written $\sigma \frown \sigma'$, be defined to be the simplex with

$$\langle x \rangle \lesssim \sigma \frown \sigma' \text{ if and only if } \langle x \rangle \lesssim \sigma \text{ and } \langle x \rangle \not\lesssim \sigma'.$$

It follows that $\sigma \frown \sigma' = \sigma \frown (\sigma \cap \sigma')$, so the difference between σ and σ' is the same as σ with the shared face removed.

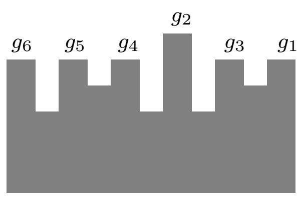
Let the eccentricity of a simplex with respect to another be:

$$\text{ecc}\,(\sigma|\sigma') \stackrel{\text{def}}{=} \frac{|\sigma \frown \sigma'|}{|\sigma|} = \frac{\text{number of } \sigma \text{ vertices not shared with } \sigma'}{\text{number of vertices of } \sigma}$$

Let the eccentricity of a simplex with respect to a family of simplices F be

$$\text{ecc}\,(\sigma|F) \stackrel{\text{def}}{=} \min\{\ \text{ecc}\,(\sigma|\sigma')|\ \ \sigma' \text{ belongs to } F\}$$

The Q-analysis of the glasses suggests that $\sigma(g_2)$ and $\sigma(g_6)$ are less integrated in F than the other simplices. As Figs. 15 and 16 show, these have the highest eccentricities (0.5 and 0.4 compared to 0.2 for the other simplices).



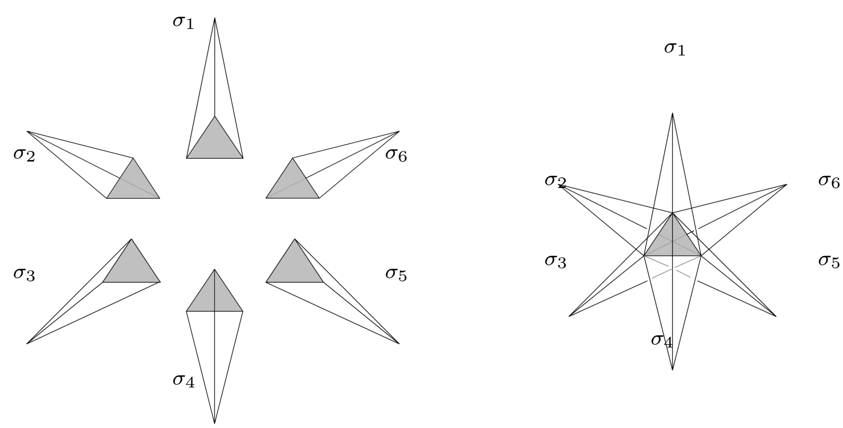
ecc(g_1) = (5-4)/5 = 0.2
ecc(g_2) = (6-3)/6 = 0.5
ecc(g_3) = (5-4)/5 = 0.2
ecc(g_4) = (5-4)/5 = 0.2
ecc(g_5) = (5-4)/5 = 0.2
ecc(g_6) = (5-3)/5 = 0.4

Figure 16: The skyscraper diagram and eccentricities for the glasses Q-analysis

Neither connectivity nor eccentricity are absolute concepts. Adding vertices or simplices to a simplicial family can change either, *e.g.* adding $\langle v_0, v_1, ..., v_n \rangle$ to the simplicial family with vertex set $\{v_0, v_1, ..., v_n\}$. "swamps" all the other simplices since they all become faces of this new simplex with eccentricity zero.

Similarly, adding vertices can change the structure, *e.g.* adding the vertex $\langle$Sherry_Glass$\rangle$ increases the dimensions of $\sigma(g_1)$, $\sigma(g_5)$, and $\sigma(g_6)$ and changes their connectivity and eccentricities. This illustrates that connectivity is sensitive to the vertices used to represent the system, and using an inappropriate vocabulary to describe a system can cause distortion.

9 Stars and hubs



(a) six 3-simplices with a common triangular face (b) the simplices in a star-hub configuration

Figure 17: A star-hub configuration

Figure 17(a) shows six 3-simplices as tetrahedra sharing a common triangular face. Figure 17(b) shows these simplices brought together into what will be called a *star-hub* configuration.

Let F be a simplicial family. *The hub of* F is defined as

$$hub\,(F) \stackrel{\text{def}}{=} \cap_{\sigma\in F}\ \sigma$$

When the hub is non-empty, $hub(F) \neq \emptyset$, F is said to be the *star of* $hub(F)$. Given a face $\langle v_0, ..., v_p\rangle$ of any simplex in F, its *star* is defined as

$$star\ \langle v_0, ..., v_p\rangle \ \stackrel{\text{def}}{=}\ \{\sigma \in F \,|\, \langle v_0, ..., v_p\rangle \lesssim \sigma\}$$

These definitions allow F to be any simplicial family. Suppose F is a subfamily of a simplicial family $\mathcal{F}$ and that $hub(F) = \langle v_0, ..., v_p\rangle$. Then it is possible that there exists a simplex σ in $\mathcal{F}$ with $\langle v_0, ..., v_p\rangle$ as a face, but σ does not belong to F. Thus in general

$$F \subseteq\ star(hub(F))$$

and for some F

$$F \subset\ star(hub(F)).$$

For example, in Fig. 17, let $\mathcal{F} = \{\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6\}$ and let $F = \{\sigma_1, \sigma_2, \sigma_3\}$. Then $hub(F)$ is the shaded triangle, but $star((hub(F))$ also includes the simplices σ_1, σ_2, and σ_3 so that

$$F\ = \{\sigma_1, \sigma_2, \sigma_3\}\ \subset\ \{\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6\}\ =\ star((\text{hub}\ (F))$$

When $star(hub(F)) = F$ the family F will be called a *maximal star*.

Let $\langle v_0, ..., v_p\rangle$ be any face of a simplex in family F. Then, by definition, $star(\langle v_0, ..., v_p\rangle) = \{\sigma \,|\, \langle v_0, ..., v_p\rangle \lesssim \sigma\}$. It is possible that $hub(\ \{\sigma \,|\, \langle v_0, ..., v_p\rangle \lesssim \sigma\})$ is "larger" than $\langle v_0, ..., v_p\rangle$.

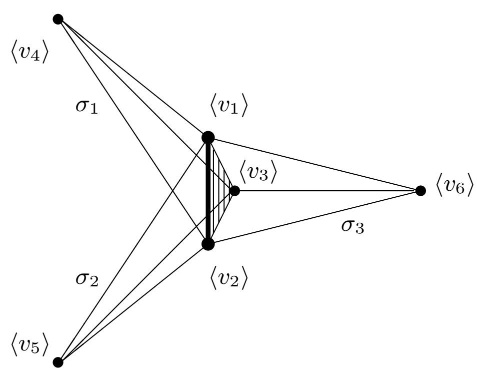


Figure 18: $\langle v_1, v_2\rangle \lesssim hub(star\langle v_1, v_2\rangle) = \langle v_1, v_2, v_3\rangle$

In Figure 18 the star of the 1-simplex $\langle v_1, v_2\rangle$ is $\{\sigma_1, \sigma_2, \sigma_3\}$. But the intersection of these simplices is the triangular face $\langle v_1, v_2, v_3\rangle$. Thus in general

$$\langle v_0, v_1, ..., v_p\rangle\ \subset\ hub(star\langle v_0, v_1, ..., v_p\rangle)$$

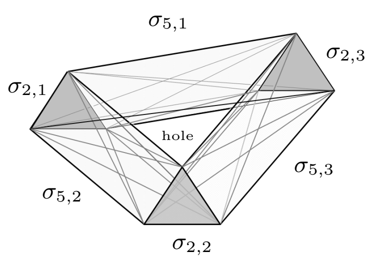
When $\langle v_0, v_1, ..., v_p\rangle\ =\ hub(star\langle v_0, v_1, ..., v_p\rangle)$, the simplex $\langle v_0, v_1, ..., v_p\rangle$ is said to be a *maximal hub*.

10 Q-graphs

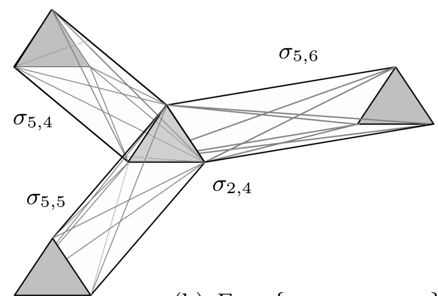
Let the *q-graph* of a simplicial family have a vertex representing each simplex and an edge with weight p between σ and σ' if they are p-near, $p \geq q$.

The simplicial families, $F_1 = \{\sigma_{5,1}, \sigma_{5,2}, \sigma_{5,3}\}$ and $F_2 = \{\sigma_{5,4}, \sigma_{5,5}, \sigma_{5,6}\}$ in Figure 19 each have three 5-dimensional simplices. The simplices of F_1 are all pairwise 2-near sharing the triangles $\sigma_{2,1} \stackrel{\text{def}}{=} \sigma_{5,1} \cap \sigma_{5,2}$, $\sigma_{2,2} \stackrel{\text{def}}{=} \sigma_{5,2} \cap \sigma_{5,3}$, and $\sigma_{2,3} \stackrel{\text{def}}{=} \sigma_{5,3} \cap \sigma_{5,1}$.

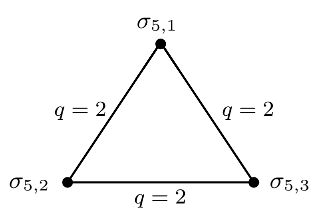
The simplices of F_2 are also pairwise 2-near sharing the triangles $\sigma_{2,4} \stackrel{\text{def}}{=} \sigma_{5,4} \cap \sigma_{5,5}$, $\sigma_{2,5} \stackrel{\text{def}}{=} \sigma_{5,5} \cap \sigma_{5,6}$, and $\sigma_{2,6} \stackrel{\text{def}}{=} \sigma_{5,6} \cap \sigma_{5,4}$. However, they are also three-wise 2-near since $\sigma_{2,4} = \sigma_{2,5} = \sigma_{2,6}$.



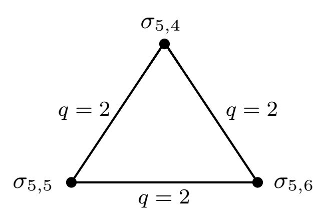
(a) $F_1 = \{\sigma_{5,1}, \sigma_{5,2}, \sigma_{5,3}\}$



(b) $F_2 = \{\sigma_{5,4}, \sigma_{5,5}, \sigma_{5,6}\}$



(c) the 2-graph of F_1



(d) the 2-graph of F_2

q = 5,4,3	$\{\sigma_1\}, \{\sigma_2\}, \{\sigma_3\}$
q = 2,1,0	$\{\sigma_1, \sigma_2, \sigma_3\}$

(e) Q-analysis of F_1

q = 5,4,3	$\{\sigma_3\}, \{\sigma_4\}, \{\sigma_5\}$
q = 2,1,0	$\{\sigma_3, \sigma_4, \sigma_5\}$

(f) Q-analysis of F_2

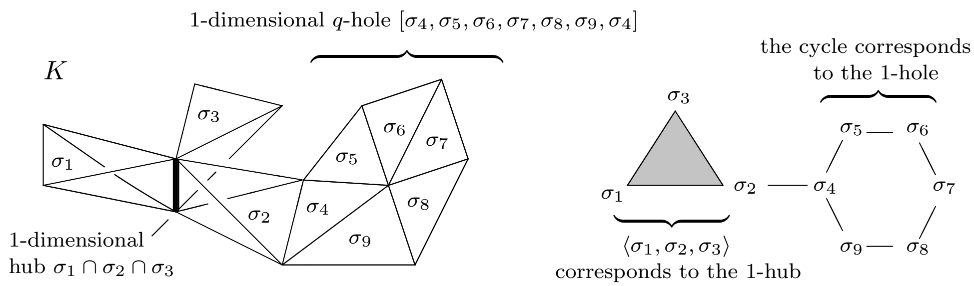
Figure 19: Q-graphs cannot discriminate different topologies

Let two q-graphs G and G' be *equivalent* if there is a bijection ϕ between their vertices such that $\langle v_1, v_2 \rangle \in G$ if and only if $\langle \phi(v_1), \phi(v_2) \rangle \in G'$.

The q-graphs of F_1 and F_2 are equivalent , *e.g.* let $\phi(\sigma_{5,1}) = \sigma_{5,4}$, $\phi(\sigma_{5,2}) = \sigma_{5,5}$, and let $\phi(\sigma_{5,3}) = \sigma_{5,6}$. Also Figs. 19(e) and (f) show that the Q-analysis of F_1 is the same as that of F_2. However, these simplicial families have different topologies because the simplices of F_1 form a configuration with a "hole" while those of F_2 are all connected by the same triangular face, $\sigma_{2,4}$. This common face acts as a *hub* of the star-like configuration.

11 From the q-graph to the q-complex

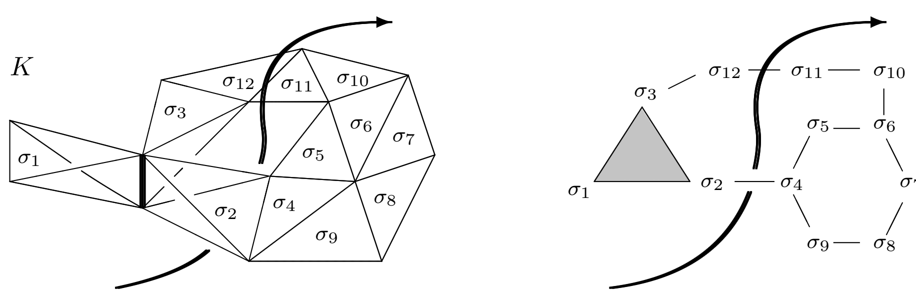
The ambiguity in the q-graph between holes and hubs in q-graphs is easy to rectify by defining the *q-complex* of a simplical family F to be the simplicial complex with simplices $\langle \sigma_1, \sigma_2, ... \rangle$ where $|\sigma_1 \cap \sigma_2 ...| \geq q$. This augments the edges of q-graphs which denote two simplices being q-near by simplices which denote that sets of simplices have a common p-dimensional face, $p \geq q$. This solves the problem of ambiguity in the q-graph.



(a) K has a 1-hub and a $1-$hole (b) the q-complex of K extends the q-graph

Figure 20: The q-complex disambiguates hubs and holes

On the left of Fig. 20(a) are three tetrahedra σ_1, σ_2 and σ_3 where $|\sigma_1 \cap \sigma_2 \cap \sigma_3| = 1$, *i.e.* these simplices have a 1-dimensional hub. This is represented by a solid triangle in the q-complex (Fig. 20(b)). Let a 1-dimensional cycle in the q-complex be defined to be a a *q-loop*, *e.g.* $\sigma_4 - \sigma_5 - \sigma_6 - \sigma_7 - \sigma_8 - \sigma_9 - \sigma_4$ in Figure 20(b). These are related to Atkin's notion of 'pseudo-homotopy', or *shomotopy*, which identifies 'q-holes' and distinguishes them from '0-holes', corresponding to the intuitive notion of 'hole'. Furthermore, a "homological" always has associated 0-loops (Fig. 21).



(a) (b)

Figure 21: Homological holes are always cycles in the q-complex

12 Galois Families

Let F be a family of simplices, $\{\sigma(a_1), \sigma(a_2), ..., \sigma(a_m)\}$ with vertices $B = \{b_1, b_2, ..., b_n\}$. Let $A = \{a_1, a_2, ..., a_m\}$. Then there is a bipartite relation R between A and B defined as $a \, R \, b$ if $\langle b \rangle \lesssim \sigma(a)$ for a in A and b in B. By an abuse of notation we will write $b \; R \; a$ if $a \; R \; b$. Let A' be any subset of A. Then

$$R(A') \;\stackrel{\text{def}}{=}\; \cap_{a \in A'} \sigma(a) \;\stackrel{\text{def}}{=}\; \sigma(A') \;\stackrel{\text{def}}{=}\; hub(A'),$$

and

$$R^2(A') \;\stackrel{\text{def}}{=}\; star(hub(A')), \quad \text{where } A' \subseteq R^2(A').$$

Similarly

$$R(B') \;\stackrel{\text{def}}{=}\; \{\sigma(a) \,|\, B' \lesssim \sigma(a)\} \;\stackrel{\text{def}}{=}\; star(B').$$

and

$$R^2(b') \;\stackrel{\text{def}}{=}\; hub(star(B')) \quad \text{where } B' \subseteq R^2(B').$$

The following hold:

For all $A' \subseteq A$, $R^2(A')$ is a maxmal star.

For all $B' \lesssim \sigma(a)$ for any a in A, $R^2(B')$ is a maximal hub.

The maximal stars $R^2(A')$ and maximal hubs $R^2(B')$ are in 1-1 correspondence.

The 1-1 correspondence is $R^2(A') \leftrightarrow R(A')$ or, equivalently, $R(B') \leftrightarrow R^2(B')$ is a *Galois connection* and $R^2(A') \leftrightarrow R(A')$ and $R(B') \leftrightarrow R^(B')$ are called *star-hub Galois pairs*. The animal-characteristics relation in Fig. 22 has the Galois pair

$\langle$brown, vegetarian, quadruped$\rangle \longleftrightarrow \langle$deer, hare, mouse, camel$\rangle$.

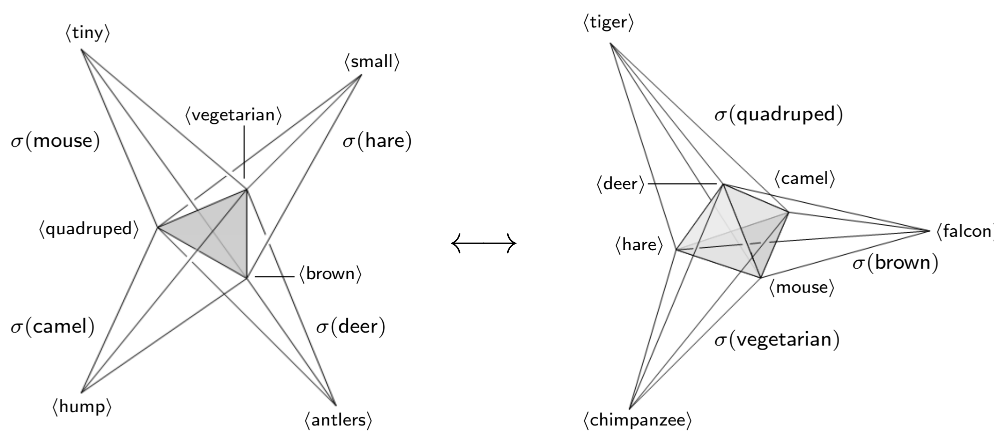
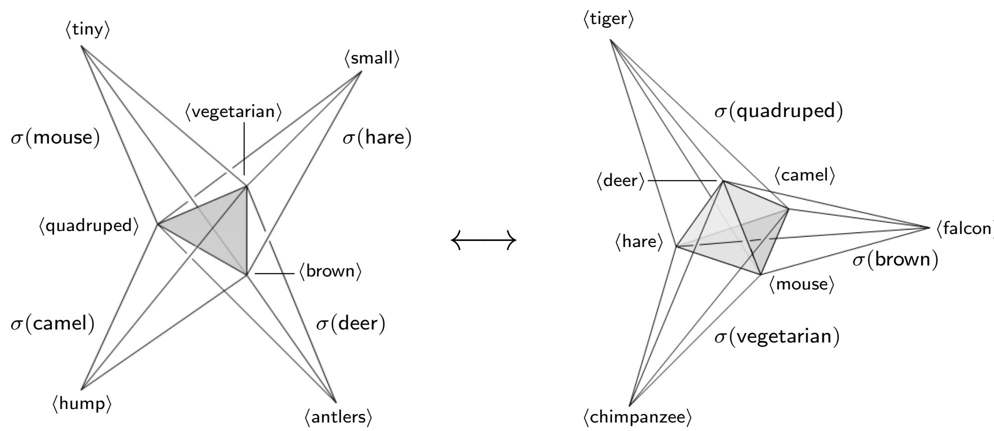
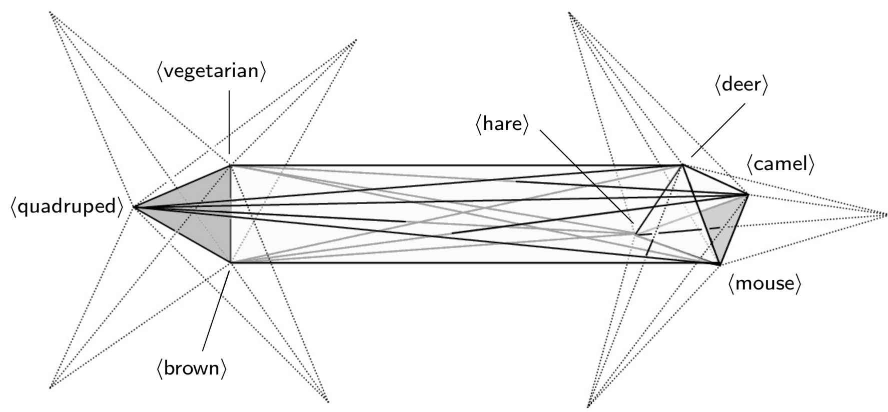


Figure 22: The Galois pair $\langle$brown, vegetarian, quadruped$\rangle \longleftrightarrow \langle$deer, hare, mouse, camel$\rangle$.

13 Galois prisms



(a) Dual star-hub pairs



(b) $hub(\{\text{hare, deer, camel, mouse}\}) \diamond hub\{\text{vegetarian, quadruped, brown}\}$
$= \langle$vegetarian, quadruped, brown, hare, deer, camel, mouse$\rangle$

Figure 23: The Galois prism formed from the hubs of dual stars

The *prism* between σ and σ', written $\sigma \diamond \sigma'$, is defined to be the simplex with the property that $\langle x \rangle \lesssim \sigma \diamond \sigma'$ if and only if $\langle x \rangle \lesssim \sigma$ or $\langle x \rangle \lesssim \sigma'$ or both. The *Galois prism* of a Galois pair $\sigma \leftrightarrow \sigma'$ is defined to be their prism, $\sigma \diamond \sigma'$.

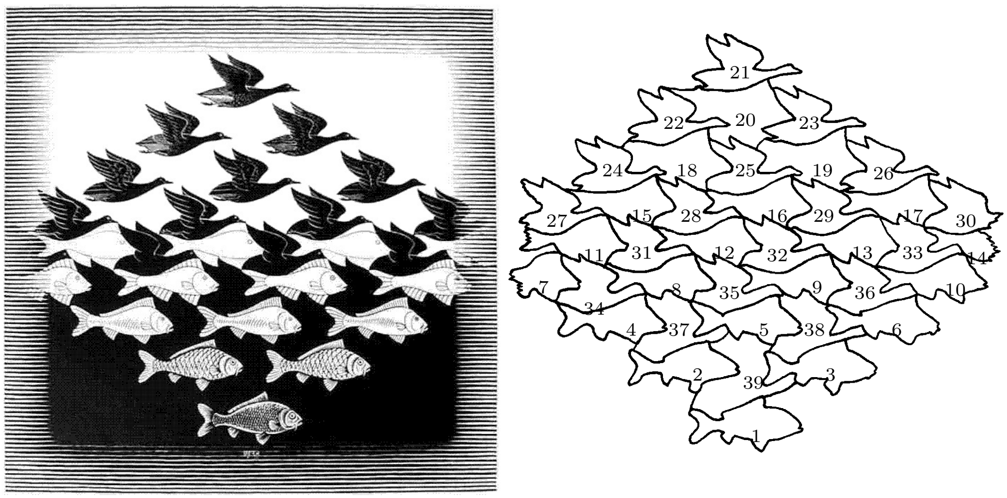
Figure 23(a) shows the star-hub pairs associated with the Galois pair

⟨hare, deer, camel, mouse⟩ $\leftrightarrow$ ⟨vegetarian, quadruped, brown, hare, deer, camel, mouse⟩.

and Figure 23(b) shows the Galois prism

⟨ hare, deer, camel, mouse ; vegetarian, quadruped, brown, hare, deer, camel, mouse⟩.

14 Example: Sky and Water



(a) Escher's *Sky and Water* (b) fish and bird shapes

Figure 24: The shapes and features abstracted from Escher's *Sky and Water*

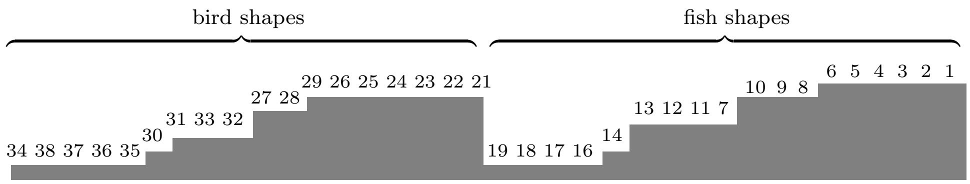
Figure 24(a) shows Escher's picture *Sky and Water* in which the birds at the top of the picture seems to change into fish at the bottom. Figure 24(b) shows the various shapes that appear in the picture. The table below shows a relation between the shapes and a set of twelve descriptors.

Shapes:	1	2	3	4	5	6	8	9	10	11	12	13	7	21	22	23	24	25	26	28	29	27	31	32	33	30	34	35	36	37	38	14	15	16	17	18	19	20	39
scales	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mouth	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
gills	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fish-tail	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
fins	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fish-shape	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0
eye	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
duck-shape	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0
two-wings	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
feathers	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
beak	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
legs	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

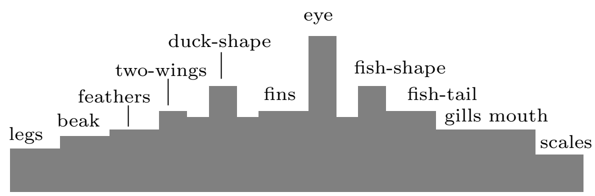
Table 4.7 The relation between descriptors and shapes for Escher's *Sky and Water*

Inspection of the incidence matrix in Table 4.7 reveals a number of maximal rectangles corresponding to star-hub Galois pairs, including:

$\langle 1, 2, 3, 4, 5, 6 \rangle \quad \longleftrightarrow \quad$ ⟨scales, mouth, gills, fish-tails, fins, fish-shape, eye⟩
$\langle 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13 \rangle \quad \longleftrightarrow \quad$ ⟨fish-tails, fins, fish-shape, eye⟩
$\langle 21, 22, 23, 24, 25, 26, 28, 29 \rangle \quad \longleftrightarrow \quad$ ⟨eye, duck-shape, two-wings, feathers, beak, legs⟩
$\langle 21, 22, 23, 24, 25, 26, 28, 29, 27, 31, 32, 33 \rangle \quad \longleftrightarrow \quad$ ⟨eye, duck-shape, two-wings⟩



(a) Skyscraper diagram for the Q-analysis of F_{Shapes}(Descriptors).



(b) Skyscraper diagram for the Q-analysis of $F_{\text{Descriptors}}$(Scales).

Figure 25: The skyscraper diagrams for the shapes – features Q-analyses

Of course there are many more Galois pairs than this, *e.g.*

$$\langle 1,2,3,4,5,6,8,9,10\rangle \quad\longleftrightarrow\quad \langle\text{mouth, gills, fish-tails, fins, fish-shape, eye}\rangle$$
$$\langle 21,22,23,24,25,26,28,29,27\rangle \quad\longleftrightarrow\quad \langle\text{eye, duck-shape, two-wings, feathers, beak}\rangle$$

Some of the columns of the incidence matrix have been swapped to make the maximal rectangles more obvious. Even so there are other Galois pairs not forming maximal rectangles in this version, for example

$$\langle 1,2,3,4,5,6,8,8,10,7\rangle \longleftrightarrow \langle\text{mouth, gills, fins, eye}\rangle$$

Figure 25(a) shows the skyscraper diagram for the Q-analysis of the shape-descriptor family. As can be seen, the shapes fall into two major components corresponding to bird shapes and the fish shapes. Figure 25(b) shows the congugate Q-analysis with σ(eye) having the largest dimension ($q = 24$) followed by σ(duck-shape) and σ(fish-shape) at $q = 16$.

Removing the "eye" descriptor creates two disconnected subfamilies, one with fish shapes and the other with duck shapes. Thus the transition from ducks at the top of Escher's picture to the fish at the bottom does not involve morphing from one shape to the other. Instead the picture is tiled by shapes, half of which get more duck-like towards the top and half of which get more fish-like towards the bottom.

14.1 Example: The Wisdom of Crowds

Groups of people often collectively give reliable answers to questions, even when some are uncertain. To investigate this, consider a mathematics test given to forty five students $\{s_1, s_2, ..., s_{45}\}$. Each member q_j of the set of questions, $\{q_1, q_2, ..., q_{20}\}$, had seven possible answers denoted $\mathtt{A}_j$, $\mathtt{B}_j$ $\mathtt{C}_j$, $\mathtt{D}_j$, $\mathtt{E}_j$, $\mathtt{F}_j$ and $\mathtt{G}_j$ in Table 4.10.

	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9	q_{10}	q_{11}	q_{12}	q_{13}	q_{14}	q_{15}	q_{16}	q_{17}	q_{18}	q_{19}	q_{20}
s_1	C_1	B_2	A_3	G_4	C_5	E_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	F_{15}	D_{16}	G_{17}	C_{18}	C_{19}	F_{20}
s_2	C_1	D_2	A_3	G_4	C_5	E_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	C_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_3	C_1	B_2	A_3	G_4	C_5	F_6	C_7	C_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	F_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_4	C_1	B_2	A_3	G_4	C_5	E_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	F_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_5	C_1	D_2	A_3	G_4	C_5	F_6	C_7	C_8	E_9	C_{10}	F_{11}	C_{12}	B_{13}	D_{14}	B_{15}	B_{16}	G_{17}	D_{18}	E_{19}	G_{20}
s_6	C_1	B_2	A_3	G_4	C_5	F_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	A_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_7	C_1	D_2	A_3	C_4	C_5	F_6	C_7	D_8	D_9	A_{10}	E_{11}	C_{12}	G_{13}	B_{14}	F_{15}	D_{16}	G_{17}	D_{18}	B_{19}	F_{20}
s_8	C_1	B_2	A_3	G_4	C_5	E_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	C_{14}	F_{15}	D_{16}	G_{17}	C_{18}	C_{19}	F_{20}
s_9	C_1	D_2	A_3	A_4	C_5	F_6	C_7	B_8	G_9	G_{10}	F_{11}	C_{12}	B_{13}	B_{14}	E_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_{10}	C_1	D_2	A_3	G_4	C_5	E_6	C_7	C_8	D_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	F_{15}	D_{16}	G_{17}	D_{18}	D_{19}	F_{20}
s_{11}	C_1	B_2	A_3	D_4	C_5	F_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	E_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_{12}	C_1	B_2	A_3	B_4	C_5	F_6	A_7	A_8	F_9	F_{10}	A_{11}	E_{12}	G_{13}	C_{14}	E_{15}	A_{16}	C_{17}	D_{18}	B_{19}	A_{20}
s_{13}	C_1	B_2	A_3	G_4	C_5	F_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	F_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_{14}	C_1	D_2	A_3	E_4	C_5	F_6	C_7	C_8	E_9	A_{10}	F_{11}	C_{12}	G_{13}	D_{14}	F_{15}	B_{16}	G_{17}	B_{18}	A_{19}	F_{20}
s_{15}	C_1	B_2	A_3	G_4	C_5	F_6	C_7	D_8	F_9	C_{10}	F_{11}	C_{12}	G_{13}	D_{14}	F_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_{16}	C_1	B_2	A_3	G_4	C_5	E_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	F_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_{17}	C_1	B_2	A_3	A_4	C_5	F_6	C_7	D_8	E_9	C_{10}	F_{11}	B_{12}	B_{13}	C_{14}	E_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_{18}	C_1	D_2	A_3	G_4	C_5	F_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	C_{14}	E_{15}	D_{16}	G_{17}	E_{18}	F_{19}	D_{20}
s_{19}	C_1	B_2	A_3	G_4	C_5	E_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	F_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_{20}	C_1	B_2	A_3	G_4	C_5	F_6	C_7	A_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	F_{15}	F_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_{21}	C_1	B_2	A_3	G_4	C_5	E_6	C_7	D_8	E_9	C_{10}	B_{11}	C_{12}	B_{13}	C_{14}	F_{15}	D_{16}	G_{17}	B_{18}	C_{19}	G_{20}
s_{22}	C_1	D_2	A_3	G_4	C_5	F_6	D_7	D_8	E_9	G_{10}	F_{11}	C_{12}	B_{13}	C_{14}	F_{15}	A_{16}	G_{17}	A_{18}	C_{19}	F_{20}
s_{23}	C_1	D_2	A_3	C_4	C_5	E_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	F_{15}	D_{16}	G_{17}	A_{18}	C_{19}	F_{20}
s_{24}	C_1	B_2	A_3	G_4	C_5	F_6	C_7	B_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	E_{15}	B_{16}	G_{17}	C_{18}	C_{19}	F_{20}
s_{25}	C_1	B_2	A_3	G_4	C_5	E_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	F_{15}	D_{16}	G_{17}	D_{18}	G_{19}	F_{20}
s_{26}	C_1	D_2	A_3	B_4	C_5	F_6	C_7	B_8	F_9	G_{10}	D_{11}	C_{12}	G_{13}	G_{14}	B_{15}	B_{16}	E_{17}	C_{18}	B_{19}	F_{20}
s_{27}	C_1	B_2	A_3	G_4	C_5	E_6	C_7	C_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	F_{15}	D_{16}	G_{17}	D_{18}	E_{19}	F_{20}
s_{28}	B_1	B_2	A_3	G_4	C_5	F_6	D_7	D_8	D_9	A_{10}	F_{11}	C_{12}	G_{13}	D_{14}	F_{15}	D_{16}	G_{17}	D_{18}	C_{19}	E_{20}
s_{29}	C_1	B_2	A_3	G_4	C_5	E_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	E_{15}	D_{16}	G_{17}	D_{18}	B_{19}	F_{20}
s_{30}	C_1	B_2	A_3	G_4	C_5	F_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	A_{14}	E_{15}	B_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_{31}	C_1	B_2	A_3	G_4	C_5	E_6	C_7	C_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	C_{14}	E_{15}	D_{16}	G_{17}	D_{18}	D_{19}	F_{20}
s_{32}	C_1	B_2	A_3	G_4	C_5	E_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	F_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_{33}	C_1	B_2	A_3	G_4	C_5	E_6	X_7	B_8	D_9	A_{10}	A_{11}	C_{12}	E_{13}	D_{14}	E_{15}	B_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_{34}	C_1	B_2	A_3	G_4	C_5	E_6	C_7	C_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	C_{14}	F_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_{35}	C_1	B_2	A_3	G_4	C_5	F_6	C_7	D_8	E_9	B_{10}	F_{11}	C_{12}	B_{13}	D_{14}	F_{15}	A_{16}	G_{17}	D_{18}	B_{19}	F_{20}
s_{36}	C_1	B_2	A_3	B_4	C_5	E_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	F_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_{37}	C_1	D_2	A_3	G_4	C_5	E_6	A_7	A_8	E_9	A_{10}	B_{11}	C_{12}	B_{13}	B_{14}	E_{15}	D_{16}	G_{17}	C_{18}	D_{19}	F_{20}
s_{38}	C_1	D_2	A_3	D_4	C_5	F_6	C_7	F_8	E_9	B_{10}	B_{11}	A_{12}	D_{13}	G_{14}	B_{15}	B_{16}	G_{17}	C_{18}	C_{19}	A_{20}
s_{39}	C_1	B_2	A_3	G_4	C_5	F_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	G_{15}	D_{16}	G_{17}	D_{18}	E_{19}	G_{20}
s_{40}	C_1	B_2	A_3	G_4	C_5	F_6	C_7	E_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	B_{14}	F_{15}	B_{16}	G_{17}	D_{18}	B_{19}	A_{20}
s_{41}	C_1	D_2	A_3	D_4	C_5	F_6	C_7	B_8	D_9	C_{10}	X_{11}	A_{12}	G_{13}	D_{14}	E_{15}	A_{16}	G_{17}	D_{18}	C_{19}	A_{20}
s_{42}	C_1	B_2	A_3	G_4	C_5	E_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	F_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_{43}	C_1	B_2	A_3	G_4	C_5	E_6	C_7	D_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	F_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}
s_{44}	C_1	B_2	A_3	G_4	C_5	F_6	C_7	A_8	E_9	A_{10}	F_{11}	C_{12}	B_{13}	D_{14}	E_{15}	B_{16}	D_{17}	C_{18}	C_{19}	F_{20}
s_{45}	D_1	B_2	A_3	G_4	C_5	E_6	C_7	B_8	E_9	A_{10}	F_{11}	C_{12}	G_{13}	D_{14}	F_{15}	D_{16}	G_{17}	D_{18}	C_{19}	F_{20}

Table 4.10. The relation between students and their answers

Can the correct answers be abstracted from this table with no further information? To test this the relation R is defined between the students and their answers. Student s_i is R-related to answer a_j if this is the answer they give to q_j. For example, student s_1 is R-related to $\mathtt{C}_1$, $\mathtt{B}_2$, $\mathtt{A}_3$, $\mathtt{G}_4$, and so on.

Answer	Students	Answer	Students	Answer	Students	Answer	Students
q_1 - $\mathtt{C}_1$	43	q_6 - $\mathtt{F}_6$	24	q_{11} - $\mathtt{F}_{11}$	37	q_{16} - $\mathtt{D}_{16}$	31
q_2 - $\mathtt{B}_2$	32	q_7 - $\mathtt{C}_7$	40	q_{12} - $\mathtt{C}_{12}$	41	q_{17} - $\mathtt{G}_{17}$	42
q_3 - $\mathtt{A}_3$	45	q_8 - $\mathtt{D}_8$	26	q_{13} - $\mathtt{B}_{13}$	35	q_{18} - $\mathtt{D}_{18}$	33
q_4 - $\mathtt{G}_4$	34	q_9 - $\mathtt{E}_9$	36	q_{14} - $\mathtt{D}_{14}$	30	q_{19} - $\mathtt{C}_{19}$	30
q_5 - $\mathtt{C}_5$	45	q_{10} - $\mathtt{A}_{10}$	34	q_{15} - $\mathtt{F}_{15}$	26	q_{20} - $\mathtt{F}_{20}$	36

Table 4.11. The most popular answers selected by the 45 students.

For each question the most frequently given answers are shown in Table 4.11. The first column shows that all students except two gave the answer $\mathtt{C}_1$ to question q_1, making it highly likely that this is the correct answer. In general one would expect the majority response to be correct.

At first sight the answers in Table 4.11 are correct, since in all cases more than half the students gave these responses. For most of the questions the students overwhelmingly agree, but for some the agreement is not so clear. For example, for question q_6 the answer $\mathtt{F}_6$ was selected by 24 students (53%). The answer $\mathtt{D}_8$ to questions q_8 was given by 26 students (58%), and the answer $\mathtt{F}_{15}$ to question q_{15} was also selected by 26 students (58%). How certain can one be that the most popular answers are really correct in these cases?

Of particular interest is 21 students answering $\mathtt{E}_6$ (47%) to q_6, compared to 24 (53%) for $\mathtt{F}_6$. Is the majority correct? To answer this question, consider the students viewed as relational simplices, *e.g.* $\sigma(s_1) = \langle \mathtt{C}_1, \mathtt{B}_2, \mathtt{A}_3, \mathtt{G}_4, \mathtt{C}_5, \mathtt{E}_6, \mathtt{C}_7, \mathtt{D}_8, \mathtt{E}_9, \mathtt{A}_{10}, \mathtt{F}_{11}, \mathtt{C}_{12}, \mathtt{B}_{13}, \mathtt{D}_{14}, \mathtt{F}_{15}, \mathtt{D}_{16}, \mathtt{G}_{17}, \mathtt{C}_{18}, \mathtt{C}_{19}, \mathtt{F}_{20} \rangle$. $H_S(Q;R)$ is the family of the relational simplices $\sigma(s_1), ..., \sigma(s_{45})$.

The Q-analysis of the hypernetwork $H_S(Q,R)$ in Fig. 26 shows the component $\{s_{42}, s_{16}, s_{43}, s_4, s_{32}, s_{19}\}$ at $q = 19$, meaning that each of these students gave exactly the same answers to all twenty questions, *i.e.* the simplex $\sigma = \langle \mathtt{C}_1, \mathtt{B}_2, \mathtt{A}_3, \mathtt{G}_4, \mathtt{C}_5, \mathtt{E}_6, \mathtt{C}_7, \mathtt{D}_8, \mathtt{E}_9, \mathtt{A}_{10}, \mathtt{F}_{11}, \mathtt{C}_{12}, \mathtt{B}_{13}, \mathtt{D}_{14}, \mathtt{F}_{15}, \mathtt{D}_{16}, \mathtt{G}_{17}, \mathtt{D}_{18}, \mathtt{C}_{19}, \mathtt{F}_{20} \rangle$. Its vertices are exactly the same as the list of most frequently occurring answers given in Table 4.11, with the exception of $\mathtt{E}_6$, instead of $\mathtt{F}_6$. Have these six students answered all the questions correctly with the exception of q_6?

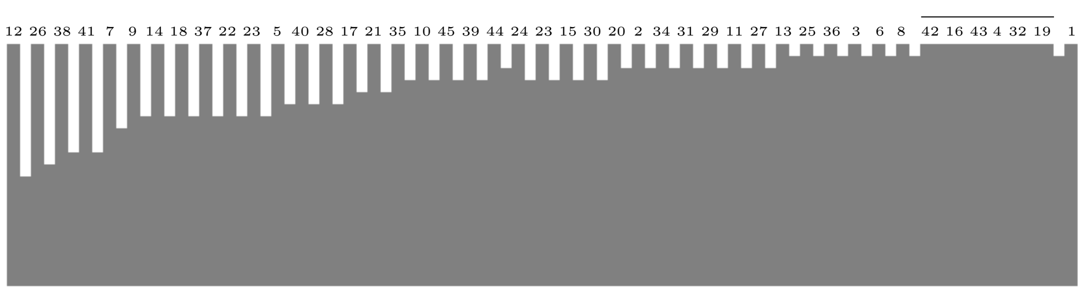


Figure 26: Q-analysis of the student-questions relation, $H_S(Q,R)$ showing connected students

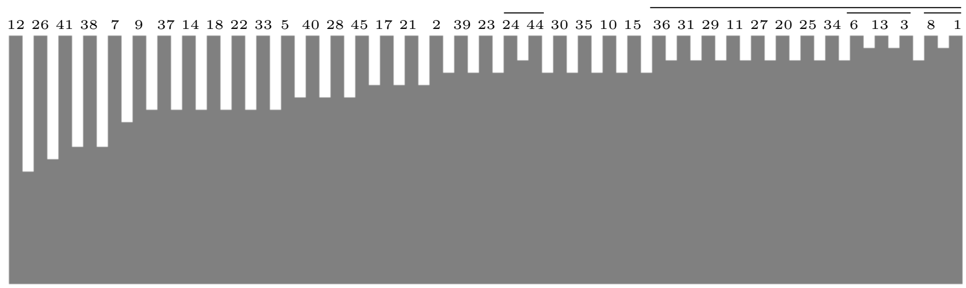


Figure 27: Q-analysis of $H_S(Q, R)$ with students s_{42}, s_{16}, s_{43}, s_4, s_{32} and s_{19} removed

In this system, if more than two students get all the answers right then they will have identical answer simplices. This is the first indication that $\mathtt{E}_6$ is correct rather than the more popular answer $\mathtt{F}_6$. Since there is only one non-trivial 19-component, all the other students have at least one vertex different to any other, which means that if $\mathtt{F}_6$ is correct only one student, s_{13}, got all the answers right.

Consider the weaker students in this cohort. They will get a number of the answers wrong, not only missing the correct answer but also giving a scattering of incorrect answers, and they will tend to be more eccentric in the answers they give. In other words, one expects the better students to be more highly connected and the weaker students to be less highly connected.

To investigate the lower level connectivities, students $s_4, s_{16}, s_{19}, s_{32}, s_{42}$, and s_{43} were removed from the system, and the Q-analysis rerun (Figure 27).

At $q = 18$ there are three components, $\{s_{24}, s_{44}\}$, $\{s_6, s_{13}, s_3\}$ and $\{s_8, s_1\}$. Of these students, s_{24}, s_{44}, s_6, s_{13}, and s_3 gave the answer $\mathtt{F}_6$ while s_8 and s_1 gave the answer $\mathtt{E}_6$. Thus, five of the most highly connected students favoured $\mathtt{F}_6$ while, including the six removed for this analysis, eight favoured $\mathtt{E}_6$ (62%).

A larger component emerges at $q = 17$, with students s_1, s_3, s_6, s_8, s_{11}, s_{13}, s_{20}, s_{25}, s_{27}, s_{29}, s_{31}, s_{34}, and s_{36}. Eight of these students favour $\mathtt{E}_6$ while five favour $\mathtt{F}_6$. Combined with the previous six, this means that 14 of the most highly connected students favour $\mathtt{E}_6$ (74%) while 5 favour $\mathtt{F}_6$.

What about the most disconnected students at the left of Fig. 27? Examination of Table 4.10 shows that s_{12}, s_{26}, s_{41}, s_{38} and s_7 all gave the answer $\mathtt{F}_6$. Assuming these are the weakest students, this is another strong indication that $\mathtt{F}_6$ is wrong. This is a strong indication that $\mathtt{E}_6$ is the correct answer to q_6.

Thus, although $\mathtt{F}_6$ is the most popular answer for q_6, the most highly connected students overwhelmingly prefer $\mathtt{E}_6$. Assuming that the most highly connected students will be the best, this is a very strong indication that $\mathtt{E}_6$ is the correct answer.

Having reached this conclusion without any information about the questions or answers other than that given in Table 4.10, the conclusion can be tested by reference to the examination paper. Question 6 reads as follows: "A body moves in such a way that its speed (in miles per hour) after t hours is $4t^3$. How far has it travelled after 3 hours?" It gives the options (A_6)16 miles, (B_6)

27 miles, (C_6)54 miles, (D_6) 64 miles, (E_6) 81 miles, (F_6) 108 miles, and (G_6) 243 miles. The stronger students correctly realised that they had to integrate $4t^3$ and substitute 3 into t^4 to give 81 miles (E_6), while the weaker student incorrectly substituted 3 directly into $4t^3$ to obtain 108 miles (F_6).

This example shows how multidimensional connectivity can be used to reason about systems. It also illustrates that the "wisdom of crowds" may be more subtle than majority decision making, and that the way individuals cluster together through their connectivity can be significant.

References

[Atkin, 1977] . Atkin, R. H., *Combinatorial Connectivities in Social Systems*, Birkhäuser (Basel), 1977.

[Johnson, 2014] Johnson, J. H., *Hypernetworks in the science of complex systems*, Imperial College Press (London), 2014.

[Katz, 1951] Katz, P., *Gestalt Pyschology*, Metuen & Co. Ltd (London), 1951.