Tangles in the Social Sciences

*A new mathematical model to identify types and predict behaviour**

Reinhard Diestel

Traditional clustering identifies groups of objects that share certain qualities. Tangles do the converse: they identify groups of qualities that often occur together. They can thereby identify and discover *types*: of behaviour, views, abilities, dispositions.

The mathematical theory of tangles has its origins in the connectivity theory of graphs, which it has transformed over the past 30 years. It has recently been axiomatized in a way that makes its two deepest results applicable to a much wider range of contexts.

This expository paper indicates some contexts where this difference of approach is particularly striking. But these are merely examples of such contexts: in principle, it can apply to much of the quantitative social sciences.

Our aim here is twofold: to indicate just enough of the theory of tangles to show how this can work in the various different contexts, and to give plenty of different examples illustrating this.

Contents

Introduction

Suppose we run a survey S of fifty political questions on a population P of a thousand people. If there exists a group of, say, a hundred like-minded people among these, there will be a 'typical' way of answering the questions in S in the way most of those people would. Quantitatively, there will exist a subset X of P, not too small, and an assignment τ of answers to all the questions in S such that, for most questions $s \in S$, some 80% (say) of the people in X agree with the answer to s given by τ. (Which 80% of X these are will depend on the choice of s.) We call this collection τ of views – answers to the questions in S – a *mindset*. Note that there may be more than one mindset for S, or none.

Traditionally, mindsets are found just intuitively: they are first guessed, and only then established by quantitative evidence from surveys designed to test them. For example, we might feel that there is a 'socialist' way σ of answering S. To support this intuition, we might then check whether any sizable subset $X \subseteq P$ as above exists for this particular $\tau := \sigma$.

Tangles can do the converse: they will identify both X and τ without us having to guess them first. For example, tangle analysis of political polls in the UK in the years well before the Brexit referendum might have established the existence of a mindset we might now, with hindsight, call the 'Labour-supporting non-socialist Brexiteer': a mindset whose existence few would have guessed intuitively when Brexit was not yet on the agenda. And similarly in the US with the MAGA[1] mindset before 2016, or that of a 'conservative Green' in the early 1970s. Tangles can identify previously unknown patterns of coherent views or behaviour.

There are two main mathematical theorems about tangles. One of them extracts, in the context of the above example, from the set S of all questions and the way the people in P answered them a small set T of 'critical' questions that suffice to distinguish most of the mindsets that exist in the population surveyed by S.[2] For every two such mindsets σ and τ there will be a question in T on which σ and τ disagree, and which can thus be used to distinguish σ from τ. As an immediate application, T would make a good small questionnaire for a larger study if S was a pilot study designed to test a large number of questions on a smaller population.

More fundamentally, if P is large enough to be representative for some larger population, the tangle-distinguishing quality of T means that for any individual represented by P (but not necessarily in P), one can predict from their answers to the questions in T how they would answer the much larger set of questions in S. Such predictions will be reliable as soon as the individual's views are aligned with *some* mindset – and the theory of tangles can determine how likely this is to be the case.

---

[1] Make America Great Again; Donald Trump's 2016 presidential campaign slogan.

[2] The questions in T will either themselves be in S or be combinations of questions from S.

The gain in conditioning predictions for the unknown answers to questions in S on the known answers to the questions in this particular set T, rather than an arbitrary small subset of S, is that T is so special: the answers to T already identify which mindset on S, if any, an individual belongs to, which enables us to base our predictions for this individual's views (on all of S) on this mindset.

The other fundamental theorem about tangles is about cases when no mindset exists for the questions in S. The algorithm for this theorem is designed to find in such cases a small subset T of S that has the following special property: for every possible consistent[3] way τ of answering T there will be a small set $S_\tau \subseteq T$ of questions, perhaps three or four, whose answers by τ are not shared by enough people for τ to be a mindset. Hence S_τ certifies that this τ, although consistent, cannot be a mindset (on T). Since mindsets are consistent, and hence would be such sets τ, this means that there are no mindsets on T. But then there are no mindsets on S either, because they would include mindsets on T. So our set T, together with all those small sets S_τ, certifies that no mindset on S exists.

This second theorem, therefore, furnishes 'negative' poll returns with a verifiable proof that mindsets not only were not found but do not in fact exist. As with the first theorem, such special sets T and S_τ of questions cannot simply be found by trial and error, since there are too many subsets T of S, too many consistent ways τ of answering T, too many subsets S_τ of every such τ, and too many people in P to test on each of those on S_τ.

In both theorems, the sets T and S_τ are highly valuable in both senses of the word: knowing them gives us a lot of valuable information that we would not otherwise have, and finding them requires some nontrivial mathematics. The theory of tangles provides this, and in many instances T and the S_τ can be computed easily.

This paper is organized as follows. Chapter 1 describes the basic idea behind tangles. It is still written informally, but in a slightly more abstract language than used so far, so that other examples can be expressed in it too.

Such examples will be introduced in Chapter 2. These are by no means a complete cross-section of where tangles can be used. My aim in selecting these examples was to choose potential applications where using tangles can make a particularly striking difference. And not too many: readers are encouraged to take the examples discussed here simply as templates for possible applications in their own field. The fields touched upon in Chapter 2 are:

- Sociology
- Psychology
- Political science
- Education
- Linguistics and analytic philosophy
- Economics

---

[3] Consistency is an easily checkable property of any mindset. It will be formalized later.

Further examples can be found in [2]. They include applications in text analysis, medical diagnostic expert systems, DNA sequencing, image processing, and generic clustering in large data sets. As here, those examples are just intended as pointers: pointers to possible uses of tangles in the empirical sciences that include, but go beyond, the social sciences discussed here.

In Chapter 3 we describe a more formal setup for tangles. It is not the most general formal setup,[4] but it is general enough to express formally the various types of tangle discussed in this paper, and probably most or all the types of tangles with applications in the social sciences.

Chapter 4 then states, in the formal language introduced in Chapter 3, the two fundamental tangle theorems: the *tree-of-tangles* theorem and the *tangle-tree duality* theorem hinted at earlier. It discusses tangle algorithms corresponding to these theorems, and their potential applications for predictions and traditional clustering. Formality is kept to the minimum needed to enable readers to check precisely where tangles might, in principle, play a role in their own fields.[5]

Chapter 5 then revisits the examples from Chapter 2 in the light of the additional precision gained in Chapters 3 and 4. The aim now is to re-cast these examples, whose essence is already understood from their informal treatment in Chapter 2, in a sufficiently formal way that they can be used as templates for similar applications. We shall also be able to describe some deeper aspects of these examples that could not be expressed earlier at the informal level.

1. The idea behind tangles

Consider a collection V of objects and a set $\vec{S}$ of features[6] that each of the objects in V may have or fail to have. Given such a (potential) feature $\vec{s} \in \vec{S}$, we denote its negation by $\overleftarrow{s}$. The pair $\{\vec{s}, \overleftarrow{s}\}$ of the feature together with its negation is then denoted by s, and the set of all these s is denoted by S.

For example, if V is a set of pieces of furniture, then $\vec{s}$ might be the feature of being made of wood. Then $\overleftarrow{s}$ would be the feature of being made of any other material, or a combination of materials, and s could be thought of as the question of whether or not a given element of V is made of wood.

In our example from the introduction, V would be the population P of people polled by our survey S (which, for simplicity, we assume to consist of

---

[4] This is presented concisely in [3], which includes further references for the mathemically inclined to where the main tangle theorems are proved.

[5] Once this has been established in principle, my hope is that such readers get in touch to see how this can be implemented. We have developed some generic algorithms that compute the solutions offered by those theorems on a wide range of potential input data. However for best performance it will be best to fine-tune them to the concrete application intended.

[6] Logicians may prefer to say 'predicates' instead of 'features' here. That would be correct, but I am trying to avoid any (false) impression of formal precision at this stage.

yes/no questions). Then $\{\, \vec{s} \mid s \in S \,\}$ might be the set of 'yes' answers to the questions in S, while $\overleftarrow{s}$ would denote the 'no' answer to the question s.[7]

1.1 Features that often occur together

Tangles[8] are a way to formalize the notion that some features typically occur together. They offer a formal way of identifying such groups of 'typical' features, each 'type' giving rise to a separate tangle. One particular strength of tangles is that the sets of *objects* (elements of V) whose features largely agree with a given type need not be clearly delineated from each other. As in most real-world examples, these sets may be 'fuzzy': in order for certain features to form a type, or tangle, it is *not* necessary that there exist a group of objects in V that have precisely these features – indeed there need not even exist a single such object.

Let us return to the example where V is a set of pieces of furniture. Our list $\vec{S}$ of possible features (including their negations) consists of qualities such as colour, material, the number of legs, intended function, and so on – perhaps a hundred or so potential features. The idea of tangles is that, even though $\vec{S}$ may be quite large, its elements may combine into groups that correspond to just a few types of furniture as we know them: chairs, tables, beds and so on.

The important thing is that tangles can identify such types without any prior intuition: if we are told that a container V full of furniture is waiting for us at customs in the harbour, and all we have is a list of items v identified only by numbers together with, for each number, a list of which of our 100 features this item has, our computer – if it knows tangles – may be able to tell us that our delivery contains furniture of just a few types: types that we (but not our computer) might identify as chairs, tables and beds, perhaps with the beds splitting into sofas and four-posters.

1.2 Consistency of features

To illustrate just how our computer may be able to do this, let us briefly consider the inverse question: starting from a known type of furniture, such as chairs, how might this type be identifiable from the data if it was *not* known?

A possible answer, which will lead straight to the concept of tangles, is as follows. Each individual piece of furniture in our unknown delivery, $v \in V$ say,

---

[7] The arrow notation has nothing to do with vectors. For the moment, it is just a convenient way of associating with a question s its two possible answers 'yes' and 'no', or with a feature $\vec{s}$ its negation $\overleftarrow{s}$. In Chapter 3 the arrow notation will acquire an additional layer of meaning that does have to do with pointing, but not in any geometric sense such as that of vectors.

[8] The word was first introduced by Robertson and Seymour [17] in their ground-breaking work on graph minors. They use it for a region of a graph that hangs together in an intricate way. Intricate in that, while being close-knit in the sense of being difficult to separate, it still does not conform to the usual graph-theoretic notions of being a highly connected subgraph, minor or similar. The abstract notion of tangles that underlies our discussions in this paper was first introduced in [3].

has some of the features from our list $\vec{S}$ but not others. It thereby *specifies* the elements s of S: as $\vec{s}$ if it has the feature $\vec{s}$, and as $\overleftarrow{s}$ otherwise.[9] We say that every $v \in V$ defines a *specification* of S, a choice for each $s \in S$ of either $\vec{s}$ or $\overleftarrow{s}$ but not both. We shall denote this specification of S as

$$v(S) := \{ v(s) \mid s \in S \},$$

where $v(s) := \vec{s}$ if v specifies s as $\vec{s}$ and $v(s) := \overleftarrow{s}$ if v specifies s as $\overleftarrow{s}$.

Conversely, does every specification of S come from some $v \in V$ in this way? Certainly not: there will be no object in our delivery that is both made entirely of wood and also made entirely of steel. Thus no $v \in V$ will specify both r as $\vec{r}$ rather than $\overleftarrow{r}$, and s as $\vec{s}$ rather than $\overleftarrow{s}$, when $\vec{r}$ and $\vec{s}$ stand for being made of wood or steel, respectively. In plain language: no specification of S that comes from a real piece of furniture can contain both $\vec{r}$ and $\vec{s}$, because these features are inconsistent.

Let us turn this manifestation in V of logical inconsistencies within $\vec{S}$ into a definition of 'factual' inconsistency for specifications of S in terms of V. Let us call a specification of S *consistent* if it contains no inconsistent triple, where an *inconsistent triple* is a set of up to three[10] features that are not found together in any $v \in V$. Specifications of S that come from some $v \in V$ are clearly consistent. But S can have many consistent specifications that are not, as a whole, witnessed by any $v \in V$.[11]

Tangles will be specifications of S with certain properties that make them 'typical' for V. Consistency will be a minimum requirement for this. But since any specification of S that comes from just a single $v \in V$ is already consistent, tangles will have to satisfy more than consistency to qualify as 'typical' for V.

1.3 From consistency to tangles

It is one of the fortes of tangles that they allow considerable freedom in the definition of what makes a specification of S 'typical' for V – freedom that can be used to tailor tangles precisely to the intended application. We shall describe

---

[9] For mathematicians: note that this is well defined even when both $\vec{s}$, $\overleftarrow{s}$ are in $\vec{S}$, as we assume.

[10] It might seem more natural to say 'two' here, as in our wood/steel dichotomy above. Our definition of consistency is a little more stringent, because the mathematics behind tangles requires it. Note that, formally, the elements of an inconsistent 'triple' need not be distinct; an 'inconsistent pair' of *two* features $\vec{r}$, $\vec{s}$ not shared by any $v \in V$, for example, also counts as an 'inconsistent triple', the triple $\{\vec{r}, \vec{r}, \vec{s}\} = \{\vec{r}, \vec{s}\}$.

[11] Here is a simple example. Suppose some of our furniture is made of wood, some of steel, some of wicker, and some of plastic. Denote these features as $\vec{p}$, $\vec{q}$, $\vec{r}$, $\vec{s}$, respectively, and assume that $S = \{p, q, r, s\}$. Then the specification $\sigma = \{\overleftarrow{p}, \overleftarrow{q}, \overleftarrow{r}, \overleftarrow{s}\}$ of S is consistent, because for any three of its elements there are some items in V that have none of the three corresponding features: those that have the fourth. But no item fails to have all four of these features. So the consistent specification σ of S does not come from any one $v \in V$.

this formally in Chapter 3. But we are already in a position to mention one of the most common ways of defining 'typical', which is just a strengthening of consistency.

To get a prior feel for our (forthcoming) formal definition of 'typical', consider the specification of S in our furniture example that is determined by an 'ideal chair' plucked straight from the Platonic heaven: let us specify each $s \in S$ as $\vec{s}$ if a typical chair has the feature $\vec{s}$, and as $\overline{s}$ if not.[12] This can be done independently of our delivery V, just from our intuitive notion of what chairs are. But if our delivery has a sizable portion of chairs in it, then this phantom specification of S that describes our 'ideal chair' has something to do with V after all.

Indeed, for every triple $\vec{r}$, $\vec{s}$, $\vec{t}$ of features of our 'ideal chair' there will be a few elements of V, at least n say, that share these three features. For example, if $\vec{r}$, $\vec{s}$, $\vec{t}$ stand for having four legs, a flat central surface, and a near-vertical surface, respectively, there will be – among the many chairs in V which we assume to exist – a few that have four legs and a flat seating surface and a nearly vertical back.

By contrast, if we pick twenty rather than three features of our ideal chair there may be no $v \in V$ that has all of those, even though there are plenty of chairs in V. But for *every* choice of three features there will be several – though which these are will depend on which three features of our ideal chair we have in mind.

Simple though it may seem, it turns out that for most furniture deliveries and reasonable lists S of potential features this formal criterion for 'typical' distinguishes those specifications of S that describe genuine types of furniture from most of its other specifications.[13] But in identifying such specifications as 'types' we made no appeal to our intuition, or to the meaning of their features.[14]

So let us make this property of specifications of S that describe 'ideal' chairs, tables or beds into our formal, if still ad-hoc, definition of 'typical': let us call a specification of S *typical* for V if every subset of at most three of its elements is shared by at least n elements of V, where n is now a fixed parameter on which our notion of 'typical' depends, and which we are free to choose.

Crucially, this definition of 'typical' is purely intrinsic: it depends on V, but it makes no reference to what a typical specification of S 'is typical of'.

---

[12] Let us ignore for the moment the possibility that the question s may not have a clear answer for chairs, as would be the case, say, for questions of colour rather than function. This is an issue we shall have to deal with, but which tangles can indeed deal with easily.

[13] ... of which there are many: if S has 100 elements, there are 2^{100} specifications of S.

[14] This is not to say that the use of tangles is free of all preconceptions, biases etc.: the choice of S, for example, is as loaded or neutral as is would be in any other study that starts with a survey. The statement above is meant relative to the given S once chosen. In Chapter 3 we shall discuss how the deliberate use of preconceptions, e.g. by declaring some questions in S as more fundamental than others, can help to improve tangles based on such preconceptions. We shall also see how to do the opposite: how to find tangles that arise naturally from the raw data of S and V, without any further interference from ourselves.

Specifications of ideal chairs, tables or beds are all typical in this sense: they all satisfy the same one definition.

Equally crucially, a specification of S can be typical for V even if V has no element that has *all* its features at once. Thus, we have a valid and meaningful formal definition of an 'ideal something' even when such a thing does not exist in the real world, let alone in V.

Relative to the definition of 'typical' we can now define tangles informally:

> A <u>tangle</u> of S is any typical specification of S.

Since our ad-hoc definition of 'typical' is phrased in terms of small subsets of $\vec{S}$, sets of size at most 3 (of which there are not so many), we can compute tangles without having to guess them first.

In particular, we can compute tangles of S even when V is 'known' only in the mechanical sense of data being available (but not necessarily understood), and S is a set of potential features that are known, or assumed, to be relevant but whose relationships to each other are unknown. Tangles therefore enable us to *find* even previously unknown 'types' in the data to be analysed: combinations of features that occur together significantly more often than others.

The two fundamental tangle theorems mentioned in the introduction then allow us to either structure these types by refinement and determine a small set of critical features that suffice to distinguish them, or else produce verifiable evidence that the data is too unstructured to contain significant types at all.

1.4 The predictive power of tangles

The general idea of trying to predict a person's likely behaviour in a future situation from observations of his or her actual behaviour in some past situations is as old as humanity: as we learn 'how someone ticks', we are better able to make such predictions.[15]

If we have the chance to choose, or even design, those earlier situations, e.g. by selecting them carefully from a collection of past situations in which our individual's behaviour was observed, or by devising a test consisting of hypothetical situations the response to which our individual is willing to share with us, we have a chance of improving our predictions by choosing particularly relevant such past or hypothetical scenarios. Tangles can help to identify these.

Let us think of V as a set of people in whose likely actions we are interested, and of $\vec{S}$ as the set of potential such actions. To keep things simple, let us consider the example from the introduction, where $\vec{S}$ is a set of potential views of the people in V,[16] i.e., we wish to predict how a person would answer a question s from S. Let us assume that, as the basis for our prediction, we

---

[15] We are talking about 'predictions' here at a quantitative level of, at best, weather forecasts; not about absolute predictions whose failure, even once, invalidates the entire theory.

[16] Thus, 'holding the view $\overline{s}$' would count as an 'action' in our new context; if desired, think of it as the action of answering the question s as $\vec{s}$.

are allowed to quiz that person on some small set T of questions, and base our prediction for their answer to s on their known answers to the questions in T. Our aim is to choose T so as to make these predictions particularly good.

As indicated in the introduction, the tree-of-tangles theorem will likely produce a particularly valuable and much smaller set T of questions whose answers entail more predictive power for S than an arbitrary subset of S of that size would. This is because T consists of just enough questions to distinguish all the tangles of S, and tangles represent the typical ways to answer S. The answers of an individual to the questions in T thus determine exactly one such type: there is only one typical way of answering all the questions in S that includes these particular answers to T. This typical way of answering S, a tangle of S, is a particularly good prediction for the answers of our individual of whom only his or her answers to T are known, because he or she is more likely to answer in some, and hence this, typical way than in an arbitrary other way.

It may seem that there is a practical problem with this approach in that we are trying to base our predictions for the answers to questions in S to be expected from the people in V on being able to compute all the tangles of S first, which may be possible only if we already know what we are trying to predict: the answers of all the people in V to all the questions in S. However, there is no such problem. In any real-world application we would indeed compute these tangles based on full knowledge of how the elements of V specify S. But we would be interested in predictions about individuals outside the set V. For this to be possible, V would have to be representative for that larger population – which, however, is a standard assumption one always has to make (and justify).

A classical application would be that S is a pilot study run on a small subset V of a larger population P, and T is a study with fewer but particularly relevant questions selected from S, to be run on P. Then the answers to T of an individual in P can justify predictions on how this individual would answer the rest of S, and T is a particularly well-chosen subset of S for this purpose.

1.5 Witnessing sets and functions

In the example discussed in the introduction, where S is a political survey and tangles are mindsets – collections of views that are often held together – we quantified the meaning of 'often held together'. We did so in terms of a subset X of all the people in $V = P$ that should not be too small (but comprise at least, say, $1/10$ of the population P) and whose views are typified by the given tangle τ in that, for every question $s \in S$, some 80% of X answer s the way τ does.[17] In our furniture example, the tangle of being a chair would be witnessed in this way by the set X of chairs in V: every feature of our 'ideal chair' τ will be shared by some 80% of all the chairs in V, though not all by the same 80%.

---

[17] Recall that τ, being a tangle of S and thus a specification of S, 'answers' every question $s \in S$ by choosing either $\vec{s}$ or $\overleftarrow{s}$.

Formally, we say that a set $X \subseteq V$ *witnesses* a specification τ of S if, for every $s \in S$, there are more v in X that specify s as τ does than there are $v \in X$ that specify s in the opposite way. If these majorities are greater than 2/3, then τ is likely to be a tangle for most definitions of 'typical' (see Section 1.3).

More generally, a 'weight' function $w \colon V \to \mathbb{N}$ *witnesses* τ if, for every $s \in S$, the collective weight of the $v \in V$ that specify s as τ does exceeds the collective weight of the $v \in V$ that specify s in the opposite way.[18]

Much of the attraction and usefulness of tangles stems from the fact that, in practice, most of them have such witnessing sets or functions [12]. But it is important to note that the definition of a tangle, even our preliminary definition from Section 1.3, does not require that such sets or functions exist. It relies only on notions of consistency and[19] type, which are defined by banning small subsets of $\vec{S}$ deemed 'inconsistent' or 'atypical' from occuring together in a tangle.

In some contexts, tangles of S can even be defined without any reference to V at all. In our furniture example we could have defined the consistency of a set of features, or predicates, about the elements of V in purely logical or linguistic terms that make no appeal to V. If $\vec{r}$ stands for 'made entirely of wood' and $\vec{s}$ stands for 'made entirely of steel', then the set $\{\vec{r}, \vec{s}\}$ is inconsistent. But we have two ways of justifying this. The reason we chose to give was that no object in V is made entirely of wood and also made entirely of steel. But we might also have said that these two predicates are logically inconsistent – which implies that there is no such object in V but which can be established without examining V.

The way consistency and type are defined formally [3] as part of the notion of abstract tangles is something half-way between these two options: it makes no reference to V but refers only to some axiomatic properties of $\vec{S}$ which reflect our notion that $\vec{S}$ is a set of 'features'. In this way it also avoids any appeal to logic or meaning.

For the rest of this paper the only important thing to note about witnessing sets or functions is that while many tangles have them, tangles can be identified, distinguished by the first tangle theorem, or ruled out by the second tangle theorem without any reference to such sets or functions. *Pars pro toto:* the mindset of being socialist can be identified without having to find any actual socialists, let alone delineating these as a social group against others.

---

[18] Formally, $w \colon V \to \mathbb{N}$ *witnesses* τ if $\sum \{ w(v) \mid v(s) = \tau(s) \} > \sum \{ w(v) \mid v(s) \neq \tau(s) \}$ for every $s \in S$.

[19] Our example definition of 'typical' in Section 1.3 implied consistency. But there are other useful notions of type that do not, in which case we simply add consistency to the requirements made of a tangle.

2. Examples from different contexts

The aim of this chapter is to illustrate the range of potential uses of tangles by a handful of further examples from different contexts, still at the level of informality adopted in Chapter 1. After introducing the formal setup for tangles and their two main theorems in Chapters 3 and 4, we shall then return to these examples in Chapter 5 and see how they are made precise in that setup.

The examples given below are mostly from the social sciences. Since I am not an expert in any of these, I shall not attempt more than to indicate the kind of use that tangles might find in these disciplines. Any actual use will require genuine expertise in the respective field, but my hope is that the examples below indicate enough of their potential to entice the experts to pursue this further.

Further examples including text analysis, diagnostic expert systems, image processing or compression, and DNA or protein sequencing can be found in [2].

The default format for describing the examples below will be to...

- name the set V of objects studied;[20]
- describe the set S of their potential features;
- discuss when a specification of S is deemed to be typical;
- describe what tangles – typical specifications – of S mean in this setup;
- indicate some uses and applications of tangles in this context.

When we return to these examples in Chapter 5, we shall further describe what the two tangle theorems mean in the context of each of these examples: how they help us organize the tangles into a global structure (e.g., how some tangles refine others), how tangles can be distinguished by some small subset of particularly crucial questions or features, and how to obtain quantifiable evidence of the non-existence of tangles whenever our data is too unstructured.

Consistency is always defined as in Chapter 1.2. The definition of when a specification of S is deemed to be typical is always made by declaring some small sets of elements of $\vec{S}$ as atypical, and then calling a specification of S *typical* if it is consistent and contains none of these 'forbidden' subsets. The collection of these 'forbidden' atypical sets will usually be denoted by $\mathcal{F}$.

Tangles, then, are typical specifications of S (as earlier): consistent sets of features that contain exactly one of $\vec{s}$ and $\overleftarrow{s}$ for every $s \in S$ and have no subset in $\mathcal{F}$.

2.1 Sociology: discovering mindsets, social groups, and character traits

Mindsets were the example of tangles we discussed in the introduction, where

---

[20] As indicated in Chapter 1.5, for tangles in their most general framework there is not even a need for objects: tangles can be made to work with just a set $\vec{S}$ and some axiomatic assumptions about $\vec{S}$ that reflect properties of 'features of objects' sufficiently to enable us to formally define consistency. Then tangles can be defined as before: as consistent specifications of S, and the two tangle theorems can be proved even in this very abstract framework [3, 7, 9].

we have a set V of people polled with a questionnaire S.[21] We may assume for our model that S consists of yes/no questions: if it does not, we can still use it, and simply translate its answers into answers of an equivalent imaginary questionnaire of yes/no questions, which we then use as the basis for computing our tangles.[22]

Specifying the collection $\mathcal{F}$ of 'forbidden' sets of features is our main tool for determining how broad or refined the mindsets will be that show up as tangles. If we forbid no feature sets, i.e., leave $\mathcal{F}$ empty, then every specification of S returned by one of the people polled will determine a tangle – because it will be consistent and hence typical if $\mathcal{F} = \emptyset$.[23]

Such tangles would be too 'fine' to be helpful. If we do not wish to influence which tangles are returned by our algorithms other than by determining how 'broad' the mindsets they define should be, we can define $\mathcal{F}$ as in Chapter 1.3: as the set of all triples $\{\,\vec{r},\,\vec{s},\,\vec{t}\,\}$ shared by no more than n people polled, where n is a number we can experiment with at the computer and see how it influences the number (and fineness) of tangles it returns.

But we can also decide to add some sets of features to $\mathcal{F}$ that we think of as inconsistent because of what these features mean. What these are will be up to us: we might add sets of features that are intuitively inconsistent, or sets of features that will occur together as answers on the same questionnaire only if someone tampered with it or tried to influence the survey.

We might even use $\mathcal{F}$ to deliberately exclude some types of mindsets from showing up as tangles, e.g., mindsets we are simply not interested in. In order to exclude such mindsets we simply add to $\mathcal{F}$ the feature sets that make them uninteresting: then no specifications of S that include these feature sets will show up as tangles.[24] Of course, subsets we forbid for this reason must be specific enough that they occur *only* in mindsets that are not of interest to us.

We already discussed in the introduction what tangles mean in this example – namely, mindsets – and what the two main tangle theorems offer: a small set of critical questions that suffices to distinguish all the existing mindsets and

---

[21] In the language developed since then, the mindsets discussed there were tangles with witnessing sets. From now on, we shall use the word 'mindset' for any tangle in a questionnaire scenario, regardless of whether that tangle has a witnessing set.

[22] For example, if a question s in S asks for a numerical value between 1 and 5 on a scale from 'do not agree at all' to 'agree entirely', we can replace s with five imaginary yes/no questions $s_1, \ldots, s_5$ asking whether the value of s is $1, \ldots, 5$, respectively, or with the four yes/no questions s_i' asking whether the value of s is greater than i, for $i = 1, \ldots, 4$.

[23] This is not the same as saying that all possible specifications of S will be consistent (and hence typical). Specifications that contain an inconsistent triple will be inconsistent even then, and hence not be tangle. But these specifications are not among those returned in the poll – recall how consistency and inconsistent triples were defined.

[24] For example, we might be interested in the opinions about hooligans among football crowds, but want to exclude hooligans themselves from this survey. Since we may not be able to identify them when we hand out the questionnaires, but know some answer patterns they are likely to give, we can add these patterns to $\mathcal{F}$ to ensure that the tangles found are mindsets of spectators that are not themselves hooligans.

on which predictions can be based, and verifiable evidence that no mindsets other than those found exist, possibly none. In Chapter 5.1 we shall add another aspect of the tree-of-tangles theorem: it enables us also to structure the mindsets found hierarchically, into broader mindsets and more focused ones refining these.

Discovering social groupings (not groups, see below) is similar to discovering mindsets, except that S is different now: it can still be thought of as a questionnaire, but the 'questions' it contains may be answered by an observer rather than the subjects $v \in V$ themselves. Thus, its specifications $\vec{s}$ will be more like the 'features' discussed in our furniture example in Chapter 1. What makes tangles special in this application is again that they can identify such groupings without identifying actual groups in terms of their members: tangles are patterns of behaviour (etc.) often found together, not groups of similar people.

While this may sound like a truism, it does differ fundamentally from the classical approach that seeks to find groups as clusters of people. Being defined directly in terms of the phenomena that define social groups, rather than indirectly in terms of the people that display them, tangles bypass many of the usual problems that come with traditional distance-based clustering, such as the fact that an individual is likely to belong to more than one social group. However, tangles can help even in finding those groups – which takes us on to character traits.

Imagine we wish to use tangles for a matchmaking algorithm. As before, they can help us to identify, from the answers people have given to a questionnaire S of character-related questions, some combinations of traits that are typically found together.[25] Thus, tangles in this context will be 'types of character'.

However, for our matchmaking algorithm it is not enough to know such types: it will also have to match individuals. So we will need a metric that tells us which pairs of people are 'close in character' – assuming that matching like individuals is our aim (which, of course, can be disputed). The simplest example of a distance function on the set of people that returned our questionnaire would be to consider two people as close if they answered many questions identically.[26] But tangles can be used to define more subtle distance functions. We shall discuss some of these in Chapter 4.4, and refer to them when we revisit the topic of matchmaking in Chapter 5.1.

To summarize, let us emphasize again what makes the tangles approach to the study (and discovery) of mindsets, social groupings, or traits of character different from traditional approaches. It is that we can mechanically find these mindsets (etc.) from observational data without any prior intuitive hypothesis of what they may be, and that we can find them without having to group the

---

[25] We have done this with data designed to test for the 'big five' personality traits, and got some rather suprising results; see [14].

[26] For mathematicians: this is the Hamming distance in the hypercube 2^S.

people observed accordingly. However, knowing the tangles can then help us also to determine these groups; this will be discussed further in Chapter 4.4.

2.2 Psychology: understanding the unfamiliar

In a way, this example is no more than a special case of the mindset example from (2.1). But the special context may lend it additional relevance.

While it is interesting to search for combinations of, say, political views that constitute hitherto unknown mindsets that can have an impact on political developments, it is not only interesting but crucial to try to *understand* minds that work in ways very different from our own. This is particularly relevant in doctor-patient relationships where the doctor seeks to offer an individual with such a different mind a bridge to society, or even just to their particular environment: to enable them to understand the people around them, and to help these people to understand them.

Tangles can already help bridge this gap at the most fundamental level, the level of notions into which we organize our perceptions. We all have a notion, for example, of 'threat'. But some patients' notion of threat may be different from ours: they may be scared by things we would not see as threatening. And what these are may well come in types: typical combinations of perceptions of everyday phenomena which, for people with a certain psychiatric condition, may combine to a perception of a threat, and thereby form their notion of 'threat' that may well differ from ours.

Tangles are designed to identify such types. The set $\vec{S}$ would consist of various possible perceptions that experience has shown are relevant, and a tangle would identify the combinations of these that are typical in the sense of Chapter 1.3. The definition of 'typical' can be made sufficiently flexible to allow tangles to capture notions, unfamiliar to us, that consist of perceptions that typically occur together in some patients. Once these have been made explicit – recall that every tangle of S will be one specific set of possible perceptions – we can train our intuition on them in an effort to *understand* our patients rather than just collect lists of unrelated symptoms. We shall get back to how tangles can help identify meaning in Section 2.5.

At a higher level, tangles can help to identify psychological syndromes as such, and maybe discover hitherto unknown syndromes. Indeed, psychological syndromes appear to be exactly the kind of thing that tangles model: collections of features that often occur together.

This would be true for any medical condition, or even for mechanical 'conditions' that lead to the failure of a machine. And indeed, there is a corresponding application of tangles for such cases, where they are used to build expert systems for medical (or mechanical) diagnosis [2]. But what makes psychiatric conditions even more amenable to the use of tangles is that the symptoms of which they are combinations are so much harder to quantify. Tangles come into their own particularly with input data that cannot be expected to be reliably precise.

So how would we formalize the search for hitherto unknown psychological syndromes? Our ground set V would be a large pool of patients in some database. The set $\vec{S}$ would be a set of possible symptoms. Tangles will be collections of symptoms that typically occur together: medical, or psychological, conditions or illnesses.

We shall see in Chapter 5.2 that the set T of tangle-distinguishing features returned by the tree-of-tangles theorem will consist of 'critical' symptoms or combinations of symptoms that can be tested on a patient in the process of diagnosis. Every consistent specification of T defines a unique tangle: a unique condition that has all the symptoms in this specification of T. In most cases, this will be the correct diagnosis.

Note, however, that our emphasis here is on *finding* psychiatric conditions: our aim is to discover combinations of symptoms that constitute an illness. It is not on diagnosing a given patient with one of these conditions (although checking the symptoms in T can play an important part on this), but on establishing what are the potential conditions to look for in the diagnosis. It seems to me that a lot has happened in psychology here in recent years, in that conditions are now recognized as illnesses that were not even thought of as typical combinations of symptoms (i.e., tangles) not so long ago.

2.3 Politics and society: appointing representative bodies

As discussed in Chapter 1, the tangle approach to clustering is that we do not primarily seek to divide our set V into groups based on similarity between its elements. This is true even if similarity is measured by how the elements of V specify S, in which case we might group $u, v \in V$ together if they specify many $s \in S$ in the same way (as $\vec{s}$ or $\overline{s}$): if $u(s) = v(s)$ for many $s \in S$.

Rather, determining tangles is more about grouping features: not in the simple sense that we find traditional distance-based clusters in the set $\vec{S}$, but in that we look at how the elements of V specify S and find particular such specifications that are typical for the elements of V. In the mindsets application from Section 2.1, for example, tangles are ways of answering all the questions in S that are typical for how the various $v \in V$ answered them.

Once we have found the tangles of S, however, we can use them to group the elements of V after all. In terms of the mindsets scenario, on which we shall base our further discussion here for better intuition, we would thus be looking for ways to group the elements of V according to their mindsets.

This can be done in various ways, which we shall look at in Chapter 4.4. For example, we might associate each $v \in V$ with the mindset τ of S that represents its views best: the tangle τ of S for which the number of elements s of S that τ specifies as v does is maximum. Conversely we may seek, for each mindset τ found, the person $v \in V$ closest to τ in this sense, and think of these v as best representing the views held amongst the members of V.

If we add our assumption from Section 2.1 that V was chosen so as to represent some larger population P well, we shall then have found a small group

of people that ideally represent the views held in our population P on matters explored by S. This process might be used, then, to appoint delegates to a body whose brief is to make decisions likely to find maximum consensus in P.

This contrasts with the usual democratic process of electing the delegates to such a decision-making body by majority vote. In a first-past-the-post system with constituencies, this can generate parliaments with large majorities even when, in each constituency, the majority of the successful candidate was small but the population is homogeneous enough across constituencies that in most of these the successful candidate is of the same political hue.

In systems with proportional representation this is avoided, but such systems require the previous establishment of political parties. Even when these exist, they may have developed historically in contexts that are less relevant today. Finding the tangles of a political questionnaire is like *finding* the political parties that ought to exist today for the elections at hand. People $v \in V$ representing these virtual 'parties' as described above could then be delegated to our decision-making body straight away.

If we wish to appoint more delegates for tangles, or virtual parties, with a larger following, we could elect representatives from these virtual parties by a standard vote using proportional representation. Alternatively we could simply refine this tangle into smaller tangles (see Chapters 3.2 and 5.3), so that all tangles end up with roughly the same amount of support and could thus be represented by one delegate per tangle.

This approach may be even more relevant outside politics, where there are many situations on a smaller scale in which we seek to appoint a decision-making body. This might be the governors for a school, or a steering committee for a choir. At such a smaller scale there will be no constituencies, and there may be no established parties relevant to the brief of that body. But appointing for every tangle τ of S the member of V (amongst those willing to stand) whose views on S are closest to that tangle would produce a committee likely to represent the views held in V well.

2.4 Education: combining teaching techniques into methods

This application starts from the assumption that different teaching techniques work well or less well with different students: that a given technique is not necessarily better or worse than an alternative for all students at once, but that each may work better for different sets of students.

Ideally, then, each student should be taught by precisely the set of techniques that happen to work best for him or her. Of course, this is impractical: there are so many possible combinations of techniques that most students would end up sitting in a class of their own. But our aim could be to group techniques into, say, four or five groups, to be used in four or five classes held in parallel, so that each student can then attend the one of these four or five classes whose techniques suit him or her best.

The question then is: how do we divide the various techniques into the groups that correspond to the classes? Just to illustrate the problem, consider a very simple example. Suppose first that some students benefit most from supervised self-study while others are best served by a lecture followed by discussion in class. Suppose further that some students understand a grammatical rule best by first seeing motivating examples that prepare their intuition, while others perfer to see the rule stated clearly to begin with and examples only afterwards. So there are four possible combinations of techniques, but maybe we can only have two classes. So we have to group our techniques into pairs.

How shall we select the pairs? Shall we have one class whose teacher lectures and motivates rules by examples first, or shall we group the lecture-plus-discussion class with the technique of introducing rules before examples?

This is where tangles come into their own. Think of V as a large set of students evaluated for our study, and of $\vec{S}$ as a set of teaching techniques. Every tangle of S will be a particular combination of techniques (a specification of S) which, in the language of our furniture example in Chapter 1 (where $\vec{S}$ consists of features), 'typically occur together'. What does this mean in our context?

What it means formally depends on the definition of 'typical' that is determined by our choice of the set $\mathcal{F}$ introduced in Section 2.1. Informally it implies, for example, that if the tangle is witnessed by a set $X \subseteq V$ (see Chapter 1.5) then this X is a group of students that would benefit particularly well from the combination of techniques specified by the tangle: for every technique in that tangle, a majority of the students in X prefer that technique to its converse. Thus, X would be an ideal population for the class in which the techniques from this tangle are used, while a set Y witnessing another tangle would consist of students best served by the teaching techniques specified by that tangle.

Once more, speaking about tangles in terms of witnessing sets helps to visualize their benefits, but it is not crucial as such: the benefit arises from finding the tangles as such, and setting up those four or five classes accordingly. We then *know* that this serves our students best collectively – and can happily leave the choice of class to them.

Let us, from now on, refer to tangles of teaching techniques as *teaching methods*: combinations of techniques that work particularly well together for some students.

When we return to this example in Chapter 5.3, we shall have to address the question of how to set our parameters in such a way that, even if the set $\vec{S}$ of potential teaching techniques is large, we still end up with a desired number of tangles – four or five in our case – that may be dictated by the formal environment, school etc., in which the teaching is taking place. We shall also see how, if we do not wish to let the students choose their class themselves, the tree-of-tangles theorem can help us devise an entry test that assigns students to the class that benefits them best.

2.5 Linguistics and analytic philosophy: how to determine meaning

The aim of this section is to indicate how tangles can help us identify *meaning*. Inasmuch as meaning is constituted in a social process this discussion belongs within the scope of this paper. It also has direct implications for the teaching of languages: in education, but also in machine learning. In Chapter 5.5 we shall discuss an application in which tangles identifying the meaning of a word are used to steer a user of an interactive thesaurus towards the desired word.

To keep the discussion focused we shall concentrate on the meaning of words. There are obvious wider analogues, of course, and indeed the relevance of social aspects to the constitution of meaning will be more relevant still when we talk about the meaning of entire phrases, perhaps depending on their contexts, or the meaning of behaviour in non-verbal communication.

Meaning of words, however, is not only easier to talk about: we have already done much of this when we discussed the furniture example in Chapter 1. There we observed that tangles can identify types of furniture, such as chairs, tables and beds, from lists of features of concrete specimen.

This contrasts with the naïve approach of trying to define words by a list of predicates, as classical dictionaries used to do. The idea there is that something warrants being referred to by that word if and only if it satisfies all the predicates on the list.[27] In other words, the list should be long enough that no things other than those we want to use our word for have all these properties, and it should at the same time be short enough that, conversely, all the things we wish to name by our word do indeed have all the properties from the list.

Put another way, if we assume that every predicate on the list describes a well-defined set of objects,[28] we would like the set of objects to be referred to by the word we are trying to define to be exactly the intersection of those sets: no larger and no smaller. As the furniture example demonstrates well, such a list is unlikely to exist: it is probably impossible to come up with a list of potential features (of furniture) such that the things that have all these features are precisely all the chairs.

The same problem arises more generally if we try to define the meaning of a word by any logical formula of previously defined predicates.

Wittgenstein [18] recognized this dilemma and argued that meaning cannot be captured by a simple taxonomic approach. Instead, he argued, the things referred to by a word behave more like members of a family: individuals who resemble each other but cannot be identified by any list of features that precisely they share.

---

[27] . . . or, more generally, satisfies some logical formula in terms of these predicates

[28] This is another dilemma with this naïve approach: it requires that there exists a hierarchy of predicates, where some are defined 'before' others and can hence be used in the list for their definition. Tangles require no such hierarchy.

Tangles offer a way to quantify Wittgenstein's family resemblances.[29] The chair example in Chapter 1 describes how.

In particular, tangles show that the intrinsic extensional imprecision in most attempts to define the meaning of a word in terms of known predicates simply by way of a logical formula does not make it futile to look for precise alternative ways to define meaning (in other ways than by logical formulas).[30] As we shall see, there are a number of parameters we can use to quantify the notion of a tangle.[31] So even if we take the view that the notion of 'meaning' can be formalized as 'tangle of predicates', as I would argue, there will not be one ultimate such notion.

Rather, there are many precise ones, one for every choice of our quantitative tangle parameters. Then in every 'context' S, i.e., for every list $\vec{S}$ of predicates deemed relevant to the definitions we are trying to make, every tangle of S constitutes an instance of that notion of 'meaning': a potential definition of what one particular word means.[32] Each of these notions will be completely precise: *tangles are a precise way to define extensionally imprecise meanings of words.*

After this rather abstract discussion let us return once more to our concrete example, and summarize how exactly the notion of 'chair' is given by a tangle. We start with a list S of potential features of furniture. Any tangle of S is one particular specification of S: a choice of either $\vec{s}$ or $\overleftarrow{s}$ for every $s \in S$, either confirming the feature $\vec{s}$ or confirming its converse $\overleftarrow{s}$. The combination of features that we might think of as decribing 'the perfect chair' would be a tangle – no matter whether such a chair exists in the world or not.

Although every specific tangle in this context is a concrete list of predicates, the notion of 'tangle' as such is purely formal and makes no reference to such predicates: a *tangle* is a specification of S that satisfies certain formal requirements – consistency and being typical – which say that it must not contain certain small sets of features (which in our context are predicates). These requirements ensure that tangles define 'good' notions rather than 'bad' ones. The settings of certain parameters in these requirments also determine whether the corresponding tangles capture broad or narrow notions.

Thus, whether or not a combination of potential features (a specification of S) is a tangle is decided by a precise definition. This does not imply, however, that it must be clear for every piece of furniture whether or not it 'belongs to' a given tangle, e.g., whether it is a chair. Indeed this is *not* clear for real pieces

---

[29] Thanks to Nathan Bowler for pointing this out to me.

[30] It seems that this difference has occasionally been overlooked.

[31] For example, by choosing the parameter n from our discussion of the term 'typical' in Chapter 1.3 – we shall meet this again in the sets $\mathcal{F}_n$ in Chapter 3.1 – or by choosing an order function as described in Chapter 3.4.

[32] In our earlier furniture example, these were the meanings of 'chair', 'table', 'bed' and so on. By our choice of quantitative tangle parameters we can influence how broad or narrow that meaning is going to be.

of furniture in the real world, and it is one of the strengths of our approach that tangles offer a precise way to capture even extensionally imprecise notions such as this one.

2.6 Economics: identifying customer and product types

The context of economics offers a wide range of interesting choices for our set V of 'objects' studied together with some potential 'features' listed in a set S. This is illustrated particularly well by a pair of examples described below, which describes two complementary, or 'dual', aspects of the same scenario.

In our first example, let V be a set of customers of an online shop, and S the set of items sold at this shop. Let us assume that each customer v makes a single visit to the shop, specifying $s \in S$ as $\vec{s}$ if v includes s in his or her purchase, and as $\overleftarrow{s}$ if not. We shall think of the specification $v(S)$ of S as v's 'shopping basket', dividing as it does the set of items into those bought and those not bought.

An arbitrary specification of S, then, is a hypothetical shopping basket, and a tangle of S is a *typical* shopping basket for this set of customers.[33] If we mentally identify a customer v with his or her shopping basket $v(S)$, we may also think of a tangle of S as a (hypothetical) *type of customer* whose 'features' are his or her purchases $v(s) = \vec{s}$ and non-purchases $v(s) = \overleftarrow{s}$.[34]

Alternatively, we may think of S, the set of items of our shop, as our set of *objects* (which would normally be denoted by V, but this letter is taken now), and of the customers in V as potential 'features' of these items: a customer v becomes a feature $\vec{v}$ of precisely those items s that v bought, and a feature $\overleftarrow{v}$ of those items s that v did not buy. Each item s then defines a specification $s(V)$ of V, specifying those v that bought it as $s(v) := \vec{v}$, say, and those that did not buy it as $s(v) := \overleftarrow{v}$. We may think of $s(V)$ as something like the 'popularity score' of the item s with the customers in V.

Note that the information encoded here is no different from that described by our earlier setup, where every customer specified each item s as $\vec{s}$ or $\overleftarrow{s}$. We thus have two ways now of describing the same set of customer preferences.[35]

What are the tangles in our second setup? They are tangles, or typical specifications, of V: hypothetical popularity scores with the customers in V

---

[33] This is not the same as just a group of items 'often bought together', as are already suggested by some online shops today. A shopping basket defined by a tangle is typical in a more subtle way; for example, it is also typical in what it does *not* include.

[34] This is analogous to thinking of the chair tangle as a type of furniture, or of a mindset tangle as a type of person holding such views. In each case, the tangle is a set of features typical for the elements of V, and we may think of it as a hypothetical 'typical' element of V that has exactly these features.

[35] For mathematicians: there is a formal duality here in that $v(s) = s(v)$ with the obvious interpretations of v as a function $S \to \{0, 1\}$ and s as a function $V \to \{0, 1\}$. The customer preferences described in two ways can be formalized as the edge set of the bipartite graph with vertex classes V and S and edges vs whenever v bought s. This edge set can be described alternatively as a list of neighbourhoods of the vertices in V or of those in S.

that are typical for the items in S. If we mentally identify an item s with its popularity score $s(V)$, we may think of a tangle of V as a *type of item* whose 'features' are its fans $s(v) = \vec{v}$ and non-fans $s(v) = \bar{v}$.

Let us look at some examples of both kinds of tangle, and of how they might be used.

Tangles of S are typical (if hypothetical) shopping baskets, collections of items that are typically bought or avoided together. For example, there might be a tangle, or 'typical shopping basket', full of ecological items but containing no environmentally harmful ones. Another might contain mostly inexpensive items and avoid luxury ones.[36]

Note that 'inverting' these tangles will not normally give a tangle. Indeed, a shopping basket full of luxury goods and avoiding cheap items is unlikely to be typical, because customers that like, and can afford, luxury good will not necessarily shun inexpensive items. Similarly, inverting the ecological tangle will not produce another tangle, since there is no unifying motivation amongst shoppers to buy environmentally damaging goods, and probably no unifying motivation to avoid 'green' products either.

If we personify these tangles as indicated earlier, and think of them as (hypothetical) people putting together these hypothetical shopping baskets, we could think of our second tangle as the budget-oriented customer type, one that prefers inexpensive brands and cannot afford or dislikes unnecessarily expensive ones, and of the first as the ecological type that prefers organic foods and degradable detergents but avoids items wrapped in plastic. And, crucially, our tangle analysis might throw up some unexpected customer types as well – perhaps one that prefers items showing the picture of a person on the packaging.

Applications of tangles of S might include strategies for grouping goods in a physical shop, or running advertising campaigns targeted at different types of customer.

Tangles of V, on the other hand, are typical popularity scores, ways of dividing V into fans and non-fans that occur for many items simultaneously. If what interests us about the items in S is mainly how they appeal to customers, we may think of tangles of V as types of items.

For example, assume that the ecological goods in our shop are substantially more expensive than non-green competing goods. Then V splits neatly into a 'green fan set' of customers, those likely to buy these green items even though they are pricier, and the rest of all customers, who will avoid them because they are more expensive. This division of V, or hypothetical popularity score, is borne out in sufficient numbers[37] by the green items in S to form a tangle

---

[36] In our current informal setup, these would be tangles of subsets of S, not of S itself – for example, of the subset of the ecologically critical or the unusually priced items. When we revisit this example in Chapter 5.6, we shall find a way of capturing all groups of items, such as the ecological or the inexpensive ones, by tangles of the same set.

[37] This is a reference to the number n used in Chapter 1.3 to define tangles informally. It will be referred to again when we define tangles formally in Chapter 3.1.

of V, since these items are both liked by the 'green' customers and disliked by the others for their price.

An application of finding tangles of V could be to set up a discussion forum for each tangle and invite the fans for that tangle to join. Since they have shown similar shopping tastes, chances are they might benefit more from hearing each others views than could be expected for an arbitrary discussion group of customers.

One last remark, something the reader may be puzzling over at this point. Since our two types of tangle are 'dual' to each other, it so happens that witnessing sets for the first type are the same kinds of sets as tangles of the second type: they can both be thought of as sets of customers. But they are not the same. A set $X \subseteq V$ witnessing a tangle of S is a set of customers such that for every item $s \in S$ either most of the people in X bought s (if the tangle specifies s as $\vec{s}$) or most of the people in X did not buy s (if the tangle specifies s as $\overleftarrow{s}$). By contrast, a tangle of V (thought of as the set of those v which it specifies as $\vec{v}$)[38] is a group of customers such that every three of them jointly bought some sizable set of items (at least n).[39]

Put more succinctly: in the first case, most people in some set of customers (not too few) agree about every single item, while in the second case the tastes of every few (e.g., three) customers are witnessed by some common set of items (not too few), regardless of whether their tastes coincide on these items or not.

3. The formal setup for tangles

The aim of this chapter is to give a more formal definition of tangles: just formal enough to enable readers to try them out in their work if they wish.[40] The definitions here will be less general than is possible, but general enough to model all the examples of tangles we shall be looking at here. More general definitions, for which all the theorems still hold, can be found in [3].

We shall make precise the following notions introduced informally so far:

- *(potential) features* s of (elements v of) V
- *specifications* $\vec{s}$, $\overleftarrow{s}$ of such s
- *consistency* of specifications
- *types* (or *typical* specifications of S)
- *tangles*

Readers consulting the mathematical papers on tangles referenced here should be aware that, for historical reasons, these things have different names there:

---

[38] Unlike in other contexts, every v in this example has a default specification $\vec{v}$: the set of purchases of v rather than his non-purchases. The set $\{\, \vec{v} \mid v \in V \,\}$, therefore, is well defined.

[39] Indeed, more is true: for every three customers v, regardless of how the tangle τ specified them, there exists a set of at least n items that are each bought by v if $\tau(v) = \vec{v}$ *and* not bought by v if $\tau(v) = \overleftarrow{v}$.

[40] If you do, please feel free to contact me for support at any time.

- Features of (elements of) V are called *separations* of V.
- Specifications of features are *orientations* of those separations.
- *Consistency* is defined more narrowly than here: specifications of S that are consistent in the sense defined formally below are treated in [7] as *robust profiles*, and in [6] as *abstract tangles*.
- Typical specifications of S are those that *avoid* some given collection $\mathcal{F}$ of subsets of $\vec{S}$, in that they have no subset that is an element of $\mathcal{F}$.
- *Tangles* are the same as here: they are typical consistent specifications of S. If reference is desired to the set $\mathcal{F}$ underlying the notion of 'typical' used for a given tangle, it is called an $\mathcal{F}$-tangle.[41]
- The operations $\wedge$ and $\vee$ have their meanings swapped.

In the interest of readability, these traditional names will not be used in the rest of this paper.

The two main tangle theorems will be stated more formally in the next chapter, in terms of the notions introduced below. The discussion of *predictions* is also deferred to the next chapter, since predictions based on tangles make reference to the first of these theorems, the *tree-of-tangles* theorem.

Also included in this chapter will be some basic facts about tangles that can help the reader develop an intuition for them. We shall see how some tangles evolve out of others, and die again to spawn new tangles – just as the detailed notions of 'armchair' and 'dining chair' refine the more basic notion of 'chair', which at their level of detail recedes and becomes less meaningful.

Importantly, we discuss how to *influence* this 'evolution of tangles' according to our needs and aims, by organizing features into hierarchies of 'orders'.

We shall also make precise the notion of *duality* for tangles, of which we already caught a glimpse in Chapter 2.6. This, too, adds substantially to our toolkit for applications.

3.1 Tangles of set partitions

A *partition* of a set V, for the purpose of this paper, is an unordered pair $\{A, B\}$ of disjoint subsets of V, possibly empty, whose union is V. Its two elements, the complementary sets A and B, are its *specifications*.[42] The sets A and $B = V \smallsetminus A$, which we also denote by $\overline{A}$, are *inverses* of each other.

When S is a set of partitions of V, we denote the specifications of its elements s as $\vec{s}$ and $\overleftarrow{s}$,[43] and write $\vec{S}$ for the set of all the specifications of its

---

[41] When tangles are mentioned without any $\mathcal{F}$ being specified, then this $\mathcal{F}$ is either arbitrary, or it is some specific $\mathcal{F}$ traditionally considered in the given context.

[42] For example, V might be a set of people whom we asked a yes/no question. Then V is made up of the set A of the people that answered yes and the complementary set B of the people that answered no. The partition $\{A, B\}$ still records the information of how the vote was split across V, but 'forgets' which of its two elements corresponds to which vote.

[43] It is meaningless here to ask which is which. For example, if $s = \{A, B\}$ as above, whether $\vec{s} = A$ and $\overleftarrow{s} = B$ or $\vec{s} = B$ and $\overleftarrow{s} = A$. This is because the two elements of $s = \{A, B\}$ are 'born equal': the expressions '$\{A, B\}$' and '$\{B, A\}$' denote the same set or

elements. A subset of $\vec{S}$ is a *specification of S* if for every $s \in S$ it contains either $\vec{s}$ or $\overleftarrow{s}$ but not both.

The elements of S are called *potential features* of (the elements of) V, those of $\vec{S}$ are their *features*. An element of V *has* a feature A if it lies in A. Given two features $\vec{r} = A$ and $\vec{s} = C$, we write

$$\vec{r} \vee \vec{s} := A \cup C \quad \text{and} \quad \vec{r} \wedge \vec{s} := A \cap C.$$

Note that $\vec{r} \vee \vec{s}$ and $\vec{r} \wedge \vec{s}$ need not be elements of $\vec{S}$, i.e., features of V in terms of our given set S of potential features. But they become features if we add to S the partitions $\{A \cup C, \overline{A \cup C}\}$ and $\{A \cap C, \overline{A \cap C}\}$ as potential features.

We read $\vee$ as 'or' and $\wedge$ as 'and'. This corresponds to the fact that the elements of V that have the feature $\vec{r} \vee \vec{s}$ (if it is a feature) are those in $A \cup C$, which are precisely those elements of V that have the feature $\vec{r}$ or the feature $\vec{s}$ (or both), and similarly with $\wedge$.

A subset of $\vec{S}$ is *inconsistent* if it contains three[44] features $\vec{r}$, $\vec{s}$, $\vec{t}$ such that no element of V has them all, i.e., such that $\vec{r} \wedge \vec{s} \wedge \vec{t} = \emptyset$. Otherwise the subset is *consistent*. For example, every feature $\vec{s} = A$ is inconsistent with its inverse $\overleftarrow{s} = \overline{A}$, because $\vec{s} \wedge \overleftarrow{s} = A \cap \overline{A} = \emptyset$. Similarly, $\vec{r}$, $\vec{s}$ and the inverse of $\vec{r} \wedge \vec{s}$ form an inconsistent triple of features.[45]

Given a set $\mathcal{F}$ of subsets of $\vec{S}$, a consistent specification of S that has no subset in $\mathcal{F}$ is called $\mathcal{F}$*-typical*, or an $\mathcal{F}$*-tangle* of S. Given an integer n, we write $\mathcal{F}_n$ for the set of all sets $\{\vec{r}, \vec{s}, \vec{t}\}$ of up to three features that no n elements of V have in common, i.e., which satisfy $\vec{r} \wedge \vec{s} \wedge \vec{t} = U$ for some $U \subseteq V$ with $|U| < n$:

$$\mathcal{F}_n = \{\, \{A, B, C\} \subseteq \vec{S} : |A \cap B \cap C| < n \,\}.$$

Thus, a specification τ of S is an $\mathcal{F}_n$-tangle if for every triple of features in τ there are at least n elements of V that share these three features.[46]

A *typical* specification, or *tangle*, of S is an $\mathcal{F}_n$-tangle for some n specified in the context, possibly as a function of $|V|$ or $|S|$. If no n is given explicitly it is assumed to be 1, in which case 'typical' is no more than 'consistent'.[47]

---

unordered pair. If this did not worry you, let it not worry you now.

[44] These features $\vec{r}$, $\vec{s}$, $\vec{t}$ need not be distinct. For example, our subset of $\vec{S}$ will already be inconsistent if it contains two features $\vec{r}$, $\vec{s}$ such that $\vec{r} \wedge \vec{s} = \emptyset$; this is included in the definition given, simply by taking $\vec{t} = \vec{s}$ or $\vec{t} = \vec{r}$.

[45] Indeed, if $\vec{r} = A$ and $\vec{s} = C$, say, then the inverse of $\vec{r} \wedge \vec{s}$ is $\overline{A \cap C}$. Let us denote this as $\vec{t}$. Then $\{\vec{r}, \vec{s}, \vec{t}\}$ is inconsistent, because $\vec{r} \wedge \vec{s} \wedge \vec{t} = A \cap C \cap (\overline{A \cap C}) = \emptyset$.

[46] In some contexts it may be more appropriate to require this only on average: that the number of elements of V sharing a given triple of features in τ is at least n on average, taken over all the triples in τ. We call such τ *averaging $\mathcal{F}_n$-tangles*. We might further require that τ is *stable*: locally optimal in the sense that replacing any one of its elements with its inverse reduces the average size of the intersections of the triples in τ.

[47] Indeed, $\mathcal{F}_n$-tangles for any n are also $\mathcal{F}_m$-tangles for all $m < n$, since $\mathcal{F}_m \subseteq \mathcal{F}_n$. In particular, they are $\mathcal{F}_1$-tangles, but $\mathcal{F}_1$ is just the set of inconsistent triples in $\vec{S}$.

Let us look at a couple of very simple tangles. The simplest tangles are the $\mathcal{F}_1$-tangles that are 'focused on' some fixed $v \in V$:

$$\tau_v = \{\, \vec{s} \in \vec{S} \mid v \in \vec{s} \,\}.$$

These tangles exist for every S and v, so their existence does not tell us anything about S (which, of course, is our aim in studying its tangles). Thus, $\mathcal{F}_n$-tangles are interesting only for $n \geqslant 2$.

Every $\mathcal{F}_n$-tangle of S contains all the features $\overline{s}$ for which $\vec{s}$ has fewer than n elements, because it cannot contain $\{\,\vec{s}\,\} \in \mathcal{F}_n$ as a subset.

If S is the set of *all* bipartitions of V, then the only tangles τ of S are the focused tangles τ_v.[48] Since $\{\{v\}, \overline{\{v\}}\} \in S$, this means that $\{v\} \in \tau$, and hence that τ is not an $\mathcal{F}_2$-tangle. The set S of all bipartitions of V, therefore, has no $\mathcal{F}_n$-tangles for $n \geqslant 2$. For this reason, we shall not be interested in tangles of this entire set S.

3.2 The evolution of tangles

However, we do often consider the set S of all partitions of V in increments: we study tangles of subsets $S_1 \subseteq S_2 \subseteq \ldots$ of S that ultimately exhaust S. These S_k have to be carefully chosen for this to make sense, and we look at how to do this in Section 3.4. But if they are, the tangles of the various S_k can interact in interesting ways, and studying this interaction adds greatly to our understanding of how they all together describe the features of V.

To see how this works, note that for $i < k$ every tangle τ_k of S_k *induces* a tangle $\tau_i = \tau_k \cap \vec{S_i}$ of S_i. Indeed, since S_i is a subset of S_k, each of its elements s is specified by τ_k as $\vec{s}$ or $\overline{s}$, and the set of all these specifications of elements of S_i is consistent (because its superset τ_k is). Similarly, if τ_k is an $\mathcal{F}$-tangle for some $\mathcal{F}$ then so is τ_i, since any subset of τ_i in $\mathcal{F}$ would also be a subset of τ_k in $\mathcal{F}$ (which by assumption does not exist).

Conversely, however, a tangle τ of S_i need not *extend to* a tangle of S_k in this way: there may, but does not have to, be a tangle of S_k whose elements in $\vec{S_i}$ are exactly τ. This is because extending τ to a tangle of S_k requires that we specify all the $s \in S_k$ that are not already in S_i as well; and there may simply not be a consistent way of doing that, or one that avoids creating a subset that lies in our chosen $\mathcal{F}$.

---

[48] For the mathematically inclined, here is the easy and pretty proof. If we can find any $v \in V$ such that $\{v\} \in \tau$, then $\tau = \tau_v$, because any $A \in \tau$ not containing v would be inconsistent with $\{v\} \in \tau$. To find v, consider any $A \in \tau$. Then $A \neq \emptyset$, since τ is consistent. If $|A| = 1$, we have found v. If $|A| \geqslant 2$, we partition A into two non-empty sets B and C. One of these sets must be in τ, as otherwise τ would contain the inconsistent set $\{A, \overline{B}, \overline{C}\}$. We then partition this set and proceed as we did earlier with A, until we arrive at some $\{v\} \in \tau$.

Similarly, a tangle of S_i can 'spawn' several tangles of S_k that all extend it.[49] We can thus study how tangles are 'born' at some time i (when they are one of several tangles of S_i spawned by the same tangle of S_{i-1}), then 'live on as' (extend to unique) tangles of S_k for some $k > i$, and eventually 'die' at some time $\ell > k > i$ (when they spawn several tangles of S_ℓ, or extend to none).

3.3 Stars, universes, and submodularity

A set of features is a *star* if their inverses are disjoint subsets of V. The sets $\mathcal{F}$ that we use to define tangles will often consist of stars. They also arise naturally as in the following example.

Consider a survey S whose questions allow answers on a scale from 1 to 5. To compute its tangles, we first need to convert S into a hypothetical questionnaire S' of yes/no questions that is equivalent to S in the sense that its answers can be computed from the (known) answers recorded for S, and whose (computed) yes/no answers conversely imply the original answers to S. For each $s \in S$ we let S' contain five questions s_i (where $i = 1, \ldots, 5$) asking whether the given answer to s is the ith of the five potential answers, yes or no. Let $s_i = \{A_i, \overline{A}_i\}$, where $\vec{s_i} = A_i$ consists of the yes-answers. These sets A_i are disjoint, because every $v \in V$ had to choose only one of the five possible answers to s, which makes him an element of only one A_i. So the $\overline{A}_i$ have disjoints complements, which means that $\{\vec{s}_1, \ldots, \vec{s}_5\}$ is a star.

We shall be particularly interested in sets S such that $\vec{S}$ contains, for any given $\vec{r}, \vec{s} \in \vec{S}$, also the features $\vec{r} \vee \vec{s}$ and $\vec{r} \wedge \vec{s}$. If S has this property, then $\vec{S}$ will, in fact, contain *all* Boolean expressions of features it contains already:[50] combinations of features built from others by using the symbols $\vee$ and $\wedge$ and taking inverses.[51] We then call $\vec{S}$ a *universe* of features.

If we are given a set S of potential features that lacks this property – for example, when S is given as a questionnaire[52] – we can extend S by adding to $\vec{S}$ any missing features and their inverses until it does have this property. The set S' of potential features generated from S in this way will be much larger than S and may well contain all bipartitions of V. We call $\vec{S'}$ the feature universe *generated by* S.

Finally, S and $\vec{S}$ are *submodular* if for any two features $\vec{r}, \vec{s} \in \vec{S}$ at least one of $\vec{r} \vee \vec{s}$ and $\vec{r} \wedge \vec{s}$ is also in $\vec{S}$. Submodularity is a technical 'richness'

---

[49] For example, if a tangle of order k identifies the chairs in our set V of furniture, we might find that S_{k+1} has three tangles that induce this chair tangle of order k: one for armchairs, another for dining chairs, and perhaps a third for garden chairs.

[50] Recall that $\vec{S}$, the set of all specifications of elements of S, already contains the inverses of all its elements.

[51] For example, $\overleftarrow{r} \wedge (\vec{r} \vee \vec{s})$ and its inverse are such 'Boolean expressions'.

[52] From now on, we shall assume in all examples where S 'is a questionnaire' that the people in V have already answered all its questions. Only then do the elements of S become potential features in our new formal sense, i.e., partitions of V: these partitions are, for every 'question' s, given as the sets A and B of the people in V that answered s as yes or as no, respectively, so that s becomes the partition $\{A, B\}$.

condition on S that is needed in the proofs of our two tangle theorems, and which helps with our tangle algorithms. If S is not submodular, we can make it so by adding the required potential features, as earlier when we generated a universe from S.[53] Any tangle of S' then also has to specify the newly added features. Tangles of sets S that are not submodular do not automatically extend to such larger tangles.

3.4 Hierarchies of features and order

In Section 3.2 we described how tangles evolve from each other when they are tangles not of the entire set S of partitions of V we are considering,[54] but of increasing subsets $S_1 \subseteq S_2 \subseteq \ldots$ of S. Whether or not this formal evolution of tangles corresponds to any meaningful refinement of typical features, however, will depend on how we set up this stratification $S_1 \subseteq S_2 \subseteq \ldots$ of S.

Technically, we shall do this as follows. To every $s \in S$ we shall assign an integer, to be called the *order* of s (and of $\vec{s}$ and of $\overleftarrow{s}$) and denoted as $|s|$, so that[55]

$$S_k = \{\, s \in S : |s| < k \,\}.$$

Informally, the tangles of S_k are also often said to have 'order k'.

The function $s \mapsto |s|$ is called an *order function*. We sometimes require it to be *submodular* on $\vec{S}$, which means that $|\vec{r} \vee \vec{s}| + |\vec{r} \wedge \vec{s}| \leqslant |\vec{r}| + |\vec{s}|$ for all $\vec{r}, \vec{s} \in \vec{S}$. Note that this implies that the sets S_k are submodular as defined in Section 3.3.

Ideally, then, the typical features encoded in a tangle of any given order should 'refine' those of tangles of lower order in some natural sense given by the application at hand. We might hope to achieve this by choosing for S_1 only the most basic potential features of the objects in V, for S_2 some slightly less basic features, and so on until the S_k for large k consist of quite specific potential features that are only needed to distinguish objects not distinguished by any of the more basic features.[56] In a questionnaire environment, S_1 would consist of the most basic questions in the survey, and as k increases, the questions become more and more specific.

In a more subtle approach, we might defer the choice of which potential features to include in S_{k+1} until we have computed the tangles of S_k.[57] In

---

[53] However, the submodular set $S' \supseteq S$ is often smaller than the universe generated by S.

[54] This S may be the set of all partitions of V, or just of some that are of particular interest.

[55] Thus, when we choose S_1 we assign its elements the order 0. When we increase S_1 to S_2, we assign the new elements – those in $S_2 \smallsetminus S_1$ – the order 1. And so on, as $k = 1, 2, \ldots$.

[56] For example, the potential feature of being collapsible or not may serve to distinguish some garden or dining chairs from others. It might then split the corresponding tangles in two, giving rise to a tangle of camping chairs and another of rigid garden chairs, and similarly for the dining chairs. The armchair tangle would be unaffected by this additional potential feature, since armchairs are all non-collapsible, and so the feature of not being collapsible will simply integrate into the existing armchair tangle.

[57] Our standard algorithms do indeed compute the tangles of S_k successively as k increases.

our questionnaire scenario, for example, the question of whether someone likes Tchaikovsky may be relevant when we are looking for how to split a tangle of love for classical music, but less relevant when we analyse a tangle of sport enthusiasm. With this approach, there will be several sets S_{k+1}, one for every choice of a k-tangle (which in turn may be tangles of different sets S_k), and the same feature may be given different orders depending on the choice of the tangle relative to which this order is assigned.[58]

While defining the order of potential features explicitly appears to be the most natural option, there are a few issues we should be aware of. One is that sometimes we simply do not know which potantial features are 'basic' and which are not. This may be because our set S is not a hand-designed survey but a collection of technical properties of the objects in V resulting from some measurements whose significance we cannot judge.

A more subtle issue is that by declaring some potential features as more fundamental than others we are in danger of influencing the outcome of our study more than we would like to. Indeed, which tangles arise in the S_k has everything to do with how these S_k are chosen. If we are searching for hitherto unknown types of features, we should keep our influence on how to group them to a minimum, unless we know what we are doing and do it deliberately.

Finally, there is a technical problem with choosing the S_k explicitly. This is that we would like them to be submodular (see Section 3.3): for any $\vec{r}, \vec{s} \in \vec{S_k}$ we would like at least one of $\vec{r} \vee \vec{s}$ and $\vec{r} \wedge \vec{s}$ to be in $\vec{S_k}$ too. We might simply add one of them to $\vec{S_k}$ to achieve this, but this can cause other problems. One is that $\vec{r} \vee \vec{s}$ and $\vec{r} \wedge \vec{s}$, while being well-defined partitions of V, need not lie in S and therefore may lack the 'meaning' that comes with the element of S, making it difficult to decide which of them is the appropriate one to add to S_k because it is more fundamental, or basic, than the other.

We shall return to this problem later in this section. Choosing the order of potential features explicitly remains a valuable aim in many scenarios, and there is a way to deal with the technical problem just described.

There are also some generic ways of defining an order function, ways that make no reference to the interpretation of the potential features but which do take into account the abstract way in which these features structure the set V, i.e., how the elements of S partition V. Moreover, just as the set S of all our given potential features can generally determine the order of a particular partition $s \in S$ (as we shall see below), it can determine the order of an arbitrary partition $\{A, B\}$ of V.

In the rest of this section we shall look at an example of such a generic way of assigning an order $|\{A, B\}|$ (which we shall abbreviate to $|A, B|$) to all the partitions of V by appealing to S in one way or another. It is only

---

[58] Our question about Tchaikovsky, for example, might receive order k when we consider the k-tangle of classical music enthusiasm, while also receiving order $k + 5$ relative to some $(k + 5)$-tangle extending the k-tangle for sports, such as a tangle for the love of ballet.

an example intended to demonstrate the idea; another example is given in [2]. On the whole, there is scope for some substantial and interesting mathematics here, some of which is explored in [5].

To understand better our task of assigning an order $|A, B|$ to an arbitrary partition $\{A, B\}$ of V, let us look at cases where we already have an aim for how to do this: when $\{A, B\}$ is itself in S. In that case we said – appealing to an interpretation of the elements of S, which we shall now seek to avoid – that $|A, B|$ should be small if its complementary features A and B are 'basic', or 'fundamental'. Let us try to capture this notion without interpretation, but allowing ourselves to refer to how the other elements of S partition V.

One obvious aspect of $s = \{A, B\}$ being 'fundamental' is that it does not divide V in such a way that large groups of 'similar' objects are split more or less evenly: that about half of them lie in A and the other half in B.[59] This can be expressed in an abstract way – with reference to S, but without interpretation – if being 'similar' can be. And indeed it can: we may think of two objects in V as similar if they share many of the features in S, if for many $s \in S$ they both specify s as $\vec{s}$ or they both specify it as $\overleftarrow{s}$.

So let us define a similarity function $\sigma\colon (u, v) \mapsto \sigma(u, v) \in \mathbb{N}$ on $V \times V$, the set of pairs of objects, by counting common features in this way, letting

$$\sigma(u, v) := \big|\{\, s \in S : u(s) = v(s) \,\}\big|,$$

where

$$v(s) := \begin{cases} \vec{s} & \text{if } v \text{ specifies } s \text{ as } \vec{s} \\ \overleftarrow{s} & \text{if } v \text{ specifies } s \text{ as } \overleftarrow{s} \end{cases}$$

and $u(s)$ is defined analogously. Thus, u and v have large *similarity* $\sigma(u, v)$ if they share many of the features in $\vec{S}$.

Note that there is ample scope to adjust this definition of a similarity function σ to a concrete situation. One natural adjustment can be made if the potential features $s \in S$ have default specifications $\vec{s}$ that are more distinguishing than their inverses $\overleftarrow{s}$. In our furniture example, $\vec{s}$ might denote the feature of being collapsible, which is a lot more distinguishing than not being collapsible. In this case we might want to count only these distinguishing features in the definition of σ, adjusting it to

$$\sigma(u, v) := \big|\{\, s \in S : u(s) = v(s) = \vec{s} \,\}\big|,$$

reflecting the fact that we consider two collapsible pieces of furniture as more like one another than two arbitrary non-collapsible pieces of furniture.

---

[59] For example, the feature of having (long) legs is fundamental in this sense for furniture: most chairs have legs, most tables have legs, and most beds do not have legs. The feature of being made almost entirely of wood is not fundamental in this sense: there are about as many chairs and tables that have this feature as there are chairs and tables that do not have it.

Based on such a similarity measure on V, we can now define an order function on the partitions of V that assigns low order to partitions that do not split many pairs of similar objects:

$$|A, B| := \sum_{a \in A} \sum_{b \in B} \sigma(a, b).$$

Note that, as desired, $|A, B|$ is large if there are *many* pairs (a, b) with *high* similarity $\sigma(a, b)$, and small otherwise. It is not hard to show that this order function is submodular.[60]

Although order functions as above can be defined on all partitions of V, they can equally well be used also when we quite deliberately restrict our attention to the original partitions of V from S – for example, because they come with 'meaning' which we still need in order to interpret our tangles once they are found.[61]

If desired, we can even tweak this method to mimick explicit definitions of order functions on S in a way that circumvents the problem with submodularity discussed earlier. We wish to assign low order to important features. Instead of doing this explicitly, we can give important features greater weight in the definition of our similarity function σ.

To do this formally, we choose a *weight function* $w: S \to \mathbb{N}$ that assigns large values to important potential features. Our first definition of σ then turns into

$$\sigma_w(u, v) := \sum \{ w(s) : s \in S, \ u(s) = v(s) \},$$

while our second definition turns into

$$\sigma_w(u, v) := \sum \{ w(s) : s \in S, \ u(s) = v(s) = \vec{s} \}.$$

We leave the definition of $|A, B|$ as $\sum_{a \in A} \sum_{b \in B} \sigma_w(a, b)$ unchanged.

Note that increasing the weight of a potential feature $s = \{A, B\}$ does not affect the value of $|A, B|$ itself, since $w(s)$ does not occur in the sum $\sigma_w(a, b)$ for any $a \in A$ and $b \in B$, because $a(s) \neq b(s)$ for such a and b. But raising the weight of s increases the order of other $r \in S$: of every $r \in S \setminus \{s, \{\emptyset, V\}\}$ in the case of the first definition of σ_w, and of every $r \in S$ that splits the set $\vec{s} \subseteq V$ in the case of the second definition of σ_w.[62] Hence if we increase $w(s)$, the weight

---

[60] Here is a sketch of the easy proof. Given $\vec{r} = A$ and $\vec{s} = C$, the sum $\sigma(A)$ of all $\sigma(u, v)$ with $u \in A$ and $v \in \overline{A}$ is $|r|$, and the analogously defined sum $\sigma(C)$ is $|s|$. Now computing these sums for $\vec{r} \vee \vec{s} = A \cup C$ and $\vec{r} \wedge \vec{s} = A \cap C$, we notice that the terms $\sigma(u, v)$ that count towards $\sigma(A \cup C) + \sigma(A \cap C)$ also count towards $\sigma(A) + \sigma(C)$, and equally often. Thus, $|\vec{r} \vee \vec{s}| + |\vec{r} \wedge \vec{s}| = \sigma(A \cup C) + \sigma(A \cap C) \leqslant \sigma(A) + \sigma(C) = |r| + |s|$ as desired.

[61] Tangle-based teaching methods, discussed in Chapters 2.4 and 5.4, are a good example.

[62] Indeed, let $r = \{C, D\}$. As $s \neq r$, there exist $c \in C$ and $d \in D$ that both lie in A or both lie in B. Hence $c(s) = d(s)$, and so $w(s)$ occurs in the sum of $\sigma_w(c, d)$ and therefore in the sum of $|C, D|$.

of s, the order of r will rise: a little bit if r and s are similar as partitions of V, and a lot if the they are very different. Thus, indirectly, assigning s large weight amounts to assigning it, and elements of S similar to it, low order relative to the order of the other elements of S.

However, there is an important difference to choosing orders explicitly: order functions based on a weight function $s \mapsto w(s)$ in this way are defined on all partitions $\{A, B\}$ of V, not just on S itself. In particular, they are defined on all Boolean expressions of features in S.[63] Moreover, these order functions are submodular, with the same easy proof as in the unweighted case.

Finally, tangles of sets S_k defined with respect to such an order function are usually witnessed by a weight function on V (see Chapter 1.5). Indeed, since our order function is submodular, we can extend S_k to a submodular set S'_k of partitions of V which still all have order $< k$. Then all tangles of S'_k are witnessed by a weight function on V, see [12]. Hence any tangle of S_k that extends to a tangle of S'_k will also be witnessed by the weight function on V for that tangle of S'_k, since $S_k \subseteq S'_k$, and most tangles of a set of partitions of V do extend to tangles of the submodular extensions of that set.

We thus have an indirect way of assigning the potential features in S low or high order, explicitly if indirectly, which avoids the technical problems this would cause if we did it directly. When we later say in our discussion of applications that we might choose the order of some features explicitly in one way or another, then doing it in this indirect way will often be what is meant.

Tangles based on the above type of order function have been used, for example, to classify texts by topic. Even if S is a set of very simple potential features of texts describing such basic properties as the frequency of words, one can set the parameters so that the tangles found correspond to the topics that these texts are about. See [16] for more.

3.5 Duality of set separations

In Chapter 2.6 we saw an intriguing example of a pair V, S of sets each of which could be regarded as a set of potential features of the other. Each of them had its own tangles, and these tangles described different aspects of the relationship between V and S. Now that we have defined tangles and their ingredients more formally, we can see easily that such a 'duality' between V and S exists always.

This means that whenever we encounter a natural setting for studying tangles, there is always also a dual setting that may have tangles. These may shed additional light on the situation we are studying, and may be worth exploring even when this was not initially envisaged.

In this spirit, this section is intended to throw an additional light on the applications of tangles we describe in this paper (all of them), but will not be formally needed for the rest of this paper except in Chapters 4.4 and 5.2.

---

[63] We shall need these for our tangle algorithms later.

Consider two disjoint sets, X and Y. Assume that some elements x of X are in a 'special relationship' with some elements y of Y. We can express this formally by way of a set E whose elements are sets $\{x, y\}$ with $x \in X$ and $y \in Y$: not all of them, but some.[64]

Our choice of E defines for every $x \in X$ a subset $\bar{x}$ of Y consisting of just those y with whom x has this special relationship. Similarly, for every $y \in Y$ we have the set $\bar{y}$ of all $x \in X$ with whom y has this special relationship:

$$\bar{x} := \{\, y \in Y \mid \{x, y\} \in E \,\} \quad \text{and} \quad \bar{y} := \{\, x \in X \mid \{x, y\} \in E \,\}.$$

Then for all $x \in X$ and $y \in Y$ we have

$$x \in \bar{y} \iff y \in \bar{x} \qquad\qquad (*)$$

as both these are true if $\{x, y\} \in E$ and false otherwise.

Let us define $\bar{\bar{x}}$ as the complement of the set $\bar{x}$ in Y, and $\bar{\bar{y}}$ as the complement of the set $\bar{y}$ in X. Further, let

$$\vec{X} := \{\, \bar{x} \mid x \in X \,\} \cup \{\, \bar{\bar{x}} \mid x \in X \,\}$$
$$\vec{Y} := \{\, \bar{y} \mid y \in Y \,\} \cup \{\, \bar{\bar{y}} \mid y \in Y \,\}.$$

Then $\vec{X}$ becomes a set of features of Y, and $\vec{Y}$ is a set of features of X, for the partitions $\{\bar{x}, \bar{\bar{x}}\}$ of Y and $\{\bar{y}, \bar{\bar{y}}\}$ of X as potential features. We say that these feature sets are *dual* to each other.[65]

Note that every x determines the set $\{\bar{x}, \bar{\bar{x}}\}$ and vice versa, and similarly every y determines the set $\{\bar{y}, \bar{\bar{y}}\}$ and vice versa. We may therefore choose to ignore the formal difference between x and $\{\bar{x}, \bar{\bar{x}}\}$, and between y and $\{\bar{y}, \bar{\bar{y}}\}$. Then X itself becomes a set of potential features of (elements of) Y, and Y a set of potential features of (elements of) X.

Let us now see how in general, given any set S of partitions of a set V, we can define a set $\vec{V}$ of features of S that is dual in this sense to the set $\vec{S}$ of features of V. We start by picking for every $s \in S$ a default specification, which we denote as $\vec{s}$ (rather than $\bar{s}$). This determines for every $v \in V$ the sets

$$\vec{v} = \{\, s \in S \mid v \in \vec{s} \,\} \quad \text{and} \quad \bar{v} = \{\, s \in S \mid v \in \bar{s} \,\},$$

which form a partition of S. If, as we shall assume, these partitions $\{\vec{v}, \bar{v}\}$ differ for distinct $v \in V$, they determine their v uniquely and we may think of each v as shorthand for $\{\vec{v}, \bar{v}\}$. This makes V into a set of partitions $v = \{\vec{v}, \bar{v}\}$ of S and

$$\vec{V} := \{\, \vec{v} \mid v \in V \,\} \cup \{\, \bar{v} \mid v \in V \,\}$$

---

[64] For mathematicians: E is the edge set of a bipartite graph with vertex classes X and Y.

[65] For mathematicians: this is well defined, stated symmetrically as above, since everything we have said has been symmetrical in X and Y.

into the set of all specifications of elements of V. Thus $\vec{V}$ and $\vec{S}$ form an instance of a pair $\vec{X}, \vec{Y}$ of dual feature sets as defined earlier.[66]

With this setup in place, we can now investigate how the tangles of S are related to those of V, how they or their distinguishers T compare with subsets of S that witness tangles of V, and so on. But this is not the place to do this; readers interesting in these connections may find some more inspiration in [1].

However we have seen enough now of this duality to apply it in practice: whenever we have a set V of 'objects' at hand, and a set $\vec{S}$ of features of these objects, then together with investigating the tangles of S we can also investigate the tangles of V viewed as a set of potential features of the elements of S. Both together are likely to paint a more comprehensive picture of the situation we are trying to analyse than just working with one of the two types of tangle.

4. Tangle theorems and algorithms

In this chapter we present the two most fundamental theorems about tangles, the *tree-of-tangles theorem* and the *tangle-tree duality theorem*, in their simplest forms. Yet the versions presented here are strong enough to apply to all the application scenarios we have discussed as examples, and are likely to apply equally to most other application scenarios that arise in the empirical sciences.

The tree-of-tangles theorem, of which we present two versions in Section 4.1, finds in a submodular set S of potential features a small subset T that suffices to distinguish all the tangles of S: for every two tangles of S there is a potential feature in T, not only in S, which these two tangles specify differently.

In Section 4.2 we present the tangle-tree duality theorem. If S has no tangle, this theorem finds a small subset T of S that suffices to demonstrate this fact conclusively. We say that T demonstrably 'precludes' the existence of a tangle. This can help to prove that some given data is, in a precise and quantitative sense, unstructured or contaminated.

In Section 4.3 we look at some generic algorithms that compute tangles, tangle-distinguishing features sets, and tangle-precluding feature sets.

In Section 4.4 we address the question of how the elements of V can be grouped, or clustered, 'around' the tangles of S.

In Section 4.5, finally, we discuss why the tangle-distinguishing property of the set T found by the two theorems of Section 4.1 makes it particularly valuable: we can predict from how a given $v \in V$ specifies the potential features in T how it will probably specify most of the other $s \in S$.

The following definitions will be needed throughout this chapter. A *feature system* is a set $\vec{S}$ obtained from a set S of partitions of a set V: it consists of all the specifications $\vec{s}$ and $\overleftarrow{s}$ of elements s of S. A subset T of S is a *tree*

---

[66] The 'special relationship' exists between elements $v \in V$ and $s \in S$ for which $v \in \overleftarrow{s}$ and, equivalently, $s \in \vec{v}$.

set[67] if $\{\emptyset, V\} \notin T$ and for every two partitions in T one side of each partition is contained in a side of the other partition.[68]

While S can be exponentially large in terms of $|V|$, tree sets in S are much smaller, at most linearly large in terms of $|V|$. The tangle-distinguishing tree sets T we find in Section 4.1 will be much smaller still, no larger than the number of tangles.

4.1 Tangle-distinguishing feature sets

Consider two different tangles of S. As both are specifications of S, the fact that they are different means that there exists an $s \in S$ which one of our two tangles specifies as $\vec{s}$, and the other as $\overleftarrow{s}$. We say that s *distinguishes* the two tangles. More generally, we say that a subset T of S *distinguishes all the tangles* of S if for every two tangles of S we can find in this set T an s that distinguishes them.

Our first theorem says that, if S is submodular,[69] we can find such a tangle-distinguishing subset T of S that is a tree set:

Theorem 1. (Tree-of-tangles theorem for fixed S [6])
For every submodular feature system $\vec{S}$ there is a tree set $T \subseteq S$ that distinguishes all the tangles of S.

Viewed in terms of V, the elements of T are nested partitions of V. If T is chosen minimal, these partitions of V cut it up into disjoint parts which correspond bijectively to the tangles of S. Indeed, every tangle τ of S induces a tangle of T, the tangle $\tau \cap \vec{T}$, and the sets

$$V_\tau := \bigcap \{ A \mid A \in \tau \cap \vec{T} \}$$

partition V as τ varies over the tangles of S. Each of the non-empty parts V_τ into which T cuts the set V is thus 'home' to exactly one tangle τ, but note that V_τ can be empty for some τ.

An important theoretical consequence of Theorem 1 is that S has only few tangles, even of the most general kind (with $\mathcal{F} = \emptyset$). Indeed, as T distinguishes all the tangles of S, these induce distinct tangles of T. Now a tree set of size ℓ has at most $\ell + 1$ tangles,[70] and all tree sets in S are small.[71]

---

[67] See [4] for what the structure of such sets T has to do with trees.

[68] Thus, for all $\{A, \overline{A}\}, \{B, \overline{B}\} \in T$ we have $A \subseteq B$ or $A \subseteq \overline{B}$ or $\overline{A} \subseteq B$ or $\overline{A} \subseteq \overline{B}$, and correspondingly $\overline{B} \subseteq \overline{A}$ or $B \subseteq \overline{A}$ or $\overline{B} \subseteq A$ or $B \subseteq A$. Such sets A, B are also called *nested*.

[69] If it is not, we may wish to make it submodular by adding some elements obtained as Boolean expressions of existing elements, as outlined in Chapter 3.3.

[70] For graph theoristis: it has exactly $\ell + 1$ consistent specifications, all of which are tangles of the most general kind. This is because the consistent specifications of a tree set T correspond to the nodes of a tree with $|T|$ edges [4], and a tree with ℓ edges has exactly $\ell + 1$ nodes.

[71] They have size less than $2|V|$, independently of the size of S. So the number of all tangles of S is at most $|T| + 1 \leqslant 2|V|$. By contrast, S itself can have size exponential in $|V|$, and the number of specifications of S is exponential even in $|S|$.

In practice, the implication also goes the other way. We can control directly about how many tangles we get by specifying $\mathcal{F}$, e.g., by choosing the parameter n for $\mathcal{F} = \mathcal{F}_n$. Since T, if chosen minimal, is automatically smaller than the number of tangles of S, it will then be small as a consequence.

This aspect is particularly valuable if S is not submodular (as Theorem 1 requires). To distinguish ℓ tangles we only need $\ell - 1$ elements of S, and these are quick to find. However, they may not form a tree set. Our algorithm will then build from these elements of S a tree set T, also of size $\ell - 1$, which once more distinguishes all our ℓ tangles of S. This T will not be a subset of S. But the elements of $\vec{T}$ will be Boolean expressions of specifications of those $\ell - 1$ elements of S we started with, obtained by iteratively adding features such as $\vec{r} \vee \vec{s}$ or $\vec{r} \wedge \vec{s}$ combined from existing or previously added features $\vec{r}$ and $\vec{s}$. So this T will be small whichever way we choose to measure it: in absolute terms, as $|T|$, or in terms of how many features from $\vec{S}$ are needed to build it.

There is no general rule of how many tangles to aim for when choosing $\mathcal{F}_n$: this will depend on the application. Choosing n large will give us only the most pronounced tangles, while choosing it lower will give us more refined tangles.[72] But if S satisfies a little more than submodularity, we can dramatically improve Theorem 1 to find a tree set in S that distinguishes even tangles of varying degrees of refinement, all at once. This will be our next theorem.

Assume now that $\vec{S}$ is not just a submodular feature system but even a universe of features of V with a submodular order function $s \mapsto |s|$. Any tangle of some $S_k = \{\, s \in S : |s| < k \,\}$ will be called a tangle *in* S. Two tangles in S, not necessarily of the same order, are *distinguishable* if there exists an $s \in S$ which they specify differently.[73] Such an s *distinguishes* the two tangles, and it does so *efficiently* if no $r \in S$ of lower order distinguishes them. A set $T \subseteq S$ *distinguishes all the tangles in S efficiently* if for every two distinguishable tangles in S there is an $s \in T$ that distinguishes them efficiently.

Theorem 2. (Tree-of-tangles theorem for tangles of variable order [7])
For every feature universe $\vec{S}$ with a submodular order function there is a tree set $T \subseteq S$ that distinguishes all the tangles in S efficiently.

See [7, 13] for more general versions of Theorem 2.

As earlier, every tangle τ in S induces a tangle $\tau \cap \vec{T}$. If T is chosen minimal then the sets

$$V_\tau := \bigcap \{\, A \mid A \in \tau \cap \vec{T} \,\}$$

---

[72] This a different kind of refinement than in the hierarchy of tangles discussed earlier.

[73] Such an s will always exist for distinct tangles of the same order, i.e., for distinct tangles of the same set S_k. But a tangle of order k is not distinguishable from the tangles of order $\ell < k$ that it induces on the sets S_ℓ, as defined in Section 3.2. Our assumption that s distinguishes two tangles, of orders k and ℓ, say, thus means in particular that $|s| < \min\{k, \ell\}$, and it implies that the tangle of higher order does not induce the tangle of lower order.

partition V as τ varies over the maximal[74] tangles in S.

The relative structure between the tree set $T \subseteq S$ in Theorem 1 or 2 and the tangles it distinguishes can be made visible by a graph-theoretic tree $\mathcal{T}$ whose edges correspond to the elements of T and whose nodes correspond to the (maximal) tangles τ of or in S.[75] In the case of Theorem 2, the tree $\mathcal{T}$ also displays the non-maximal tangles in S and their relative structure (such as which refines which): not by its nodes, but by some of its subtrees.[76]

4.2 Tangle-precluding feature sets

The tree-of-tangles theorems from Section 4.1 offer a geometric picture of how the various (maximal) tangles of S are organized in a tree-like way. Viewed in terms of the set V, they produce a tree set T of partitions which cut V up into disjoint parts that correspond exactly to the (maximal) tangles of S.

Our next theorem, the tangle-tree duality theorem, produces a similar tree set $T \subseteq S$ for the case that S has no tangles – understood to be $\mathcal{F}_n$-tangles for some n to be specified, as before.[77] Once more, T partitions V into disjoint parts. This time these parts are not 'home to a tangle', but the opposite: they are so small that they *cannot* be 'home to a tangle'. Indeed if 'tangle' means $\mathcal{F}_n$-tangle, they have size $< n$. Let us make this more precise.

Consider a consistent specification τ of a tree set $T \subseteq S$. Let $\sigma \subseteq \tau$ be the set of those elements $\vec{t}$ of τ for which $\vec{t} = C$ is minimal under inclusion. One can show that σ is a star of features (see Chapter 3.3). These sets σ, one for every consistent specification of T, are the *splitting stars* of T.[78] A tree set $T \subseteq S$ is *over* $\mathcal{F}$ if all its splitting stars are elements of $\mathcal{F}$.

Theorem 3. (Tangle-tree duality theorem [6])
Let $\vec{S}$ be a submodular feature system, and let $\mathcal{F} = \mathcal{F}_n$ for some integer $n \geqslant 1$. Then exactly one of the following two statements holds:

(i) S has an $\mathcal{F}$-tangle.

(ii) S contains a tree set over $\mathcal{F}$.

See [6, 8, 9] for more general versions of Theorem 3.

---

[74] Maximal in terms of inclusion as subsets of $\vec{S}$. The maximal tangles in S are those not induced by any other tangle of larger order.

[75] If $V_\tau \neq \emptyset$, then τ orients all the edges of $\mathcal{T}$ towards 'its' node by specifying their corresponding $t = \{A, B\} \in T$ as the A for which $V_\tau \subseteq A$.

[76] Given k, the tangles of S_k correspond to the nodes of the tree $\mathcal{T}_k$ for the tree set $T \cap S_k$, which in turn correspond to the components of the forest obtained from $\mathcal{T}$ by deleting the edges that correspond to elements of T in S_k. These subtrees are nested in a way that reflects the nestedness of the tangles in S as subsets. See [4, 6, 7].

[77] See Chapter 3.1 for the definition of $\mathcal{F}_n$-tangles.

[78] As remarked earlier, the 'tree-shape' of a tree set can be made visible by a graph-theoretic tree whose edges correspond to the elements of T and whose nodes correspond to the consistent specifications of T. The splitting stars of T correspond to the stars at these nodes [4].

It is easy to see why (i) and (ii) cannot happen at once, why the existence of a tree set over $\mathcal{F}$ precludes the existence of an $\mathcal{F}$-tangle.[79] The harder part of the theorem is that at least one of (i) and (ii) must always happen: that if S has no $\mathcal{F}$-tangle then it contains a tree set as in (ii), one which demonstrably precludes the existence of an $\mathcal{F}$-tangle (as just noted).

This is valuable, and it is the way in which Theorem 3 is usually applied. For example, suppose that some tangle-finding algorithm fails to find a tangle. This does not mean, yet, that no tangle exists. But if the algorithm also finds a tree set as in (ii) of Theorem 3, we can be sure that indeed no tangle exists.[80] The theorem thus assures us that whenever no tangle exists there is conclusive and easily checkable evidence for this.

4.3 Algorithms

There are a number of algorithmic aspects concerning tangles. These include the computation of the following objects:

- the set S, or sets S_k, in whose tangles we are interested;
- the tangles of S or of the S_k;
- if tangles are found, a set tree T of potential features distinguishing them (as in Theorems 1 or 2, respectively);
- if no tangles are found, a tree set T of potential features that certifies their non-existence (as in Theorem 3).

In applications where S is given directly by a questionnaire, the first task does not arise. It may happen that, in the course of computing the tangle-distinguishing feature set T for the third task, the set $\vec{S}$ is enlarged by newly added features $\vec{r} \wedge \vec{s}$ or $\vec{r} \vee \vec{s}$ for already existing features $\vec{r}$ and $\vec{s}$, but any such enlargement of S will happen only then. In particular, we do not need to generate a feature universe from $\vec{S}$ before we look for or process its tangles, or even to extend $\vec{S}$ to a submodular feature system to which our tangle theorems can be formally applied.

In other applications,[81] where we consider as S some submodular extension of a questionnaire, or where we take S to be the entire set of partitions of a set V but only look for tangles of its subsets $S_1 \subseteq S_2 \ldots$ as in Chapter 3.4, we do have to compute S or these sets S_k.

When we extend a given feature system $\vec{S'}$ that is not submodular to a submodular system $\vec{S} \supseteq \vec{S'}$ to which we wish to apply our theorems (which assume submodularity for S or the order function that determines the S_k), it

---

[79] Indeed, suppose there are both a tangle τ as in (i) and a tree set T as in (ii). As $T \subseteq S$, the tangle τ induces a consistent specification (indeed a tangle) of T. As T is over $\mathcal{F}$, the splitting star of this specification of T is an element of $\mathcal{F}$. Since it is also a subset of τ, this contradicts the fact that τ is an $\mathcal{F}$-tangle.

[80] It is easy to check that a given tree set that is claimed to be over $\mathcal{F}$ is indeed over $\mathcal{F}$: we just have to compute its splitting stars, which is easy, and check that they are all in $\mathcal{F}$.

[81] including those where we start from a questionnaire but do not take this to be our set S

can happen in rare cases that some tangles of S', those we are interested in, fail to extend to tangles of S. In such cases the tangle-distinguishing set T found for S in the third task will distinguish most of the tangles of S', those that extend to tangles of S, but maybe not all of them.

In the case of the fourth task the tangle-tree duality theorem applied to S might, theoretically, produce a tree set T certifying that S has no tangle even when S' did have tangles but these did not extend to tangles of S. In order to avoid this, we always try to compute the tangles of the feature system S' we are interested in directly, even if S' is not submodular. If no tangle is found, we then run the algorithm for the fourth task on S'. This will usually produce a tree set $T \subseteq S'$ certifying that S' has no tangle. If it does not, the algorithm will output a list of features that must be included in any tangle of S'. If S' is submodular then such a tangle will in fact exist. But even if S' is not submodular, as in our hypothetial case here, this list will be a good start to restart our search for a tangle of S', which in most cases is also found.

Although the S_k may not be large for the small values of k relevant to us, determining them precisely can take longer than is practicable. For example, determining the order of every partition of a (large) set V takes time exponential in $|V|$, since there are $\binom{|V|}{2}$ such partitions. But often there are known heuristics for the application in question that can reduce the set of relevant partitions, or the low-order partitions can be computed by known methods.[82] If that is not the case, one can use standard approximation methods, such as finding local minima close to random partitions, to find not all but enough features of order $< k$ for each relevant k to compute large enough subsets of the S_k to make their tangles significant.

Once we have decided for which S or S_k we wish to compute its $\mathcal{F}$-tangles, we can do this fairly quickly in an ad-hoc way: we go through its elements s one by one and add $\vec{s}$ or $\overleftarrow{s}$ to every tangle previously found for the subset of S already considered, as long as this does not create a forbidden subset. But checking whether a given subset of $\vec{S}$ is forbidden, i.e., in $\mathcal{F}$, is usually quick.[83]

Computing a tangle-distinguishing feature set T in our third task is more subtle. Grohe and Schweitzer [15] have turned the proof of Theorem 2 given in [7] into an algorithm with polynomial running time, which however is not intended for practical applications.[84] Elbracht, Kneip and Teegen [11] have

---

[82] For example, if our application is to capture fuzzy but discernable regions in a picture, we might run a standard clustering algorithm to find a much smaller set S of partitions of the pixel set that correspond to natural lines and do not cut right through a big cluster. If the order of our partitions of V is determined by a similarity measure σ or σ_w, as in Chapter 3.4, we can use network flow theory to find the partitions of low order quickly.

[83] For example, for $\mathcal{F} = \mathcal{F}_n$ as in Chapter 3.1 we just have to count for every two elements A and B in a partial tangle already computed how many elements of V lie in $A \cap B \cap C$ or in $A \cap B \cap \overline{C}$, where $s = \{C, \overline{C}\}$ is the new element of S under consideration.

[84] However it is of theoretical significance for computational isomorphism problems. This takes us too far afield, but any reader interested is encouraged to consult [7] and [15].

described a different algorithm, also intricate, which has an exponential worst-case running time but has shown excellent performance in practice.[85] Even if S is not submodular, it uses only $\ell - 1$ elements of S in total to build a tree set T of size $\ell - 1$ to distinguish ℓ tangles.

If we do not require our tangle-distinguisher T to be a tree set but just want it to be small, it can be found inside S and still of size $|T| < \ell$. The fact that T is a tree set, however, is crucial to a number of tangle applications. One consequence is that its tangles extend (uniquely) to tangles of S. Another is that it enables us to associate elements of V with tangles of S; see Section 4.4 below. Both these properties together form the basis for tangle-based predictions, discussed in Section 4.5.

Computing a tangle-precluding feature set T as in our fourth task may or may not be fast,[86] depending on the application; see [11] for details.

4.4 Tangle-based clustering

As we have seen, tangles of a set S of features of a set V of objects, including tangles of views prevalent in a population of people, are best at discovering what *kinds* of groupings exist in V, rather than at dividing V into such groups. However, once found, tangles of S can help with this too.

Before we explore this further, let us briefly look at two other ways of defining clusters in V that are also determined by tangles, but which we shall not pursue further here. The first of these are witnessing sets of tangles, as mentioned briefly in Chapter 1.5. These may indeed be interesting in practice, and they helped motivate tangles when they were invented. The question of when tangles have witnessing sets has not been studied sufficiently so far. But most of the tangles we have studied here do have witnessing functions [12].

Another approach we shall not pursue in detail is duality: to obtain clusters in V from the tangles of V (rather than of S) that arise when we interpret V as a set of potential features of the elements of S, as explained in Chapter 3.5. Informally, just as tangles of S are collections of features shared by many of the objects in V, tangles of V are sets of points sharing many of the features in $\vec{S}$, which on the face of it looks exactly what we would expect of a cluster.

More formally, a specification σ of V is an $\mathcal{F}_n$-tangle of V if for all triples $\vec{x}, \vec{y}, \vec{z} \in \sigma$ we have $|\vec{x} \cap \vec{y} \cap \vec{z}| \geqslant n$. By definition of $\vec{V}$ (see Chapter 3.5) this means that for all triples $v_1, v_2, v_3 \in V$ there are n elements s of S such that $v_i \in \vec{s}$ if $\sigma(v_i) = \vec{v_i}$ and $v_i \in \overline{s}$ if $\sigma(v_i) = \overleftarrow{v_i}$, where $\vec{s}$ is some fixed default specification of s chosen in the process of definition the feature system $\vec{V}$.

Given such an $\mathcal{F}_n$-tangle σ of V, we can associate with it the set

$$V_\sigma^+ := \{\, v \in V \mid \sigma(v) = \vec{v} \,\}$$

---

[85] Typical running times for a few dozen tangles of a few thousand features measure in minutes or seconds on an ordinary laptop.

[86] Our current algorithm has a worst-case runnig time of $O(|S|^4)$ for submodular S.

of those $v \in V$ which σ specifies as $\vec{v}$ rather than as $\overline{v}$. This is a cluster in V in the sense described informally above: for every three points $x, y, z \in V_\sigma$ there are at least n elements s of S such that $x, y, z \in \vec{s}$, i.e., such that x, y and z share these n features $\vec{s} \in \vec{S}$. Similarly,

$$V_\sigma^- := \{\, v \in V \mid \sigma(v) = \overline{v} \,\}$$

is a set of points sharing many of the features in $\vec{S}$: every three of them share at least n of the features of the form $\overline{s}$, or equivalently, jointly lack at least n features of the form $\vec{s}$ (in our default specification of S). In fact, neither of these sets captures the full information encoded in the tangle σ of V.

More importantly, both clusters, V_σ^+ and V_σ^-, depend not only on V and S but also on the default specification of S we chose initially. These kinds of clusters in V, therefore, are significant only if this default specification of S is particularly natural – as is the case in our example of shopping tangles in Chapter 2.6,[87] but cannot be expected in general.

Let us get back to our original topic of how to define clusters in V associated with tangles of S, a set of potential features of V. A simple way to do this is to make one group for each tangle of S and associate each $v \in V$ with the tangle closest to it under the first similarity function σ defined in Chapter 3.4: the tangle τ of S for which the number of elements $s \in S$ with $\tau(s) = v(s)$ is maximum.

However, in Chapter 3.4 we also saw how this similarity function can be used to define an order function on S, indeed on the set of *all* partitions of V, to obtain a hierarchy $S_1 \subseteq S_2 \subseteq \ldots \subseteq S$ of sets of partitions of V and their associated tangles.[88] Using this hierarchy, we can do something more sophisticated to associate the elements of V with tangles in S.

For example, we could make a group in V for every tangle in S, and associate a given $v \in V$ with the tangle τ of largest order whose features are among those of v, in that $\tau(s) = v(s)$ for all $s \in S_k$ if τ has order k. Thus, we would first look which tangle τ_1 of S_1, if any, is a subset of $v(S) = \{\, v(s) \mid s \in S \,\}$; then which tangle τ_2 of S_2 refining τ_1 is still a subset of $v(S)$; and so on until we have found a tangle $\tau_k \subseteq v(S)$ such that no tangle of S_{k+1} refining τ_k is a subset of $v(S)$. Then the views of v differ from any tangle of S_{k+1} on at least one $s \in S$ not specified by τ_k, and we associate v with this τ_k.

Note, however, that this approach is not very error-tolerant: even if some v agrees on all of S with a tangle τ of highest order, except in one s of relatively low order k, then v will not be associated with τ but only with the much broader tangle $\tau_k = \tau \cap \vec{S_k}$.

---

[87] There, specifying an item s as purchased rather than as not purchased is clearly the natural choice.

[88] Recall that we use the term 'tangle *in* S' for all tangles of the sets $S_k \subseteq S$, and that the tangles of S_k are said to be tangles (in S) *of order k*.

To mend this, we might combine the two approaches above by letting our similarity function take account of all $s \in S$ but attach more importance to those of low order. For example, we could measure the similarity of two specifications f, g of S by a vector

$$\sigma(f, g) = \big(\sigma_0(f, g), \sigma_1(f, g), \dots\big)$$

where $\sigma_k(f, g)$ denotes the number of $s \in S$ of order k such that $f(s) = g(s)$, and order these vectors lexicographically.[89]

More radically, we might define the similarity of f and g as

$$\sigma(f, g) := \sum \{\, N - |s| \,:\, f(s) = g(s) \,\},$$

where N is the largest order of any $s \in S$.[90]

Such similarity functions could be used directly for the matchmaking problem discussed in Chapter 2.1, or as input for a standard distance-based clustering algorithm on V.

However, we can also use it to group V around the tangles of S, as is our aim here, as long as we say how $\sigma(v, \tau)$ should deal with the fact that if τ is a tangle of S_i, say, it does not specify the entire S. In our last definition of $\sigma(f, g)$ extended to such partial specifications f, g of S we might treat $s \in S$ that are not specified by both f and g like those s for which $f(s) \neq g(s)$, i.e., ignore them in the sum that computes the similarity between f and g. In our previous definition of $\sigma(f, g)$, the one with the lexicographic ordering of similarity vectors, we would just set $\sigma_k(v, \tau) := 0$ if τ has order $i \leqslant k$. We could then make one group in V for every tangle in S, as before, by associating a given $v \in V$ to the tangle most similar to it, the tangle τ for which $\sigma(v, \tau)$ is maximum.

Finally, we can use the tree set $T \subseteq S$ from Theorem 1 or 2 to partition V into the 'clusters'

$$V_\tau := \bigcap \{\, A \mid A \in \tau \cap \vec{T} \,\},$$

where τ varies over the tangles of S (in the case of Theorem 1) or the maximal tangles in S (in the case of Theorem 2). As remarked after these theorems, the sets V_τ do indeed form a partition of V if T is chosen minimal, which we may assume.

4.5 Predictions

Let us now discuss how knowing a tangle-distinguishing set T as in Section 4.1 enables us to make predictions. For simplicity, we shall do this for the scenario

---

[89] For example, an element $v \in V$ is more similar in this ordering to a tangle ρ of S than to a tangle τ of S if ρ and τ agree with v on equally many s of orders 0, 1 and 2, respectively but more s of order 3 satisfy $\rho(s) = v(s)$ than $\tau(s) = v(s)$.

[90] As in our first approach, this counts the $s \in S$ which f and g specify in the same way, but elements s of low order are given greater weight in this count.

of Theorem 1: we assume that we have some fixed set S of potential features of a set V, and that we have computed the tangles of S for some $\mathcal{F}$ of our choice.

To make our discussion more intuitive, let us use the language of the questionnaire example from the introduction. Thus, S is a questionnaire, and we are interested in how an unknown person might answer it. This can be formalized, if desired, by assuming that V is a sample of some larger population P of people in whose views on S we are interested. We know only the answers given by the people in V, but have reason to believe that these represent P well.[91]

Our contention is that knowing our unknown person's answers to just the questions in T enables us to predict their answers to the other questions in S with greater confidence than knowing their answers to an arbitrary subset of S of the same size as T, and certainly better than choosing their answers ($\vec{s}$ or $\overline{s}$) to the various questions s at will.[92]

Our emphasis will lie on predicting the views of our unknown person correctly with high probability for many of the questions from S: not necessarily for all of them, nor just for single questions – which is easier and does not require tangles. But note that good predictions for many single s do not necessarily add up to a good collective prediction for one person's views on many s. For example, the collection of the most popular views on single questions s may be inconsistent and thus, in its entirety, not be held by anyone.[93]

An aspect important for applications is that although we need to compute the tangles of S and the tangle-distinguishing set T in order to make such predictions, we do not need to 'understand' the tangles we have computed, to 'make sense' of them: the tangle-based process of predicting the answer to S

---

[91] This is a heuristic assumption that needs validation, but that is a routine matter which is not our concern here. Mathematically, the 'unknown person' is simply a random specification of S: a point in our probability space of all the specifications of S. The probability distribution on this space is determined roughly (but not induced exactly) by how S partitions V, in that we require it to assign small measure to specifications of S that include a forbidden set of features for one of the tangles that we observed were induced on S by V. Applying this distribution to the abstract space $\{0,1\}^S$, i.e., 'forgetting' V, allows us to reinterpret S as a set of 'potential features' also of the elements of P, even if these potential features are not known individually as partitions of P.

[92] This latter success rate would be $1/2^{|S|}$ if we assume, for want of any information to the contrary, that all yes/no answers are equally likely and independent of each other for different questions. In mathematical terms, this is the probability of a single specification of S in the space of all its specifications with the uniform distribution. Our contention is that the existence of a tangle implies that the specifications of S are not distributed uniformly, and that knowing a tangle tells us enough about the actual distribution to make our prediction better.

[93] Here is a simple example in the notation of Capter 3.1. Consider a tripartition of V into sets A_1, A_2, A_3 of equal size, and let S consist of the three partitions $s_i = \{A_i, \overline{A_i}\}$ of V ($i = 1, 2, 3$). For each i, two thirds of the people in V answered s_i as $\overline{A_i}$, so this will be our best prediction for how an unknown person would answer s_i. But these three answers together are inconsistent: as $\bigcap_i \overline{A_i} = \overline{\bigcup_i A_i} = \emptyset$, not a single person in V answered them in this way for all i, and hence this should not (and will not) be our prediction of how an unknown person would answer S.

of our unknown person can be carried out entirely mechanically on the basis of the data available to us, the way in which the people in V have answered S.

So let us look at the question of how likely it is that our unknown person answers most of S in some specified way τ. Knowing his or her answers to some of the questions does not necessarily place us in a better position to guess their remaining answers: if the events that s is specified as $\vec{s}$ or $\overleftarrow{s}$ are independent for different $s \in S$, which in the absence of any knowledge at all is a reasonable assumption, we learn nothing from such partial answers about the remaining questions.

However, if S has tangles then these events are not independent: answers that occur in the same tangle tend to be positively correlated. (This inspired our informal description of tangles in Chapter 1 as collections of features that 'often occur together'.) Therefore, taking as our guess for the answers to S of our unknown person the specification of S given by *any* tangle is already better[94] than guessing arbitrarily. But we can do better still.

For a start, let us restrict our guesses to consistent specifications of S. This is reasonable: recall that, by definition, no $v \in V$ answered S inconsistently, and so we may assume the same for our unknown person in P. If T in Theorem 1 is chosen minimal, which we may also assume, then every consistent specification of T extends to a unique tangle of S. In other words, once we know how our unknown person has answered T (consistently), there is exactly one tangle of S that makes a possible guess for all our person's other answers! This tangle's set of answers to S, then, is our prediction for how our unknown person would answer S.

To summarize, finding the tangles of S gives us a set of better guesses than random: one for each tangle. Theorem 1 then narrows this down to a unique such guess (with an even better expected success rate) once we are able to quiz our unknown person on T.

In a typical application in the social sciences, S might be a large questionnaire used in a pilot study on a small set V of people, and T might be the main study run on P, distilled from the findings for S on V. In a typical application in economics, T might be used as a set of test criteria for properties in $\vec{S}$ which we desire to know but cannot measure directly.

Let us remark that the prediction discussed above, which answers S as the unique tangle τ of S does that induces the views of our unknown person $p \in P$ on T, is not the only reasonable prediction for $p(S)$ if we are in the scenario of Theorem 2. In this wider context, the specification $p(T)$ of T chosen by p will still extend to a unique maximal tangle in S, but this may be a tangle of low order: a tangle of S_k for some small k. In such a case we still have to decide what to predict as $p(s)$ for all $s \in S \setminus S_k$. Making this choice is related to the problem, discussed in Section 4.4, of how best to associate a given $v \in V$ with some tangle of the entire S.

---

[94] ...in the sense that it is more likely to be correct on many questions in S

Quantitative estimates for the quality of these predictions can only be established depending on context. Since feature systems and their tangles are so general, good estimates cannot be expected to apply to worst-case scenarios but will have to be probabilistic. They will have to be based on a probability distribution on the set of all feature systems of a set V, but which such distributions are reasonable to assume will depend on the intended application.

5. Applying tangles: back to the examples

In this chapter we revisit the scenarios discussed in Chapter 2 of how tangles might typically be applied in the social sciences. Having seen a more formal definition of tangles in Chapter 3 and statements of the two main tangle theorems in Chapter 4, we are now in a position to say more precisely for each of those scenarios how the tangles discussed there may be defined, and what the two theorems mean in those contexts.

Our aim is that every reader should be able to evaluate, by following these templates, how tangles might be defined in their own professional context, and gauge what their impact might be.[95]

For each of the contexts described in Chapter 2 we already discussed there how the sets S, V and $\mathcal{F}$ could be specified. All 'tangles' mentioned in this chapter will be $\mathcal{F}$-tangles, typically[96] with $\mathcal{F} = \mathcal{F}_n$ for some integer n.[97] Our aim now is to do at least some of the following, separately for each context:

- describe how Theorems 1 and 3 can be interpreted;
- discuss choices for an order function as introduced in Chapter 3.4;
- describe the evolution of the resulting tangles (Ch. 3.2), and Theorem 2;
- interpret any clustering in V based on the tangles of S, as in Chapter 4.4;
- discuss the use of the tangle-distinguishing tree set T;
- compare the tangles of S with their dual tangles of V from Chapter 3.5.

In order to avoid repetition, we shall discuss aspects that apply in most or all of these contexts only once, in the mindset example in Section 5.1. This section, therefore, is recommended reading also for readers primarily interested in another of the fields discussed later.

5.1 Sociology: from mindsets to matchmaking

In the simplest scenario for Chapter 2.1 we think of S as a questionnaire, which a sample V of a population P of people in whose views we are interested have

---

[95] As pointed out before, my group and I will be happy to collaborate on any such effort, assist with mathematical or technical support, and provide some relevant software.

[96] Other choices of $\mathcal{F}$ – e.g., meaning-driven choices as in Chapters 2.1 and 3.4 – are possible and may be discussed too where relevant.

[97] See Chapter 3.1 for the definition of $\mathcal{F}_n$.

filled in.[98] The sample V is chosen so as to represent P well. A tangle of S is a mindset or a typical pattern of behaviour, known or unknown.[99] The set $\mathcal{F}$ may include subsets of $\vec{S}$ designed to exclude the views of people we are explicitly not interested in.

Theorem 1, our tree-of-tangles theorem for a fixed feature set, finds a tree set T of questions that are critical for telling apart the mindsets found as tangles. Both the tangles found and the tree set T that distinguishes them will typically contain not only elements of S but also Boolean expressions of elements of $\vec{S}$: certain combinations of particular views on single matters queried by S.

The fact that T is a tree set is likely to correspond to logical inference. For example, S might have a tangle representing enthusiasm for skiing, and another for boxing. Since these are distinct mindsets, T will have to contain a question that tells them apart. If we assume that no-one likes both skiing and boxing, this might be the question s asking 'do you like skiing', or the question r asking 'do you like boxing'. If the yes-answers to these questions are $\vec{s} = B \subseteq V$ and $\vec{r} = C$, respectively, we would have $B \subseteq \overline{C}$ and $C \subseteq \overline{B}$.

Suppose now that there are also mindsets that entail a loathing for all sports. Then T must contain questions that distinguish these mindsets from the two sport-loving ones. Neither r nor s achieves this: r fails to distinguish those sport loathers from the skiers (since these also dislike boxing), while s fails to distinguish them from the boxers. We thus need a further question $t = \{A, \overline{A}\}$ in T to do this. This might be the question 'do you like any sports', with yes-answer $\vec{t} = A$, say. Since skiing and boxing are types of sport,[100] it will satisfy $A \supseteq B$ and $A \supseteq C$.

The splitting stars of T, see Chapter 4.2, correspond to the various mindsets that T distinguishes. Each splitting star consists of the logically strongest views in $\vec{T}$ consistent with the corresponding mindset. It thus represent this mindset 'in a nutshell'. If, in the above example, we have one comprehensive mindset of loving sport but hating both skiing and boxing (rather than several mindsets for further types of sport), the splitting star of this mindset would be the set $\{\vec{t}, \overleftarrow{s}, \overleftarrow{r}\} \subseteq \vec{T}$. If there is no such mindset, but still no further sports mindsets, then T will not contain all three questions r, s, t. It might contain t and s, in which case $\{\vec{t}, \overleftarrow{s}\}$ would be the splitting star of the mindset of loving sports but not skiing. This mindset would differ only immaterially from the boxing mindset, since there are not enough sports enthusiasts that dislike

---

[98] To avoid repitition, we subsume under this also the slightly different scenario where S is not a questionnaire but a list of aspects of behaviour that has been observed amongst the people in V, and S has been filled in by an observer.

[99] While often we shall be naturally interested in understanding the mindsets found by our theorems and algorithms in terms of some intuitive interpretation, it is important to note that the theory does not require this. In particular, we can use it to make tangle-based predictions for the views or behaviour of people in $P \smallsetminus V$ without such an intuitive understanding. Compare Chapter 4.5.

[100] This is an example of where the fact that T is a tree set reflects logical inference.

both skiing and boxing to form their own mindset. Similarly, T might contain t and r but not s, in which the skiing fans would be indistinguishable from the more general sports fans that dislike boxing.

In Theorem 3, the splitting stars of T are in $\mathcal{F}$, and are therefore each held only by a small set of the people in V, or by a set of people deemed irrelevant. The entire set V is then fragmented into these small subsets, which documents conclusively that no significant mindset in terms of S exists in the population V (and, by inference, in the population P).

In a more subtle setup we have an order function on S that assigns[101] low order to questions of a fundamental or basic nature, and higher order to questions about more detailed aspects of mindsets expressable in terms of those basic views. Indeed, views held by a mindset τ of S_k refine those of the mindset $\tau \cap \vec{S_\ell}$ of S_ℓ for each $\ell < k$.

In our earlier example, the question t about sports in general will probably have lower order than the questions r and s about boxing and skiing, since it is less specific. We might then have a tangle of low order for sport enthusiasm, and tangles of higher order for the love of skiing and boxing. These latter tangles would both refine the general sportstangle.

The splitting stars of the tangles of lowest order in the tree set T found by Theorem 2 are the collections of strongest views for the mindsets at this basic level. These mindsets branch out into more refined mindsets of which splitting stars of features of T of higher order are the sets of strongest views. Theorem 2 thus enables us to extract from the data provided a hierarchy of mindsets of varying degree of complexity, which unfold from the most basic to the most specific. We also have in T a small selection of key questions on which these mindsets differ, so that we can tell them apart just by evaluating those few questions.[102]

Finally, we have the option of not choosing the order of the questions in S explicitly, but to compute orders of all partitions of V as outlined in Section 3.4, and then to consider for Theorem 2 the sets $S_1 \subseteq S_2 \subseteq \ldots$ arising from this order function. This approach has the advantage that it is entirely deterministic once the questionnaire S has been set up, in that we do not interfere with the result – what mindsets are found as tangles – by declaring some questions as more basic, fundamental, or important than others. On the other hand, we have less influence on what kinds of possible mindsets are found.

Once we have determined the mindsets in terms of S that are prevalent in our population V we can, if desired, associate individuals $v \in V$ with mindsets that represent their views. As discussed in Chapter 4.4, there are various ways

---

[101] To ensure that our order function becomes submodular, as required by Theorem 2, we can make these assignments indirectly by way of a weight function $w \colon S \to \mathbb{N}$, as explained in Chapter 3.4.

[102] However, remember that the questions in T are not necessarily themselves in S: they may be combinations of questions from S. In order to distinguish all the mindsets on S, it is enough to ask the questions from S needed to build the composite questions in T.

of doing that, and which is best will depend on the context. Note, however, that the more sophisticated methods discussed there, which assign a person v to a tangle of S based on his or her entire set $v(S)$ of answers to S, is not available for the larger population P whose views are known only for the questions in T. For these individuals, the tangle of S that represents their views best is likely to remain the unique tangle of S that extends their views $v(T)$ on T, assuming that these are consistent.

The tangle-based approaches to clustering outlined in Chapter 4.4 can also be used for a matchmaking algorithm as envisaged in Chapter 2.1, by matching individuals that have been assigned to the same tangle or to tangles that are similar in terms of a similarity function σ.

The T-based clustering method mentioned last, for example, could be used to assign two individuals u, v the distance $\sum_{t \in \rho T \tau}(1/|t|)$, where $\rho T \tau$ is the set of questions t in T that distinguish the tangle ρ to which u was assigned from the tangle τ to which v was assigned.[103] In particular, if u and v were assigned to the same tangle, because they answered the tangle-distinguishing questions from T identically, they are likely to be matched.

If the tangles ρ and τ that u and v were assigned to are not the same, u and v will still be considered relatively close if they differ on only a few questions from T and these have relatively high order, i.e., are specific rather than fundamental questions. In the interpretation this means that u and v are deemed to be close, and are thus likely to be matched, if they share the most fundamental character traits and differ only one some more specific traits that distinguish more refined types of character, while they may differ on many other questions in S that are irrelevant to distinguishing types of character.

5.2 Psychology: diagnostics, new syndromes, and the use of duality

In Chapter 2.2 we discussed two possible applications of tangles in psychology.

The first was the formation of idiosyncratic concepts[104] from a plethora of perceptions in the mind of a particular patient: perceptions that combine loosely into a notion in this patient's mind which cannot be naively understood by the therapist, because these perceptions are not commonly associated with each other to form a concept in other people's minds. While I believe that the idea of tangles can inspire the discovery of such idiosyncratic concepts in individual patients, the implementation of this may best be left to the therapist's informed intuition: there will rarely be enough data about our individual patient's perceptions to warrant a computer-based algorithmic search for tangles defining his or her unfamiliar concepts.

---

[103] For graph theorists: these t correspond to the edges of the path in the tree $\mathcal{T}$ between the node representing ρ and the node representing τ.

[104] It is important here not to confuse concepts with words used to express them: since language is defined by interpersonal consensus, the individual's concepts and notions we are trying to understand here will typically not be associated with a word.

Our other application of tangles from Chapter 2.2, however, seems ideally suited to a large-scale computer-based approach. The idea there was to see which psychological symptoms, collected across many patients, combine into recognizable syndromes that can serve as a focus for more targeted research and the development of treatments.

An obvious possible emphasis for a large and fundamental study would lie on discovering previously unknown syndromes. However, any tangles-based definition of 'syndrome' will also encompass known syndromes, which should likewise show up in such a study. This can be used in two ways. First, to validate the tangle-based model by checking that known syndromes and psychological conditions are reliably found. Secondly, tangles will offer some quantitative underpinnings also to the ongoing study of previously known syndromes.

The set T of tangle-distinguishing features found by Theorems 1 and 2 on the basis of such a large study can then be used also for diagnostic purposes in individual patients. The idea would be to test T, rather than the much larger S, on a given patient first, and if this returns a consistent specification of T we have identified a tangle that is most likely to describe the patient's psychological condition: the unique tangle of S that extends this consistent specification of T.

On the face of it, this procedure resembles the familiar classical diagnostic approach in medicine based on 'exclusions': just as a doctor tests specific conditions in order to exclude potential illnesses as the cause of the symptoms observed, testing each potential feature in T will support some and exclude other possible tangles of S. Indeed, T may well be seen as a particularly efficient set of such tests, which evidence-based guidelines to doctors may suggest.

However we must not forget that each tangle, being a particular specification of S, is just an 'ideal' set of test results that corresponds to a syndrome. Diagnosing an individual patient with one of these syndromes is still a nontrivial process: we have to find the syndromes that best match this patient's own test results on S. There are several ways to do this – see Chapter 4.4 – and this should be a matter for learned debate within the discipline as much as for the psychologist in charge of the patient.

Finally, there is an aspect of tangle duality that is particularly useful in a psychology context. Let us again think of S as a set of questions answered by every person v in V, even though in our current context the $\vec{s} \in \vec{S}$ are more likely observations made by a psychologist than answers given directly by the client.

In Chapter 1.5 we briefly considered so-called witnessing sets for tangles: a subset X of V is said to *witness* a tangle τ of S if, for every $s \in S$, the majority of the elements of X specify s as τ does, rather than in the opposite way. If this majority subset of X has size at least $p|X|$ for every s, where $p \in [\frac{1}{2}, 1]$, let us call X a *p-decider* for S.

Note that this definition no longer needs or refers to τ: a set $X \subseteq V$ is a p-decider for S as soon as every $s \in S$ is specified in the same way by at least $p|X|$ of the elements of X. In particular, whether or not a set $X \subseteq V$

is a p-decider for S is not, as such, a property of tangles. Every $X \subseteq V$ is trivially a $\frac{1}{2}$-decider for S, and every singleton $\{v\}$ is trivially a 1-decider for S, but for given $p > 1/2$ there may be no p-decider for S in V of any non-trivial size. If there is, however, and p is close to 1, then its elements are essentially indistinguishable in terms of S: every question in S is answered by most of them in the same way.

Now recall from Chapter 3.5 that the elements $v \in V$ of the ground set of our feature system $\vec{S}$ are themselves potential features in a dual feature system $\vec{V}$ with ground set S. A p-decider for the dual system $\vec{V}$, then, is a subset S' of S such that every $v \in V$ answers most of the questions in S', at least $p\,|S'|$ many, in the same way: either yes or no.

The questions in S', therefore, can be used interchangably to some degree, which can be particularly useful in a psychology context. For example, if we are interested in how our patient would answer some particular question $s' \in S'$, but we cannot ask s' directly because of some taboos or traumas, or because asking s' might influence his or her other answers, we can instead ask some more innocuous $s'' \in S'$ and interpret its answer as an answer to s'.

5.3 Political science: appointing representative bodies

The starting point for this example was that bodies elected to make decisions on behalf of the population that elected them should be composed of delegates that represent the typical views held in the electorate, perhaps in numbers reflecting the popularity of these views. Given that tangles provide a model for precisely this, views commonly held together, one might seek to constitute a representative body by appointing as delegates people whose views best resemble the tangles found for the set S of relevant issues determining the election.

This problem was addressed, and solved, in Chapter 4.4: we discussed various similarity functions for specifications of S, and since both the tangles of S and the sets of views on S of individuals $v \in V$ are specifications of S, we can compare them and delegate to our representative body those $v \in V$ whose views are closest to the various tangles found for S.

Even if all the main views held (in the electorate V) are now represented in our representative body, the question is still whether that representation is fair: more commonly held views should be represented by more delegates.

Theorem 2 provides a neat way to solve this problem too. Recall that the algorithm which computes the tree of tangles for a hierarchy $S_1 \subseteq S_2 \subseteq \ldots \subseteq S$ of feature sets (Chapter 4.3) does so for the sets S_k in turn, as $k = 1, 2, \ldots$ increases. At each stage k it evaluates whether or not a tangle found for S_k extends to one or several tangles of S_{k+1}. At this time, the algorithm has already found the subset T_k of T needed to distinguish all the tangles in S of order up to k. The nested set T_k partitions V into the sets $V_\tau := \bigcap \{ A \mid A \in \tau \cap \vec{T}_k \}$, where τ varies over the consistent specifications (or tangles) of T_k.

At this stage we could establish our representative body by appointing for each τ a number d_τ of delegates proportional to the number $|V_\tau|$ of people

whose views are best represented by τ.

However, Theorem 2 enables us to improve on this still, by chosing the composition of the group of delegates for τ not arbitrarily (specifying only its size) but according to the tangles of higher order into which τ 'splits': the tangles of $S_{k+1}, S_{k+2}, \ldots$ that induce τ. In other words, when we run our algorithm we make a decision at stage k for every τ as above whether or not we seek to refine T_k at the node representing τ: for those τ whose $|V_\tau|$ is smallest we appoint one delegate, while for the τ with bigger $|V_\tau|$ we keep the algorithm running.

Even then we may end up with sets V_τ that are bigger than others: it can happen that a tangle τ is represented by many people but still does not split into higher-order tangles. So we shall still pick possibly different numbers d_τ of representatives for different tangles τ. But this procedure uses such arbitrary choices of representatives only as a last resort, when no alternative choices informed by higher-order tangles can be used instead.

Unlike in a party-based representative vote, the composition of delegates obtained in this way reflects not only the relative support of the major parties in the electorate, but also within each of these blocks its various shades of more refined views, all proportionately to how these views live in the electorate.

5.4 Education: devising and assigning students to methods

In Chapter 2.4 we discussed how tangles can help to group teaching techniques into 'methods' so that different types of students can benefit best from techniques that work particularly well for them. A 'method' in this sense is simply a collection of techniques that are applied at the same time in one class. The idea is that, since at a large school it may be possible or necessary to offer the same content in several parallel classes, these classes could use different such 'methods' to benefit different types of students.

Our approach was to define 'methods' as tangles of techniques: groups of techniques that often work well together, where 'often' means 'for many potential students'.

Finding these tangles will require a fundamental study involving a large set V of students. Every student in such a study will have to be tested against various techniques, say those in a set S, which requires a considerable effort over time.[105] Once these tests have been performed we can define an order function on S to distinguish techniques that made a big impact (positive or negative) on many of the students tested: such techniques will be assigned a low order to ensure that they are considered even for the most basic tangles computed. If we wish not only to find tangles of S but to apply our theorems to them, we could view the $s \in S$ as partitions $\{A, B\}$ of V and define their

---

[105] Older students might, of course, simply be interviewed and asked which teaching methods they prefer. This may be less significant, but it may help make such a study feasible.

order $|A, B|$ indirectly through weighted similarity functions, as in Chapter 3.4, giving greater weight to the techniques found to be more divisive in our tests.

Either way, the k-tangles found will be concrete specifications of some techniques, ranging from the fundamental to the more subtle as k increases. They each form a 'method' in that they lay down which of the techniques they specify should be used or not when teaching 'according to' that tangle.

All this cannot be done at a single school facing the practical problem of how to set up its classes and assign its students to these classes. Conversely, not every 'solution' found by a large study, such as a set of tangles and a tangle-distinguishing set T of critical techniques, can be implemented at any given school. Indeed, local requirements or possibilities may dictate how many classes can be taught in parallel (and hence, into many tangles or 'methods' the techniques available should be grouped), and which techniques are available may depend on the composition of the local teaching staff.

However, once we have the data from our large study, a complete specification $v(S)$ of S for every student v tested, we can run our tangle algorithms with different parameters, parameters than can be adjusted to a school's given requirements or possibilities. For example, we can adjust the parameter n in the definition of $\mathcal{F}_n$ so as to obtain a desired number of tangles to match the target number of parallel classes, or we can delete potential techniques s from S before computing the tangles if these techniques cannot be implemented at the given school.

The result of this process, then, is a tangle-based concrete way to group the techniques available into as many 'methods' as parallel classes can be offered.

But how do we assign students to these classes? One possibility is to simply describe the methods on which the various parallel classes are based, and let them choose. Another is to test, or at least ask, students which techniques from the specific set T found by the algorithm for Theorem 1 works best for them: by definition, this is a small set of 'critical' techniques, each a combination of the techniques tested directly (those in S), on which the methods found differ. Every (consistent) specification of T defines a unique tangle that extends it – one of the teaching methods that correspond to the classes offered in parallel.

As with our earlier task of determining a set of methods to be offered, the set T that critically distinguishes them can be computed locally to match the concrete school's possibilities and requirements. In this way, the results of the same large study can be used by different schools to compute a bespoke set of particularly critical teaching techniques against which students can be tested in order to assign them to the classes corresponding to the various methods.

As the number of tangles involved in these applications will be small, the way to do this will be to assign each student v to the class whose tangle τ extends his or her specification $v(T)$ of T, the class $V_\tau := \bigcap \{ A \mid A \in \tau \cap \vec{T} \}$ defined after Theorem 1 in Chapter 4.1.

5.5 Meaning: from philosophy to machine learning

Our notions determine how we see the world. They are our way of cutting the continuum of our perceptions and ideas into recognizable chunks which, to some extent, persist in time, space, and across different people. Chunks of ideas or perceptions which we bundle together again and again, in different places, and which are similar to the recurring chunks in other people's minds.

Inasmuch as we understand, remember, and communicate aspects of the world around us as structures of notions, the question of what these notions are, i.e., which ideas or perceptions we combine into notions, determines what we can understand, remember, or communicate about the world. Since in everyday life[106] we do not choose our notions consciously, understanding and quantifying them – as tangles enable us to do – is perhaps significant for our own understanding of the world only when we seek to compare how people from different cultures or backgrounds understand the world differently.

But as soon as we try to teach a computer what our notions are, perhaps through some interactive process,[107] we need some quantitative definition of a 'notion'. As we have already seen, tangles may well offer that.

More ambitiously, if we seek to enable computers to 'understand' the world by themselves, the question of what should be their own fundamental notions, not necessarily copying ours, will be on the agenda even more. This is because there are 'good' notions and 'bad' ones, judged by how they help us understand the world. Good notions cut the continuum of ideas and perceptions along lines running between phenomena which we would like to distinguish,[108] whereas bad notions cut across such phenomena and are therefore less helpful for distinguishing them. As we saw in Chapter 3, this aim of finding good notions, possibly unlike ours, is very similar to finding tangles seen as specifications of set partitions.

We discussed in Chapters 1.3 and 2.5 how the notion of 'chair' can be captured as a tangle of potential features of furniture. Note that we can choose to be more demanding regarding the quality of this notion by raising the parameter n in the definition of $\mathcal{F}_n$-tangles, since a furniture type will show up as an $\mathcal{F}_n$-tangle τ only if every three of the features in τ are shared by at least n items. Another way to be more demanding of an $\mathcal{F}_n$-tangle, in order to ensure it only identifies high-quality notions, would be to replace 'three' with some higher number: to ask that not only all triples of features in tangle are shared

---

[106] Unlike, for example, in mathematics: there, decisions about which properties of the (mathematical) objects studied should be bundled into notions to facilitate further study are part of our daily bread. And they are as important as elsewhere in life, since they determine what structures of the objects under study become visible and can therefore be explored.

[107] For example, the computer might learn by showing us pictures of groups of objects and asking which of them 'do not belong' in this group.

[108] Despite the wording here, this is not entirely a matter of taste. Very crudely, one might say that good notions facilitate the expression of theories that describe the world better than others – for example, in the sense of making more specific or substantial predictions.

by many items but, perhaps, all quintuples.[109]

For the sake of our chair example, we considered as features only properties of furniture, such as being made of wood, which may or may not apply to a particular piece at hand. However when we try to train a computer to form or recognize notions, there is no reason to be so restrictive: we can also use other parameters that help us distinguish between different notions, such as context, or the time when it was fashionable.[110]

In Chapter 3.4 we saw how to use order functions to divide potential features into hierarchies, assigning low order to basic features and higher order to more specific ones. This is nowhere more relevant than when we employ tangles to capture notions in our ideas and perceptions: some notions are clearly more fundamental than others, and some questions (potential features) are more basic than other questions and thus should have lower order.

In addition to, or as a basis for, assigning an order to potential features in terms of similarity functions for set partitions (see Chapter 3.4), there are two ways of defining order explicitly that come to mind in the context of meaning: one is relevance, the other what one might call clarity or definiteness.

As an example for a relevance-based order function we could choose to assign all colour questions high order in our search for furniture types if we consider colour irrelevant to how furniture splits into types. As an example for a clarity-based order function, suppose we are trying to distinguish hairstyles. We are likely to get more helpful answers in the clarity sense (of delineating hairstyles from each other) if we ask when they were fashionable than whether they are pretty, so the former question should receive lower order on the clarity scale than the latter.

Note that these two considerations for how to define an order function may well conflict: being pretty or not is perhaps the most relevant aspect of a hairstyle, but it may also be the least clear in that people have divided opinions about it. Note also that even if we wish to base our order function on relevance or clarity or both, we do not necessarily have to define the order $|s|$, or a weight $w(s)$, manually for every question s: both relevance and clarity can be gleaned to some extent from how the people who took our survey answered its questions, and therefore computed mechanically.

What we have described so far is how a computer can learn the meaning of words as we use them, or come up with notions of its own that are formed by observing the world. Both these are aspects of what one would call the formation of passive vocabulary. Once this has been achieved, our computer will also want to know which of its tangle-encoded notions best describe a given object presented to it. This is a problem we discussed in Chapter 4.4:

---

[109] If we are too demanding, of course, we shall simply not get enough tangles to serve as notions, so we have to strike a balance here. But the quantitative parameters of $\mathcal{F}_n$-tangles allow for precisely that – not to mention the numerous possibilities of other choices of $\mathcal{F}$.

[110] Imagine a party game where a person thinks of something and we have to guess what it is by asking yes-no questions of any kind. Every such question is a 'potential feature'.

the problem of how to match a given object to a tangle of potential features, the tangle that best captures its actual features. See there for further discussion.

Let us complete this section with a straightforward application of tangles of notions, just as an example of what may be possible: an interactive thesaurus for non-native speakers.

Roget's classical *Thesaurus* helps writers find the best word for what they are trying to say simply by grouping together some likely candidates. The relevant group of words can be found by looking up a word whose meaning is close to the intended meaning, but maybe does not quite capture it. If a word has multiple meanings, it is linked to several such groups.

The value of this lies in offering the writer a relevant choice, but the thesaurus does not help with making this choice. This is fine if the writer knows all the words on offer, and perhaps just could not think of the right one. For learners of a language, however, this falls short of their need: they, too, will know what they are trying to express, but need help in finding the best word for it. Let us see how tangles can help them.

In the simplest model, we could devise for every word field currently offered together as a group of choices a questionnaire S whose answers for any given context would enable us to choose the correct word from this field. Each of these words, then, is likely to correspond to a tangle of S – a tangle we would be able to compute before our thesaurus is published. We could then also compute the tangle-distinguishing small set T of questions from Theorem 1. These questions from T could then be put to the user trying to identify the right word for their intended notion, and answering just these will steer them to the word that best fits their intended notion.

Recall that the questions in T may be combinations of questions from S, so the questions a user really has to answer may be a little larger. But the subset of S needed to form the questions in T is still likely to be smaller than S itself. This is crucial for making a good thesaurus: it will matter to the user whether they need five questions or fifteen to be steered to the fitting word.

Devising a questionnaire S for each word field in the thesaurus, and answering all its questions for each word, may look like a lot of work. But this work has to be carried out only once, when the thesaurus is made: it is offset by a gain on the user side in that the questions asked are chosen specifically for each word field selected by the user, and are chosen particularly well for this word field.

In a more sophisticated model, we could grade the questions from S as more or less relevant for the corresponding word field, and assign them an order correspondingly. We would then obtain a hierarchy of tangles as described in Chapter 3.2. This would correspond to a hierarchy of more or less general or specific words, just as in reality.

In this last respect, our thesaurus might even learn from user interaction. Although a user may not know the meanings of the words on offer, they will be able to grade the questions put to them as more or less relevant to their specific

search: tangles, after all, reflect notions, not words, and users come with such notions in their minds. Our tangles will then have to be recomputed from time to time when enough user feedback has been collected, and re-checked editorially against the words on offer. At this point, editors might also comment on user-defined tangles that do *not* have a corresponding word to match: such tangles will exist, since notions exist that are not exactly matched by words. Compared against the notions in the minds of speakers of other languages, however, this seems even more likely and worth addressing in ongoing editorial work.

5.6 Economics: customer and product types

Let us rephrase our example from Chapter 2.6 of an online shop, and the tangles of the dual feature systems this gives rise to, in terms of set partitions as introduced in Chapter 3.1.

In the example, V was a set of customers of an online shop, and S the set of items sold at this shop. We assumed that each customer v has made a single visit to the shop, specifying $s \in S$ as $v(s) = \vec{s}$ if he or she included s in their purchase, and as $v(s) = \overleftarrow{s}$ if not.[111] In the dual setup, every item $s \in S$ specifies every $v \in V$: as $s(v) = \vec{v}$ if v bought s, and as $\overleftarrow{v}$ otherwise. Thus, $s(v) = \vec{v}$ if and only if $v(s) = \vec{s}$.

In the terminology of Chapter 3.1, every item $s \in S$ is a partition of V: the partition $s = \{\vec{s}, \overleftarrow{s}\}$ of V into the complementary set

$$\vec{s} = \{v \in V \mid v \text{ bought } s\}$$

of customers that bought s and the set $\overleftarrow{s} = V \smallsetminus \vec{s}$ of those that did not. Similarly, every customer $v \in V$ is a partition of S: the partition $v = \{\vec{v}, \overleftarrow{v}\}$ of S into the complementary set

$$\vec{v} = \{s \in S \mid v \text{ bought } s\}$$

of items that v bought and the set $\overleftarrow{v} = S \smallsetminus \vec{v}$ of items that v did not buy.

We thought of the specification $v(S)$ of S as v's 'shopping basket', the division of S into the set $\vec{v}$ of items that v bought and the set $\overleftarrow{v}$ of items that v did not buy. Similarly, we thought of $s(V)$ as the 'popularity score' of s, the division of V into the set $\vec{s}$ of fans of s and the set $\overleftarrow{s}$ of non-fans of s.

The tangles of S, then, were typical shopping baskets, or types of customer: ways of splitting the set S of items into two parts – bought versus not bought – that were reflected, by and large, by the purchasing behaviour of many customers. The tangles of V were typical popularity scores, or types of item: ways of splitting the set of customers into fans and non-fans which broadly reflects the popularity score of many items.

---

[111] Note that, unlike in most other contexts, every $s \in S$ has a default specification here: we use $\vec{s}$ to indicate that s is 'present' rather than 'absent', consistently for all $s \in S$.

More formally, an $\mathcal{F}_n$-tangle τ of S is a specification of S such that for every three elements s_1, s_2, s_3 of S there are at least n customers v that bought or failed to buy each s_i as prescribed by τ: such that $v(s_i) = \tau(s_i)$ for all i, or more explicitly, such that v bought s_i if $\tau(s_i) = \vec{s_i}$ and v did not buy s_i if $\tau(s_i) = \bar{s_i}$. Similarly, an $\mathcal{F}_n$-tangle ρ of V is a specification of V such that for every three customers v_1, v_2, v_3 in V there are at least n items s that were bought or not by each v_i as prescribed by ρ: such that $s(v_i) = \rho(v_i)$ for all i, or more explicitly, such that v_i bought s if $\rho(v_i) = \vec{v_i}$ and v_i did not buy s if $\rho(v_i) = \bar{v_i}$.

A subset U of V witnesses a tangle τ of S if, for every single $s \in S$, the tastes of most customers in U regarding s is as described by τ: if most customers in U bought s when $\tau(s) = \vec{s}$, and most customers in U did not buy s when $\tau(s) = \bar{s}$. A subset R of S witnesses a tangle ρ of V if every $v \in V$ bought or avoided most of the items in R as described by ρ: if v bought most items in R when $\rho(v) = \vec{v}$, and v avoided most items in R if $\rho(v) = \bar{v}$.

With this formal setup in place, let us now revisit the examples discussed in Chapter 2.6.

The clearest of these is the tangle ρ of V which specifies the rich customers v as $\vec{v}$, and all the others as $\bar{v}$. This is a tangle, because there is a sizable group of items which the rich customers will tend to like *and* all the other customers will tend to dislike: the expensive high-quality items. Formally, we shall find for every triple v_1, v_2, v_3 of customers n such items s to satisfy $s(v_i) = \rho(v_i)$ for all $i = 1, 2, 3$: so that, regardless of whether v_i is rich or not, their purchasing behaviour with respect to our n expensive high-quality items s will be as prescribed by ρ.[112]

If we assume that all the green items in our shop are more expensive than competing non-green items, there is a similar tangle of V which specifies a customer v as $\vec{v}$ whenever v likes ecological items (even if they are expensive), and as $\bar{v}$ if v does not care enough about ecology to pay more. The green items in S then witness this tangle ρ of V in the same weak sense as above, i.e., in terms of the definition of an $\mathcal{F}_n$-tangle: for every triple v_1, v_2, v_3 in V there are at least n items, all among the green ones, which each v_i buys if $\rho(v_i) = \vec{v_i}$ and does not buy if $\rho(v_i) = \bar{v_i}$.

Let us now change our assumptions slightly, and assume no longer that the green items are more expensive. Then ρ is no longer a tangle of V, since non-green customers might buy green items too, even if they do not prefer them: for a triple of non-green customers we may be unable to find n items that they all fail to buy.

Let us now look at tangles of S. If we assume that every $s \in S$ is either green or positively dangerous to the environment, we have a tangle of S witnessed by

---

[112] Note, however, that the expensive high-quality items do not witness the tangle of rich people in the sense of Chapter 1.5, because not every rich person will have bought, in their one purchase, more than half of all the expensive items.

the ecology-minded customers, again in the weak sense of the definition of $\mathcal{F}_n$-tangles. If we do not make this assumption, however, we may have no tangle of S, regardless of whether the green goods are more expensive or not: the ecology-minded customers will no longer witness such a tangle, because they will disagree too much on the goods that are not ecologically critical. More fundamentally, we may be unable to extend the specification (as $\vec{s}$) of the green products s to a tangle of all of S, because we may be unable to specify the environmentally neutral items in such a way that for every three of them we can find n customers whose purchases agree on these three items with our specification. In this case, therefore, we would only have a tangle of the set $S' \subseteq S$ of environmentally critical items, those that are either good or bad for the environment.

Is there a way to capture the preference for green items by a tangle even in this more realistic case that not all items in our shop are environmentally critical? Instead of looking for tangles of all of S we might try to define an order function on S so that, for example, the set S' above appears as one of the sets S_k. Then there would be a 'green' k-tangle witnessed by the ecology-minded customers. However it will be difficult or impossible to find an order function that singles out all the relevant subsets of S in this way – not to mention the fact that we want to be open to unknown tangles, for which we cannot hope to encode a suitable subset of S in the form of S_k.

However, the approach sketched in Chapter 3.4 offers a way forward. Let us consider the set U of all partitions[113] of S and define the *order* of $\{A, B\} \in U$ as

$$|A, B| := \sum_{a \in A} \sum_{b \in B} \sigma(a, b),$$

where

$$\sigma(a, b) = \left| \{ v \in V : a(v) = b(v) = \vec{v} \} \right|$$

is the number of customers that bought both a and b. Thus, very broadly, a partition of S gets high order, and is therefore disregarded when k-tangles are computed for small k, if it divides the contents $\vec{v}$ of many shopping baskets $v(S)$ – not the items missing from them – into two roughly equal parts.

Unlike our earlier idea, this unifies tangles of different subsets of S into tangles of the same set U. For example, the set S' of green items in S will witness a tangle of U: partitions of low order cannot divide S' into roughly equal parts, so most of S' will lie on the same side of – and thereby consistently specify – all such partitions.

Similarly, the set of the comparatively inexpensive items in S witnesses a tangle of U (thought not a tangle of V).[114] By contrast, the set of expensive

---

[113] Recall that *partitions* of a set, in our terminology, always split it into exactly two parts.

[114] It witnesses a tangle of U, because the pairs of inexpensive items will be bought together by many price-conscious customers, so no low-order partition of S will divide many pairs of inexpensive items, i.e. most of them lie on the same side of every partition. But as

items in S does not witness a tangle of U (let alone one of V): customers unable to afford them will not contribute to the value of σ for pairs of expensive items, while customers that can afford them are not necessarily likely to buy them more often than cheaper items. Hence, $\sigma(a, b)$ will not be greater for pairs a, b of expensive items than for arbitrary pairs of items.

This pair of examples shows neatly that tangles of U deliver what we hoped for: they describe, loosely and without any need for difficult judgments in single cases, groupings of items that provide some common motivation for purchase. In particular, they will do this for previously unknown kinds of motivation such as, perhaps, featuring a face on the packaging.

Conversely, tangles of U avoid some silly instances that can occur as tangles of S or V. For example, our online shop S has a particularly silly tangle: the empty shopping basket $\{\ \bar{s}\ \mid\ s \in S\ \}$. Indeed, if S offers any reasonable amount of choice, this is a tangle witnessed by all of V, even in the strong sense of Chapter 1.5: just because our shop has so many more items to offer than a single customer can buy, it is likely that for every item $s \in S$ there are many more customers v that did not buy s (and hence specify it as $v(s) := \bar{s}$) than there are customers that bought s. Thus, V is a witnessing set for the empty shopping basket as a tangle of S.

However, one can use Theorem 3 to show that there is no such tangle of U; see [2, 10].

References

[1] R. Diestel. Homology and duality of set separations. In preparation.

[2] R. Diestel. Tangles in the empirical sciences: indirect clustering by separation. In preparation.

[3] R. Diestel. Abstract separation systems. *Order*, 35:157–170, 2018.

[4] R. Diestel. Tree sets. *Order*, 35:171–192, 2018.

[5] R. Diestel, J. Erde, R. Pendavingh, and G. Whittle. Order functions for tangles of set separations. In preparation.

[6] R. Diestel, J. Erde, and D. Weißauer. Tangles in abstract separation systems. *J. Combin. Theory Ser. A*, to appear.

[7] R. Diestel, F. Hundertmark, and S. Lemanczyk. Profiles of separations: in graphs, matroids, and beyond. *Combinatorica*, 39(1):37–75, 2019.

---

with the ecological items, the set of inexpensive items does not witness a tangle of V that specifies precisely the price-conscious customers v as $\bar{v}$, even in the weak sense of providing the n elements needed for its triples, since customers that are not price-conscious have no motivation to avoid inexpensive items.

[8] R. Diestel and S. Oum. Tangle-tree duality in graphs, matroids and beyond. arXiv:1701.02651. To appear in *Combinatorica*.

[9] R. Diestel and S. Oum. Tangle-tree duality in abstract separation systems. Submitted, arXiv:1701.02509, 2017.

[10] R. Diestel and G. Whittle. Tangles and the Mona Lisa. arXiv:1603.06652.

[11] C. Elbracht, J. Kneip, and M. Teegen. Tangle algorithms for abstract separation systems. In preparation.

[12] C. Elbracht, J. Kneip, and M. Teegen. Tangles are decided by weighted vertex sets. arXiv:1811.06821.

[13] C. Elbracht, J. Kneip, and M. Teegen. Trees of tangles in abstract separation systems. In preparation.

[14] J. Erde and E. Gürpinar. Character traits and tangles. In preparation.

[15] Martin Grohe and Pascal Schweitzer. Computing with tangles. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pages 683–692, New York, NY, USA, 2015. ACM.

[16] NN. Tangles: a novel way to classify texts. In preparation.

[17] N. Robertson and P.D. Seymour. Graph minors. X. Obstructions to tree-decomposition. *J. Combin. Theory (Series B)*, 52:153–190, 1991.

[18] Ludwig Wittgenstein. *Philosophical Investigations*. Basil Blackwell, Oxford, 1953.