Preface

This is a book about potential applications of a new mathematical theory, written by a mathematician for a non-mathematical readership. Its style develops from an intuitively informal to a more formal level that uses basic mathematical language, just enough to make things precise. No serious mathematics is used anywhere in the main body of this book.[1]

This preface says a little about where tangles come from in mathematics, so as to indicate what is new in this book and what is not. Readers without this background are encouraged either to just skim the preface for a quick impression, or to skip straight ahead to Chapter 1. This begins with three separate introductions addressing natural scientists, social scientists, and computer scientists in turn.

The mathematical theory of tangles has its origins in the theory of graph minors developed by Neil Robertson and Paul Seymour in the final two decades of the 20th century. In a series of over twenty research papers, which culminated in the proof of one of the deepest theorems in graph theory, the *graph minor theorem*, Robertson and Seymour developed a new connectivity theory tailored specifically to the somewhat 'fuzzy' notion of their central object of study, that of a graph minor. Their new connectivity theory centred around a revolutionary new concept of high local connectivity in a graph: the notion of a *tangle*.

Loosely speaking, a tangle is 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 does not conform to the usual graph-theoretic notions of high connectivity.

Tangles constituted a shift of paradigm in what high *local* connectivity, somewhere in a graph or network, might mean. There is a standard measure of global connectivity for graphs, and the traditional way to measure their local connectivity was simply to look for regions in the graph that were highly connected in this global sense, applied to the region as a subgraph. As these highly connected regions were themselves viewed as graphs, they would be decribed in the same way as graphs are: by precisely naming their 'vertices' (or 'nodes') and the 'edges' connecting them.

Graph minors, on the other hand, the objects that Robertson and Seymour set out to study, are fuzzier substructures than subgraphs: a highly connected minor will usually persist even if the graph containing it is changed a little. Rather than describing these minors in the traditional, somewhat pedestrian, way of naming all their vertices and edges, Robertson and Seymour thought of an ingenious indirect way to capture just their essence: by declaring for every bottleneck in the graph on which of its two sides *most* of that minor lies.[2]

Such a collection of pointers at the bottlenecks of a graph came to be called a *tangle*. Of course, this is a hugely abstract kind of thing – if indeed it merits being called a 'thing' at all. However, bundling even the most complicated collections of objects and their relationships into a single notion is a process not uncommon in mathematics: it enables us to move on and describe more concisely any higher-level structures in which such composite objects occur. In our example, the collection of pointers that constitute a tangle in a graph, one at each of its bottlenecks, deliberately ignores the detail of what vertices and edges our highly connected minor consists of. Instead, it just records where most of it lies – relative to *every* bottleneck.

It turned out that this deliberate restriction of information about the highly connected minors in a graph came with a gain in clarity: the detail discarded was clutter, the information retained its essence.[3]

This development in graph theory was followed by a discovery which, quite unexpectedly, made the entire theory of graph tangles available for the analysis of highly cohesive substructure far beyond graph theory: it turned out that, not just for the notion of a tangle but also for the proofs of the deepest theorems about them, it is enough to know the relative position of those bottlenecks, rather than how exactly they divide the graph of which they are bottlenecks. This information can be encoded in some abstract way that is quite independent of graphs.

The theory of tangles has thus become applicable to a wealth of real-world scenarios. The purpose of this book is to show how this can work.

Our narrative starts with a naive discussion of what tangles mean in various real-world scenarios, and how tangle theory can make an impact there. It then takes the reader through the basic mathematical underpinnings of abstract tangle theory, just enough to enable them to set up a rigorous quantitative framework for applying tangles to their own field. It winds up by revisiting the example scenarios to show how the more formal theory plays out in these contexts.

It should be stressed that those real-world scenarios discussed are highly simplified: they are toy examples of contexts in which tangles can be applied. In reality, they can probably be applied somewhere in most of the natural and quantitative social sciences. This will require the input of experts in those fields. It is the aim of this book to put such experts in a position to try this out for themselves; generic software for this purpose is available via `tangles-book.com`.

The layout of this book is as follows. It begins in Chapter 1 with three short introductions to what tangles are, and what they are designed to achieve: in the natural sciences, in the social sciences, and more specifically in data science. These introductions can be read independently of each other, and are written so as to appeal to readers with these respective backgrounds. In this way, they provide three separate entry points to this book. However, they show aspects of the same big picture, and none of them requires any expertise in the area for which it was written. Hence readers with any background may well benefit from reading all three of them: they are all short, and they illuminate the notion of tangles from three rather different angles.

Chapter 2 develops the notion of a tangle from the intuitive picture formed in Chapter 1, still at an informal level. This will be accomplished by the end of Section 2.3. At this point, any reader who cannot wait to see some tangle applications will be sufficiently equipped to skip ahead to Part II, where applications are discussed informally on the basis of just the notion of tangles, not their theory as described later in the book.

Chapter 3 gives a first indication of the two main theorems about tangles, still not in formal mathematical language but in terms of the example settings described in the three introductions. Together, Chapters 1–3 form Part I of the book, an informal introduction to the notion and theory of tangles from three rather different application perspectives.

Part II continues with a collection of explicit example scenarios in which tangles might be applied, and describes informally what the mere notion of a tangle can already achieve there. As pointed out earlier, these example scenarios are highly simplified and, in their simplicity, artificial. The idea behind going through such a range of examples is to indicate the potential of tangles throughout the sciences, and to do so in an unassuming way that inspires readers to find tangles in their own field of expertise.

The examples in Part II were chosen to illustrate the diversity of potential tangle applications. The corresponding chapter sections may be dipped into at liberty: nothing here is required reading for any material later in the book, except for the corresponding sections in Part IV.

Part III then explains tangles a little more formally. It still does not assume any knowledge of advanced mathematics, but the description is in basic mathematical terms such as sets, subsets, functions and so on. The idea is that this more formal description of the notion of tangles, given in Chapter 7, should be precise enough to enable the reader to apply tangles to their own individual background.

Chapter 8 continues with statements of the two main tangle theorems. The first of these describes how the tangles of a large dataset lie with respect to each other: how some tangles refine others, and how the most refined tangles are separated by some particularly crucial bottlenecks which, between them, organize the dataset into a tree-like shape that displays where its main tangles lie. The second fundamental tangle theorem, which is equally important, tells us what our data looks like if it has no tangles. It offers verifiable quantitative evidence of the *lack* of structure in our data – for example, if it is polluted or inconclusive for some other reason.

The remainder of Part III describes the mathematical toolkit that enables us to tune tangles to fit an intended application (Chapters 9–10), and then describes the fundamental tangle algorithms in Chapter 11.

In Part IV, finally, we return to the examples discussed informally in Part II. Equipped with the formal notions from Part III, and having met the two main tangle theorems, readers will be able to see not just what tangles mean in those various contexts, but also how they can be structured and fine-tuned to offer insights relevant to that field.

Throughout the text, there are markers for 'footnotes' that are collected together at the end of the book. The reason I have implemented these as endnotes is that they can happily be skipped at first reading: they offer further illustrations, more detailed explanations and so on, which are not meant to interrupt the flow of reading unless the reader feels curious for more at that point already.

This book would not exist but for the inspiration and contributions in substance I received from numerous people over the past few years. The development of abstract tangle theory that underlies the applications envisaged here began with an idea of my student Fabian Hundertmark, who extracted from our then recent proof [6] of a canonical tree of tangles theorem for graphs the algebraic core of tangles that was actually needed in that proof [26]. When Sang-il Oum visited me in 2013, we found a proof also of the tangle–tree duality theorem based just on these minimalist algebraic foundations for the notion of a tangle [14, 15]. This set the scene for the development of abstract tangle theory based on [9], which was carried through in the following years mostly by various members of my Hamburg group, particularly by Sandra Albrechtsen, Johannes Carmesin, Christian Elbracht, Ann-Kathrin Elm, Raphael Jacobs, Paul Knappe, Jakob Kneip, Max Teegen, Hanno von Bergen and Daniel Weißauer.

The idea to use this abstract theory of tangles for applications outside mathematics was born when I told Geoff Whittle about it in Oberwolfach in 2016. I remember vividly his exclamation, '*surely*, as we can see structure and things in images so quickly, our brain just sees tangles!'. We then proved that in [16], most of which is now part of Section 14.6.

In the years that followed I benefited immensely from discussing tangle applications with quite a diverse set of people. Outside mathematics these include Partha Dasgupta in economics, Jane Heal in philosophy, Thomas Günther in virology, Chin Li in psychology, the CNRS group around Oliver Poch and Julie Thompson in protein sequencing, Rolf von Lüde in sociology, Ulrike von Luxburg in machine learning, as well as the people at Google including, in particular, Krzysztof Choromanski. Within mathematics they include Nathan Bowler, Joshua Erde, Jim Geelen, Rudi Pendavingh, and Geoff Whittle. To all these I extend my thanks for their ideas, enthusiasm and encouragement!

Last but not least, I thank my tangle software group of Dominik Blankenhagen, Michael Hermann, Fabian Hundertmark and Hanno von Bergen for their amazing success in bringing this pie down from the sky and rooting it firmly in fertile earth. Their generic tangle software is now available via `tangles-book.com` under an open-access licence [1]: for all who would like to play with it or just see some examples in action, to apply it in their own professional context, or to develop it further by adding their own packages to the library.

Reinhard Diestel, February 2024