

Data Science meets BGB: Eine Einführung in die juristische Datenvisualisierung

Ihar Nestsiaenia und Christian Hartz



Open Peer Review

Dieser Beitrag wurde lektoriert von:
Ferdinand Wegener & Louis Goral-Wood



Ihar Nestsiaenia ist Lead Machine Learning Engineer, spezialisiert auf die Anwendung und Implementierung von KI-Algorithmen. Er arbeitet für EPAM Systems und entwickelt für Wolters Kluwer. In seiner Verantwortlichkeit liegt die Anwendung von Machine Learning zur Lösung von Information-Retrieval-Problemen.

Christian Hartz ist Rechtsanwalt und Dozent. Bei Wolters Kluwer ist er als Legal Engineer im Team “Content Architecture & AI” als Product Owner für verschiedene nationale und internationale KI-Projekte verantwortlich.

Gesetze sind abstrakt und daher teilweise schwer verständlich. Referenzen auf andere Normen machen dies noch komplizierter. Was passiert aber, wenn man gerade diese Referenzen aus mathematischer Sicht betrachtet und sie visualisiert? Dieser Beitrag widmet sich der Visualisierung aller Zitierungen des BGB und zeigt, dass beide Seiten – sowohl Jurist:innen als auch Informatiker:innen – gewinnen, wenn sie zusammenarbeiten.

In der juristischen Ausbildung als auch in der täglichen Arbeit von Jurist:innen dreht sich vieles um Text: Gerichtliche Entscheidungen oder Kommentare durchgehen, Schriftsätze lesen und verfassen oder der berühmte Blick ins Gesetz.

Auch wenn es auf den ersten Blick nicht so scheint: Alle diese Texte haben gemeinsam, dass sie nicht linear sind. Es gibt Beziehungen innerhalb und außerhalb der Texte. Ein Schriftsatz zitiert eine wichtige BGH-Entscheidung, einen Kommentar oder eine Norm. Eine Entscheidung zitiert eine andere Entscheidung. Normen zitieren andere Normen innerhalb und außerhalb desselben Gesetzes. All diese Verknüpfungen bleiben oftmals unberücksichtigt, weil sie sich hinter dem Text verstecken. Schaut man sich die Bücher des BGB mit einem mathematischen Hintergrund an, so sticht einem die Verwendung der Graphentheorie ins Auge. In der Mathematik ist ein Graph eine mathematische Struktur, die dazu verwendet wird, Verbindungen paarweiser Informationen herzustellen. Jede Beziehung zwischen Objekten in der Welt kann so modelliert und als Graph dargestellt werden. Ein Beispiel hierfür ist ein soziales Netzwerk, bei dem jede Person ein Knoten darstellt und mit anderen Personen Beziehungen eingeht. Wenn wir diese Beziehungen visualisieren, erhalten wir einen Graphen. Sofern Beziehungen zwischen allen Knoten bestehen, nennt man dies ein Netzwerk. Auch Gesetze können als Graph dargestellt werden.

Versuchen wir uns einmal ein Gesetz als Graph vorzustellen:

Im ersten Schritt kann man sich das Inhaltsverzeichnis des Gesetzes anschauen. Diese Informationen können in einer Baumstruktur abgebildet werden. Auch das ist eine Art Graph. Gehen wir einen Schritt weiter: Zwar hilft uns das Inhaltsverzeichnis dabei, zu verstehen, wie das Gesetz aufgebaut ist, aber wir haben noch keine Ahnung, ob und wie Beziehungen der Normen untereinander aussehen.

In Gesetzen gibt es eine Vielzahl von Referenzen. Zitiert eine Norm eine andere Norm oder ein Konzept aus dieser Norm, so kann eine Beziehung zwischen diesen beiden Normen hergestellt werden. Wenn wir alle Beziehungen aller Normen her-

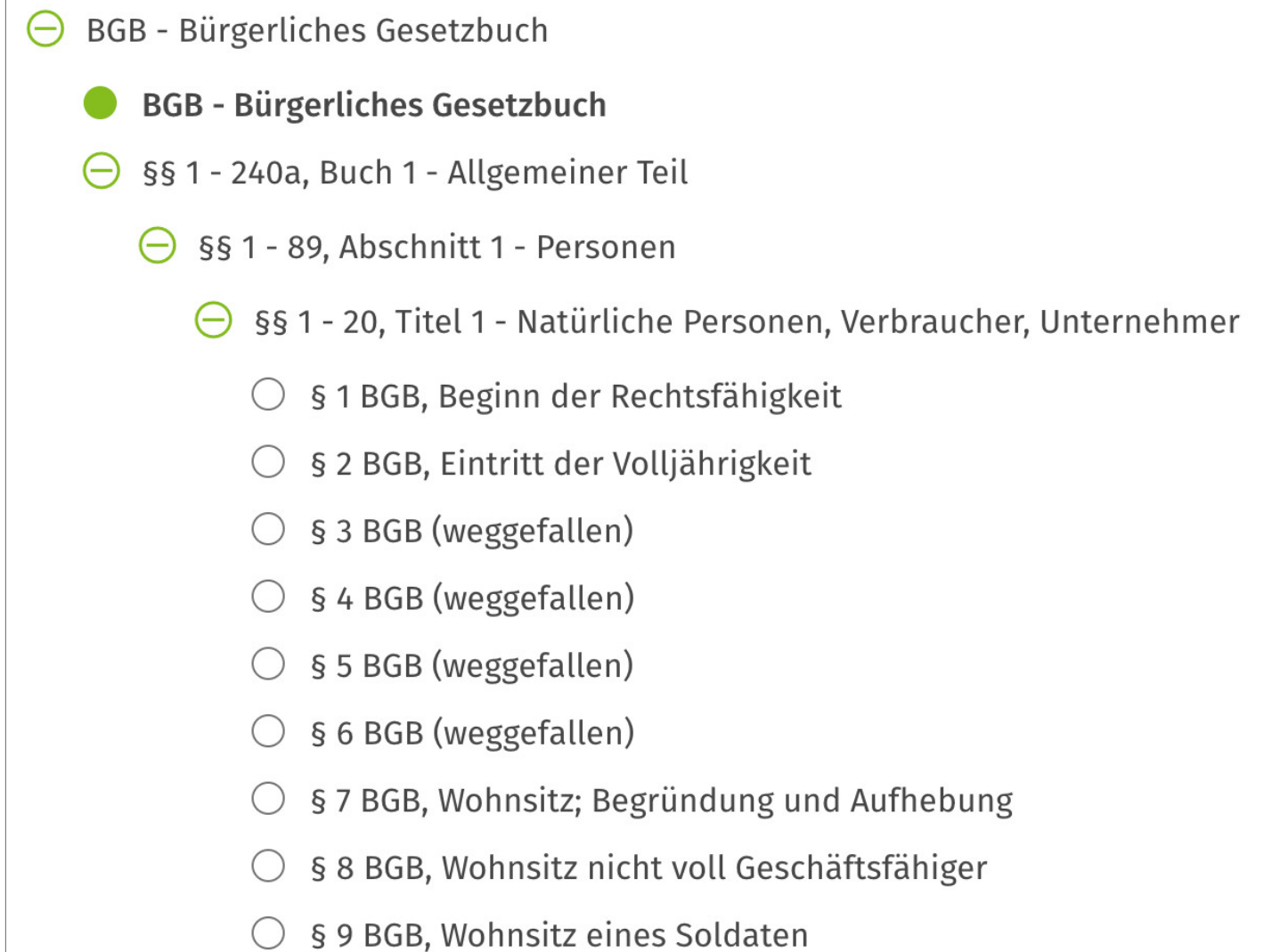


Abb. 1: Inhaltsverzeichnis des BGB aus Wolters Kluwer Online als Baumdiagramm.

stellen, so erhalten wir einen Graphen; genauer gesagt, einen gerichteten Graphen (sog. directed Graph). Gerichtet meint hier, dass wir die Richtung des Informationsflusses visualisieren. Im Falle der Norm also: Norm A zitiert Norm B (ausgehend), bzw. Norm C wird von Norm D zitiert (eingehend). Durch diese Unterscheidung und Visualisierung des Informationsflusses kann gezeigt werden, in welche Richtung der Verweis und damit das Einbinden der Information statt findet.

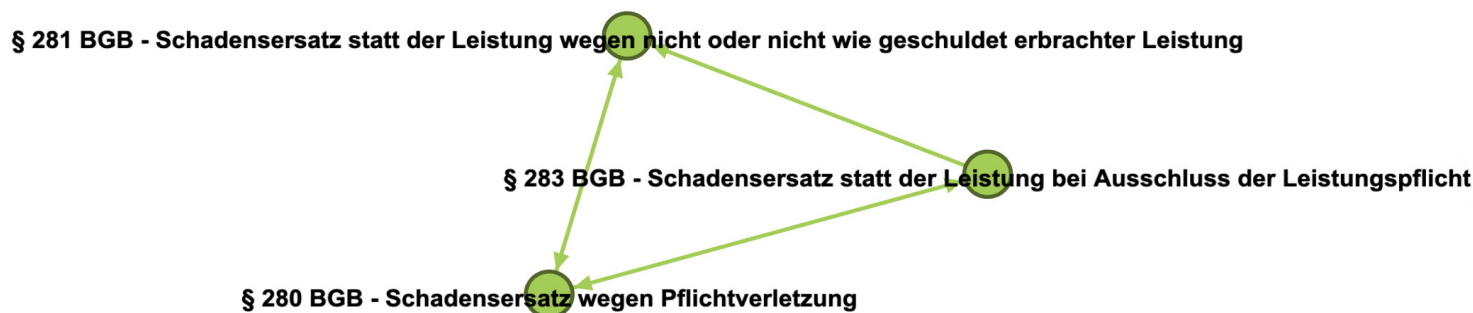


Abb. 2: Dieser Graph stellt dar, dass sowohl § 281 BGB (Knoten) als auch § 283 BGB (Knoten) direkt auf § 280 BGB (Knoten) verweisen (Kanten). § 280 BGB hingegen verweist selbst auch auf § 281 BGB und § 283 BGB. § 281 BGB hingegen verweist nicht auf § 283 BGB.

Hier sieht man eine einfache Visualisierung einer Norm, die selbst zitiert wird (§ 280 BGB). Die Norm wird als Kreis dargestellt und wird in der Regel als Knoten oder Ecke bezeichnet. Die Verbindungslinien, die bei einem gerichteten Graphen als Pfeil dargestellt werden, sind die sog. Kanten. Diese verbinden wiederum die anderen Normen mit der Norm, die sie zitieren. Wir haben allerdings nicht nur die Verlinkung von einzelnen Normen berücksichtigt, sondern auch von ganzen Büchern, Kapiteln, Titeln etc.; diese nennen wir strukturelle Knoten.

Nachdem wir zunächst erklären wollen, was Datenvisualisierung ist, warum wir sie verwenden (siehe Kapitel A.) und wie das technisch aussehen kann (Kapitel A. I.), wollen wir im zweiten Teil das Bürgerliche Gesetzbuch (Kapitel B.) etwas näher beleuchten.

A. Was ist Datenvisualisierung und warum verwenden wir sie?

Ausgangspunkt für die Analyse von Daten ist oftmals eine sogenannte Explorative Datenanalyse (EDA). Hierbei geht es nicht nur darum, Erkenntnisse zu erlangen, sondern auch ein Gefühl für die Daten zu erhalten. Mittel hierfür können statistische Auswertungen oder aber Visualisierungen sein. Dabei meint Visualisierung nicht nur die hier verwendete Darstellung der Beziehungen in Netzwerken, sondern es kann auch ein einfaches Balken- oder Tortendiagramm sein.

Verteilung der Links

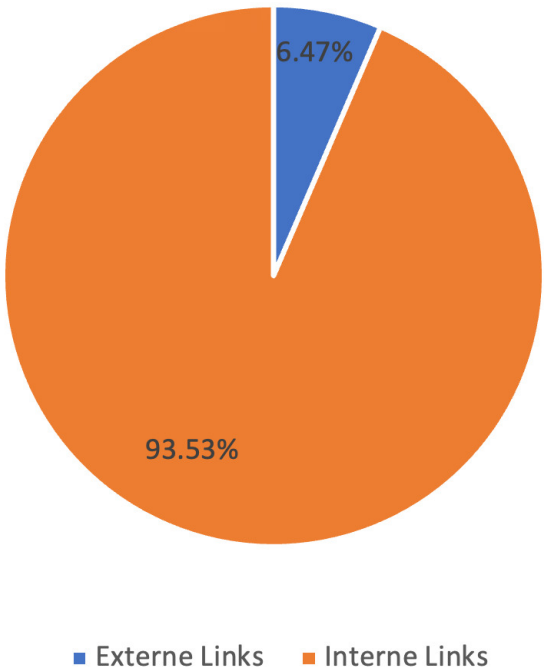


Abb. 3: EDA in Form eines Tortendiagramms, das alle Verweisungen des BGB auf andere Regelungen im BGB (interne Links; orange) und des BGB auf andere Gesetze (externe Links; blau) visualisiert.



Würde man beispielsweise alle Verweise des BGB einzeln in einer Tabelle ausgeben, wäre es bei der Anzahl an Verweisen (insgesamt 1932 Verweise, wovon 125 Externe und 1807 Interne sind) eine schier unüberschaubare Anzahl an Informationen. Mittels der Datenvisualisierung kann eine aggregierte Version erstellt werden (vgl. Abbildung 3).

Diese erlaubt es uns, auf Besonderheiten zu stoßen, die bei der Betrachtung einzelner Elemente nur sehr schwer erfolgen kann. So erkennt man etwa bestimmte Häufungen von Informationen an einem bestimmten Punkt (sog. Cluster) in einer Datenvisualisierung deutlich einfacher.

Schauen wir uns nun die technische Umsetzung etwas näher an.

Das Erstellen einer Graph-Visualisierung erfolgt in zwei Schritten: 1. Erzeugen des Graphen und 2. Auffinden des besten Weges zur Visualisierung. Die Ausgangsdaten für unsere Visualisierung stammen aus dem BGB in einem strukturierten XML-Format von **Wolters Kluwer**. In diesem Format sind alle Schlüsselemente enthalten: Titel, Text und die Referenzen.

Wir nutzen in einem ersten Schritt einen einfachen Ansatz, indem wir verschiedene Farben verwenden, um die unterschiedlichen Bücher des BGB zu unterscheiden. Gleichzeitig nutzen wir auch bestimmte Farben, um Strukturknoten, wie bspw. Abschnittsüberschriften, zu kennzeichnen.

Schließlich verwenden wir den **ForceAtlas2**-Algorithmus, um die Lage der Knoten zu bestimmen. Die Idee hinter dem **ForceAtlas2**¹ Algorithmus ist es, ein physikalisches System zu simulieren, das Netzwerke räumlich zuordnet.

¹ Für mehr Informationen über die Funktionsweise des ForceAtlas2-Algorithmus, ist das Entwicklungspapier zu empfehlen, [hier](#) abrufbar (Stand: 20.02.2022).

Aus praktischer Sicht bedeutet dies, dass in der Mitte der Visualisierung mehrere Ansammlungen (Cluster) von verbundenen Normen zu sehen sind. Im Außenbereich der Visualisierung sammeln sich Einzelnormen oder kleine Cluster. Je mittlerer der Knotenpunkt (Norm) in der Darstellung also ist, desto öfter wird dieser von anderen Knotenpunkten referenziert und je näher die Knotenpunkte räumlich sind, desto öfter referenzieren sich Knotenpunkte, sodass Cluster entstehen (vgl. Abbildung 4).

Um die Visualisierung noch aussagekräftiger zu machen, können Parameter zur Festlegung der Anziehungskraft verändert werden. Je nachdem, wie der Parameter gewählt wird, wird eine gedrungene oder weitläufige Visualisierung erstellt. Auch kann es hilfreich sein, Überlappungen zu verhindern und Knotenpunkte (sog. Hubs) zu separieren, wodurch die Visualisierung deutlich lichter wird. Dieser Ansatz erzeugt wunderbare Visualisierungen, um ein grundsätzliches Gefühl für die Daten und die Verknüpfung der Normen zu erhalten. Gleichwohl ist es jedoch ungleich schwerer, hierdurch auch konkrete Fragen zu beantworten. Daher ist es praktisch unmöglich, eine einzige Visualisierung zu erstellen, mit der alle Fragen beantwortet werden können.

Um den Graph sinnvoll nutzen zu können, wird ein Programm benötigt, mit dem die Visualisierung einfach verändert und an konkrete Fragen angepasst werden kann.

Wir begannen unsere Suche nach einer solchen Lösung mit komplett auf **Python** basierenden Bibliotheken;² diese sind allerdings allesamt nicht interaktiv. Sodann wechselten wir zu Web-basierter Technologie wie **sigma.js** und **D3**. Allerdings waren diese auch nicht ausreichend genug, um einfache Veränderungen vorzunehmen; es musste jeweils neuer Code verfasst werden.

² Bibliothek meint hier eine Sammlung von extra Modulen, die für eine Programmiersprache von Dritten entwickelt wurde und die es Programmierern erlaubt, unkompliziert und einfach auf bereits existierende Funktionen zurückzugreifen und diese in ihr Programm zu inkorporieren.

Letztlich entschieden wir uns, Gephi zu verwenden. **Gephi**³ ist eine Open-Source Bibliothek mit einem sehr breiten Funktionsspektrum.

Im Verlauf des Artikels sind verschiedene Visualisierungen zu finden, welche wir zur Beantwortung spezifischer Fragen oder zur Ermöglichung neuer Perspektiven erstellt haben. Hier eine Übersicht über Tipps und Tricks, die wir verwendet haben, um diese Repräsentationen zu erstellen:

- Knotengröße: hilfreich, um wichtige Knoten hervorzuheben. Normalerweise visualisiert die Knotengröße die Anzahl der Beziehungen. Wir nutzen verschiedene Ansätze zur Berechnung der Knotengröße: der Knoten ist größer, wenn er mehr eingehende Beziehungen hat (in-Degree), der Knoten ist größer, wenn er mehr ausgehende Beziehungen hat (out-Degree) oder eine Kombination aus ein- und ausgehend (Degree).
- Knotenfarbe: eine großartige Möglichkeit, um bestimmte Knoten hervorzuheben. Wir haben dies bspw. durchgängig verwendet, um die verschiedenen Bücher des BGB zu visualisieren.
- Kantenfarbe: kann verwendet werden, um verschiedene Arten von Referenzen hervorzuheben. Wir verwenden lediglich eine Farbe zur Kenntlichmachung der Zitate der Normen. In manchen Visualisierungen haben wir die Kantenfarbe identisch mit der Knotenfarbe gewählt, um einfacher zu verdeutlichen, von wo die Beziehung ausgeht.
- Kantengewicht: kann hilfreich sein, um die Visualisierung zu vereinfachen. Wenn beispielsweise ein Knoten mehrere Referenzen zu einem anderen Knoten hat, dann kann dies einfach durch das Kantengewicht und bspw. eine stärkere Linie hervorgehoben werden.

- Filter: diese können auch helfen, die Visualisierung zu vereinfachen. Bspw. dann, wenn wir nur Knoten mit ganz bestimmten Eigenschaften betrachten wollen.

Somit ist es ganz einfach, einen Graphen zu erstellen. Allerdings gilt es zu beachten, dass zur Beantwortung besonders spezifischer Fragen viele Parameter verändert werden müssen, um eine gute Visualisierung zu erhalten. Betrachten wir nun ein paar Beispiele:

B. Die einzelnen Fragestellungen und ihre Visualisierung

I. Die Fragestellungen und ihre Bedeutung

Eine Datenvisualisierung erfolgt in der Regel mit einem bestimmten Ziel. Das Ziel unserer Datenvisualisierung ist es, Fragestellungen aus dem BGB nachzugehen. Diese Fragestellungen kann man teilweise bereits vorab formulieren oder aber sie ergeben sich im Laufe des Prozesses der Erstellung. In unserem Fall lassen sie sich in zwei Gruppen unterteilen: allgemeine und spezielle Fragestellungen. Die allgemeinen Fragestellungen betreffen das ganze BGB, während die spezifischen Fragestellungen nur ganz bestimmte Ausschnitte oder Bezüge betreffen.

Allgemeine Fragestellungen und Ausgangshypothesen:

- Gibt es Auffälligkeiten hinsichtlich der Verweise? Hypothese: bestimmt, aber wir haben zum Startzeitpunkt (noch) keine Ahnung, was es ist.
- Gibt es bestimmte andere Gesetze, die besonders häufig zitiert werden? Hypothese: ja, bspw. das Einführungsgesetz zum BGB.
- Gibt es bestimmte Bereiche des BGB, die besonders häufig zitiert werden? Hypothese: ja, vermutlich der vor die Klammer gezogene Allgemeine Teil.

³ [Hier](#) abrufbar (Stand: 29.11.2022).

- Gibt es bestimmte Normen des BGB, die besonders häufig zitiert werden? Hypothese: ja, vermutlich eine Norm aus dem Allgemeinen Teil.

Spezifische Fragestellungen:

- Gibt es Normen, die als Zwischenschritt ganze Bücher des BGB verbinden?
- Gibt es Normen, die eine Vielzahl anderer Normen verbinden?
- Gibt es abgeschlossene Bereiche, die nur untereinander referenzieren?

Dies stellt nur eine Auswahl an Fragen dar, mit denen wir gestartet sind.

II. Allgemeine Fragestellungen

1. Gibt es Auffälligkeiten?

Die erste Visualisierung zeigt das BGB in seiner Gesamtheit. Visualisiert sind nur Verbindungen zwischen Normen innerhalb des BGB. Verschiedene Bücher sind farblich unterschiedlich markiert. Wenn es eine Verbindung zwischen den Büchern gibt, so wird sie durch eine Linie angezeigt. Auffällig sind die beiden Cluster in der Mitte. Insbesondere das lila Cluster hat einige sehr lange Verkettungen von Normen, die von Mitte unten (lila) bis Oben rechts (blau) reichen.

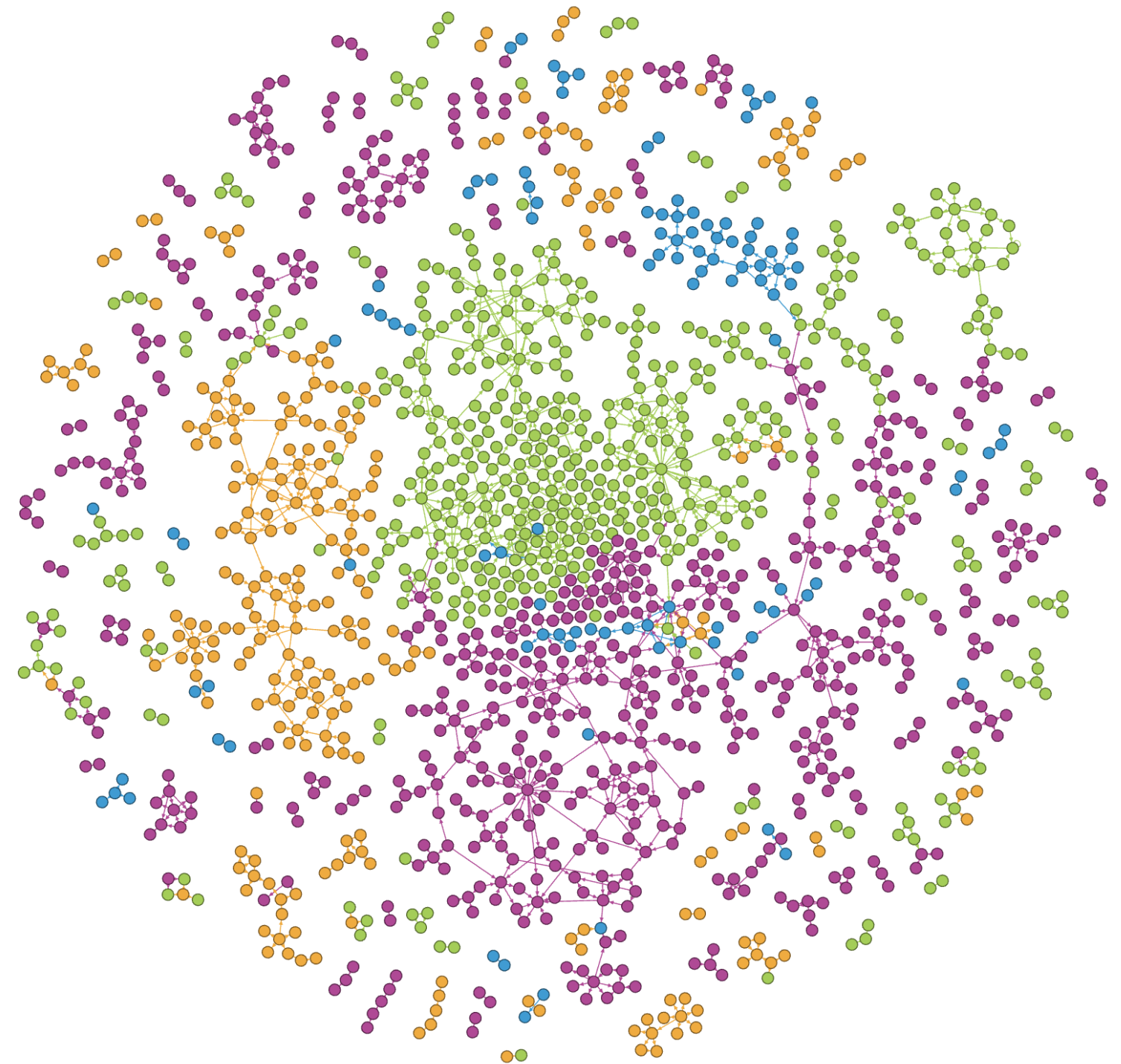


Abb. 4: Visualisierung der Cluster im BGB

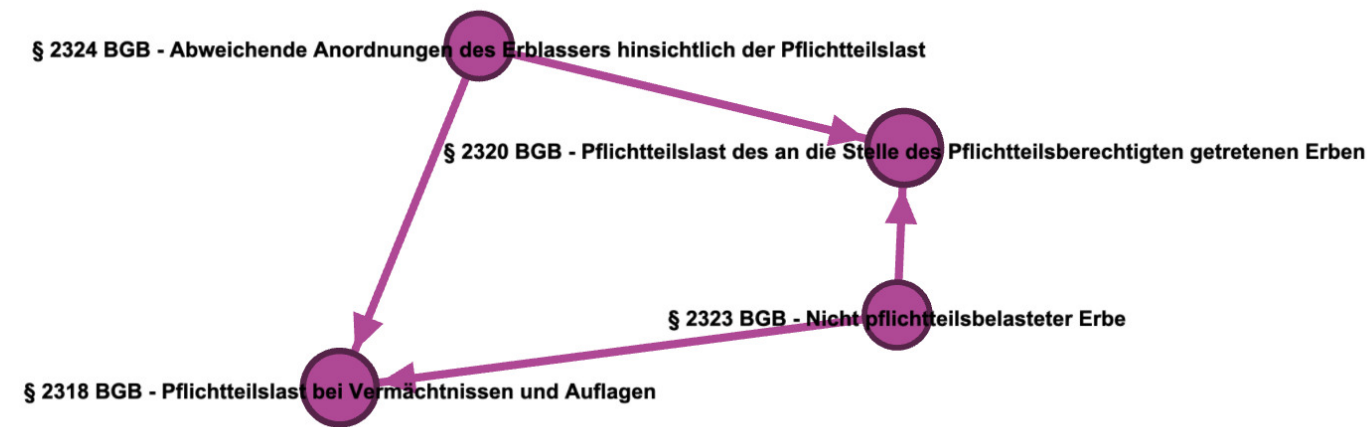


Abb. 5: Hier sieht man ein solches „Zitierviereck“.

Ein weiteres Beispiel für Auffälligkeiten sind Unter-Netzwerke, die sich mehr oder weniger selbst zitieren, bzw. eine Art Viereck bilden. Auf den ersten Blick sieht dies wie ein Zirkelschluss aus. Betrachtet man allerdings die Richtung der Zitierung, erscheint es verständlicher. Auch im folgenden Beispiel sieht die Zitierung auf den ersten Blick verwirrend aus:

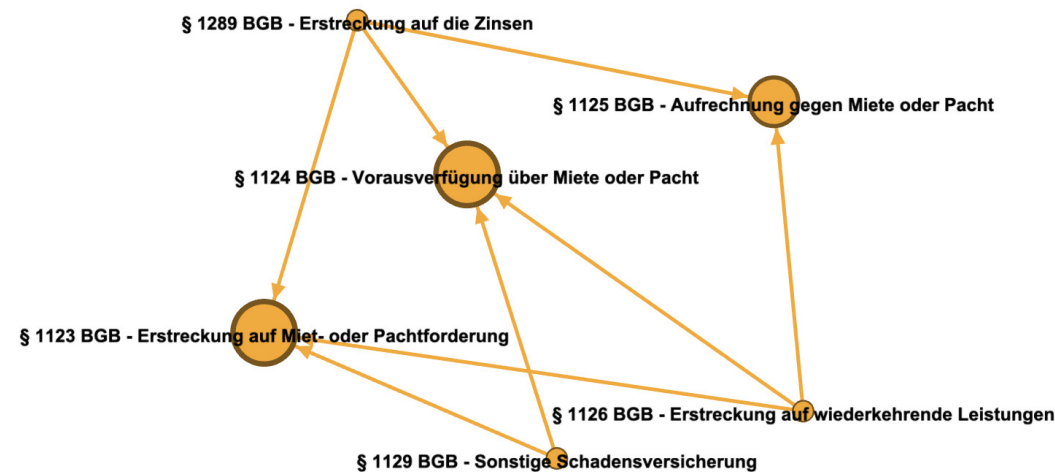


Abb. 6: Verschiedene Normen, die sich scheinbar selbst zitieren.

2. Gibt es bestimmte andere Gesetze, die besonders häufig zitiert werden?

In der folgenden Visualisierung sind nur die externen Verbindungen visualisiert, also alle Vorschriften, die eine Norm außerhalb des BGB zitieren. Die Gruppen, die hier zu sehen sind, sind Normen, die gemeinsam bestimmte Gesetze zitieren.



Abb. 7: Hervorhebung externer Gesetze in Rot

Der große rote Punkt Mitte rechts ist ein Gesetz, das besonders oft zitiert wird: das EGBGB. Damit hat sich die Hypothese bestätigt. Auch Zivilprozessordnung und Grundbuchordnung, die auf Platz 2 und 3 der meist-zitierten anderen Gesetze liegen, sind verständlich. Platz 4 für das Verschollenheitsgesetz⁴ hat uns allerdings doch etwas überrascht.

Nachfolgende Darstellungen zeigen proportional die Häufigkeit der Zitierungen.

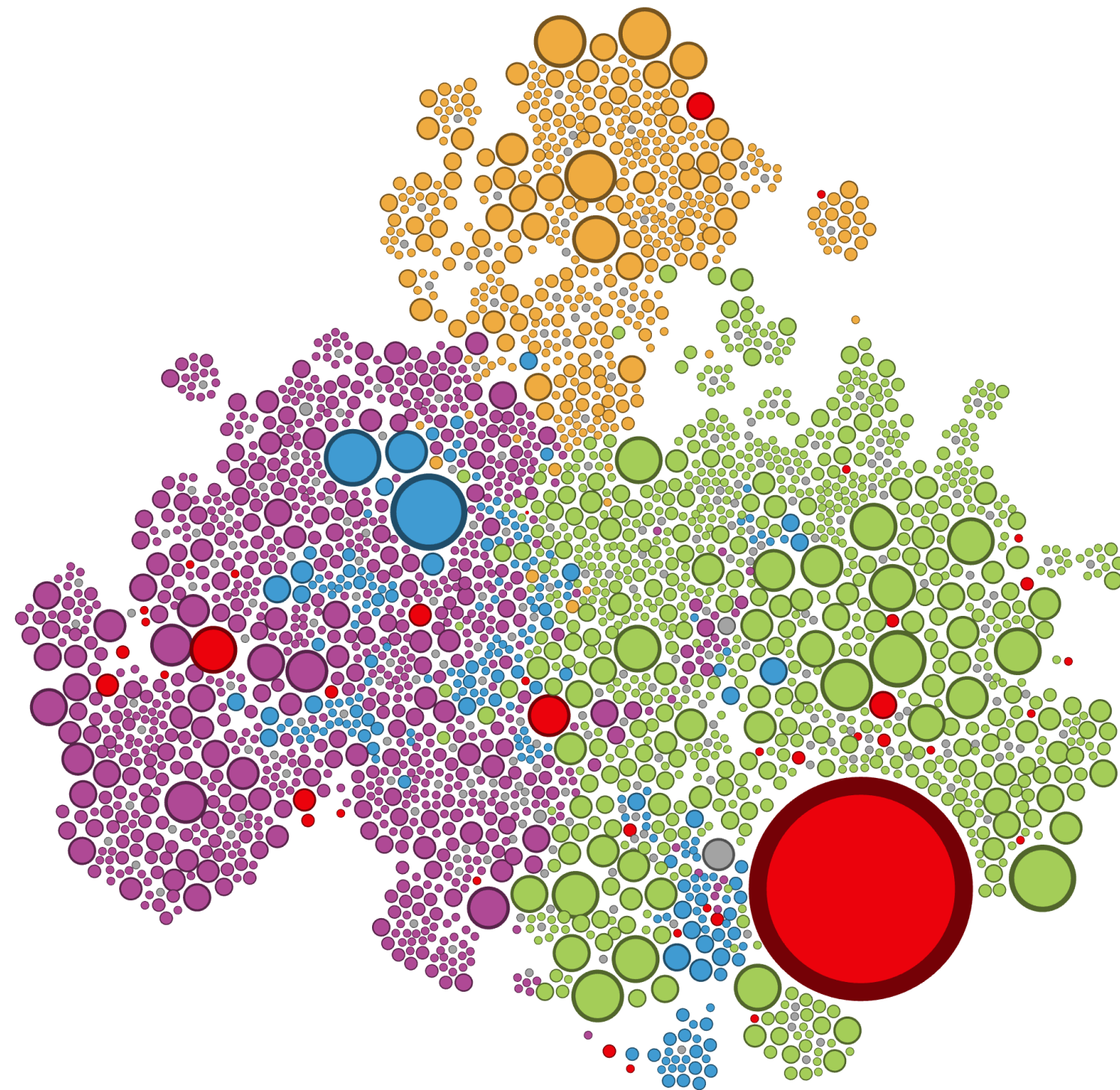


Abb. 8: Darstellung der Menge interner und externer Verweise (rot).

⁴ Das VerschG ist ein im Erbrecht häufiger relevantes Gesetz. Es regelt, wann eine Person als verschollen und damit rechtlich als tot gelten kann, obwohl man ihren Tod nie positiv feststellen konnte.

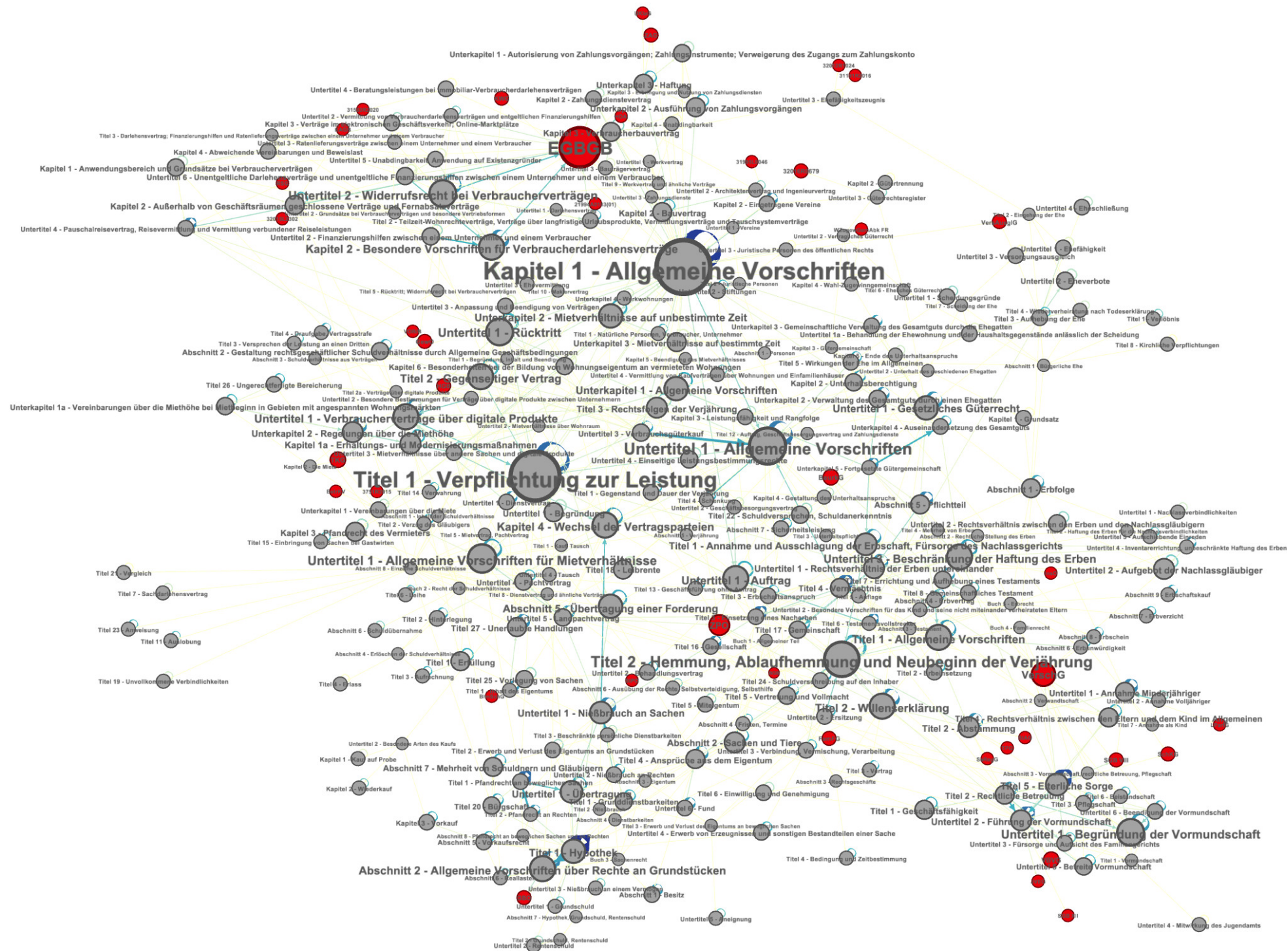


Abb. 9: Relative Darstellung der internen Verweise (graue Knoten) und externen Verweise (rote Knoten)



Data Science meets BGB: Eine Einführung in die juristische Datenvisualisierung

Diese Visualisierung betrifft die einzelnen strukturellen Knoten und visualisiert die Häufigkeit der Zitierung anderer Gesetze, wobei der Pfeil dicker wird, wenn bestimmte Gesetze von bestimmten Teilen häufiger zitiert werden (sog. Gewichtung):

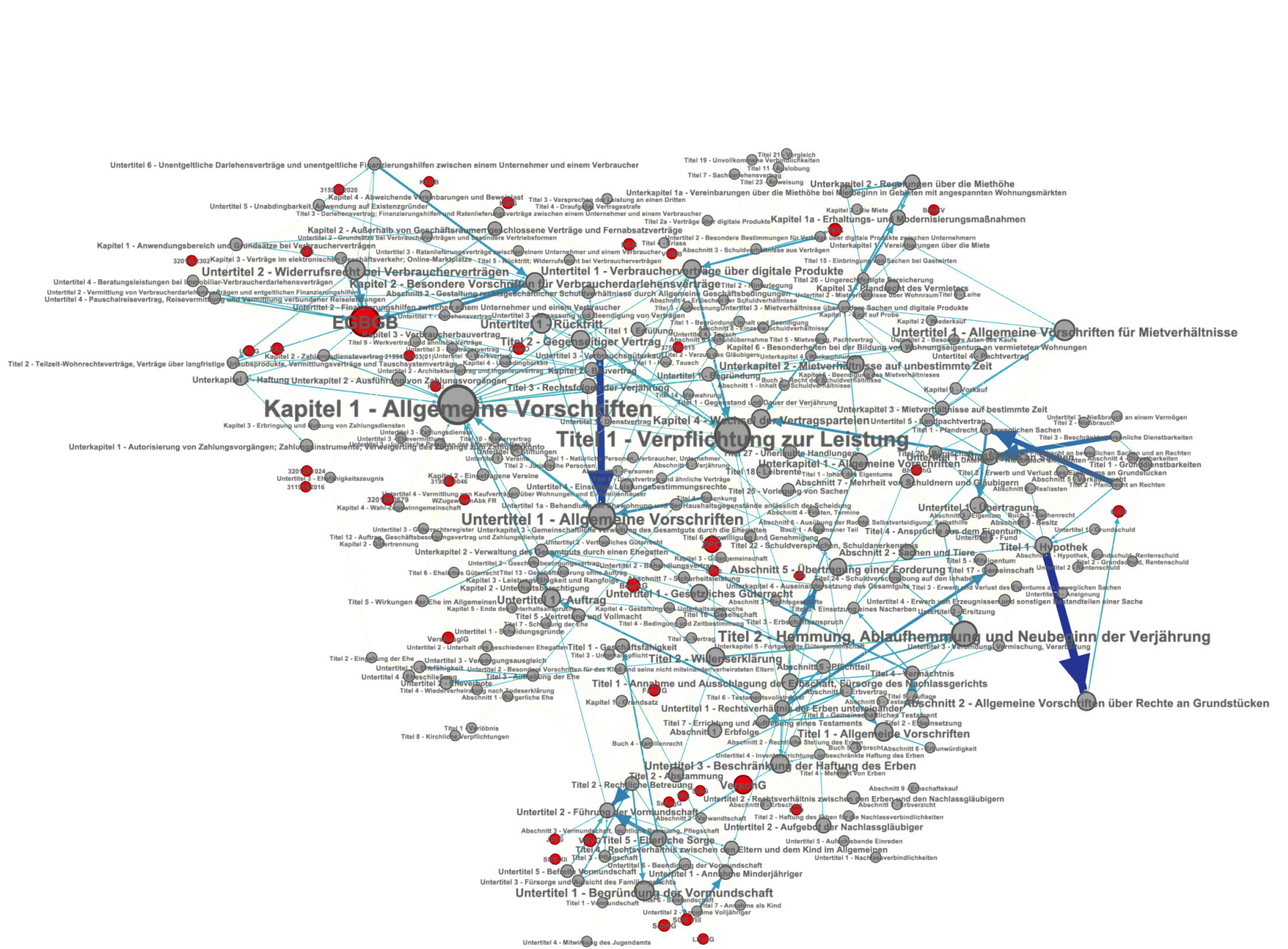


Abb. 10: Nutzung des Kartengewichts zur Darstellung mehrfacher Zitierungen einer Norm.



3. Gibt es bestimmte Bereiche des BGB, die besonders häufig zitiert werden?

In der folgenden Visualisierung wurden die eingehenden Zitierungen untersucht und je häufiger eine Norm aus einem Bereich des BGB zitiert wurde, desto größer wird der Knoten dargestellt.

Auch hier hat sich die Hypothese bewahrheitet, dass der Allgemeine Teil des BGB, insbesondere die Allgemeinen Vorschriften deutlich häufiger zitiert werden. Das „Vor-Die-Klammer-Ziehen“ im BGB ist also auch in der Datenvisualisierung klar erkennbar.

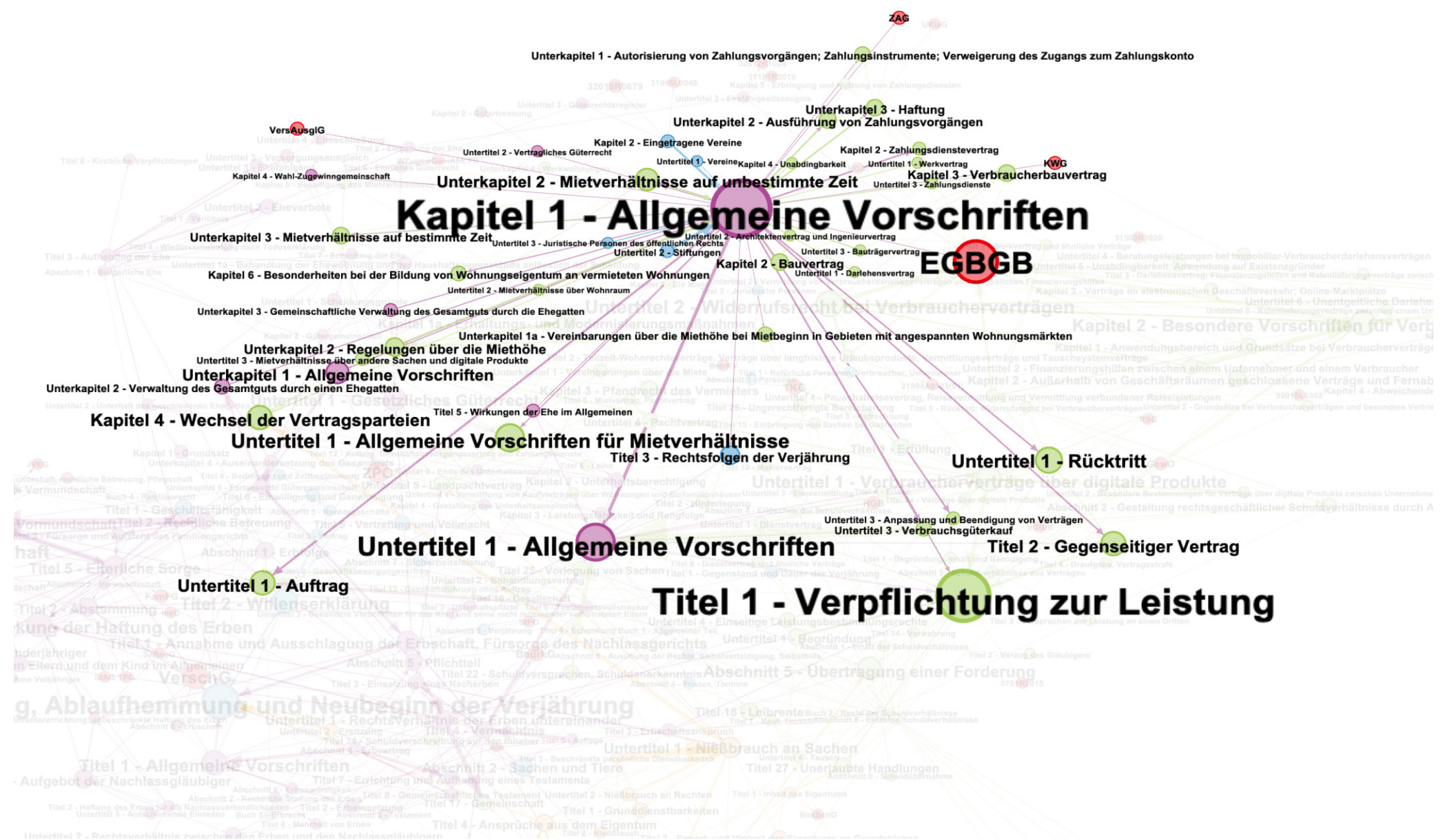


Abb. 11: Strukturelle Elemente und deren Zitierung; hier im BGB AT in Lila.

Im Vergleich hierzu der deutlich kleinere Knoten der §§ 203 ff. BGB:

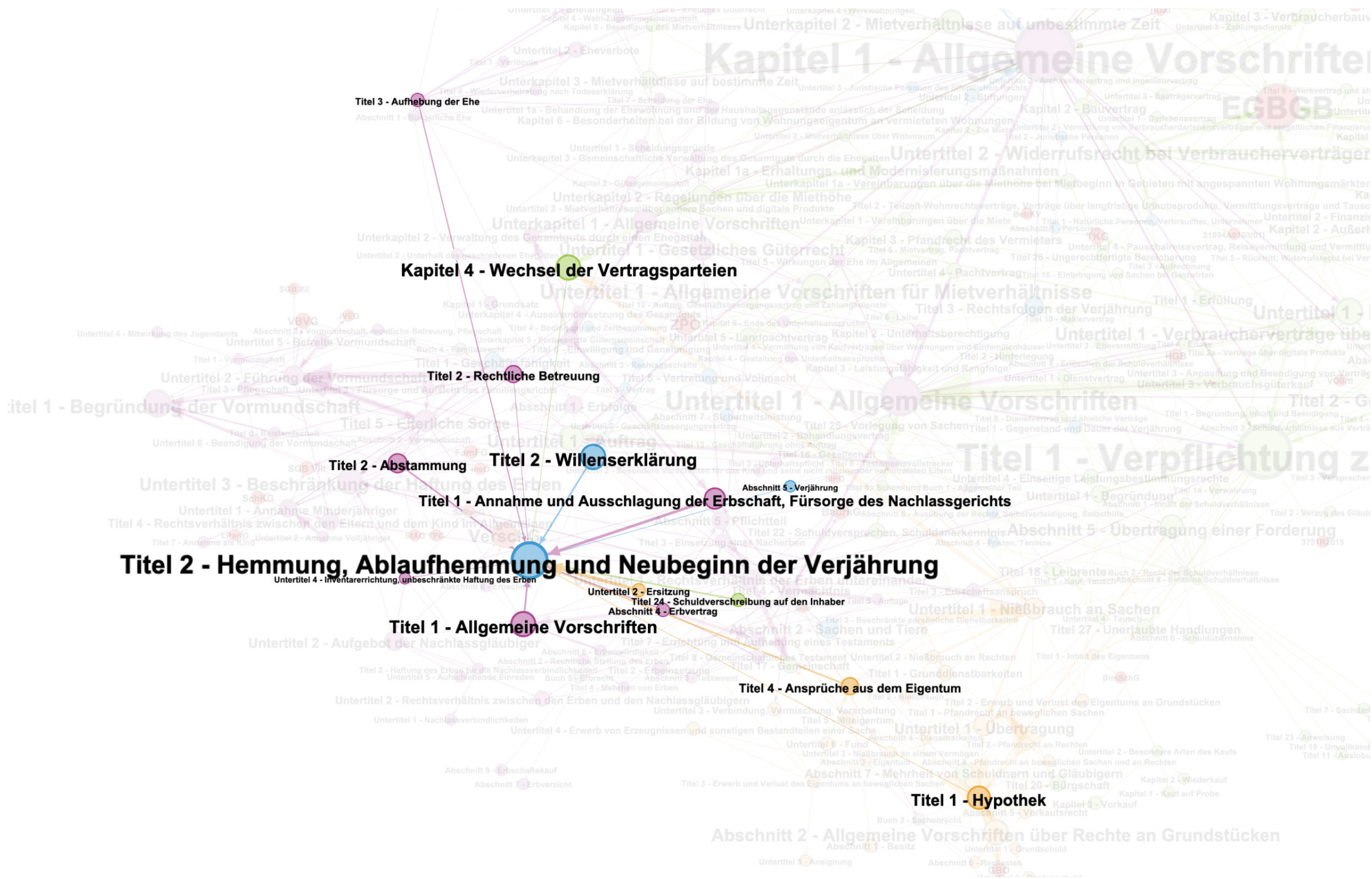


Abb. 12: Visualisierung der §§ 203 ff. BGB.



4. Gibt es bestimmte Normen des BGB, die besonders häufig zitiert werden?

Die mit am meisten zitierte Norm ist § 210 BGB, was man an den ganzen eingehenden Verbindungen sieht:

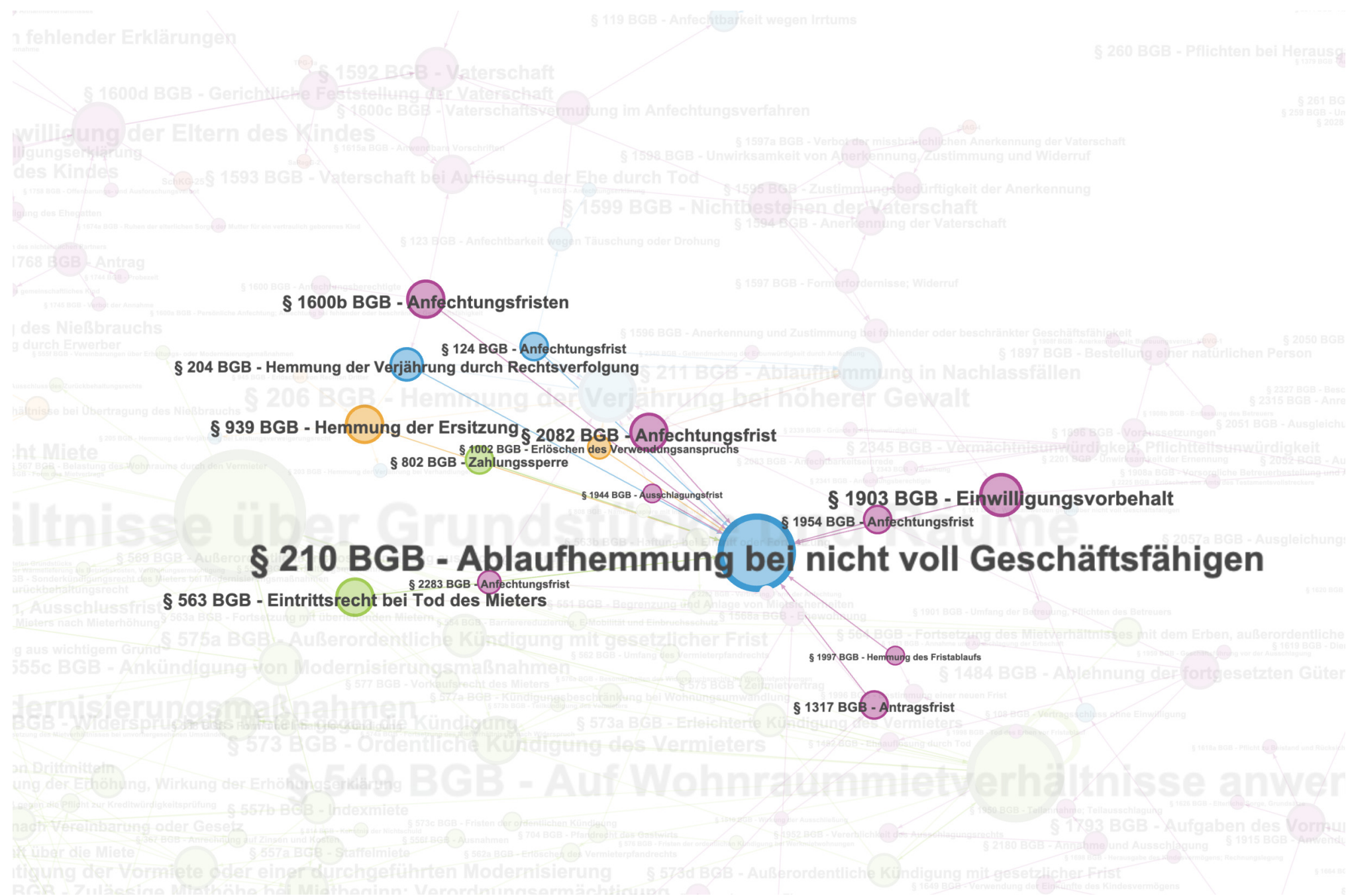


Abb. 13: Visualisierung der am häufigsten referenzierten Norm, § 210 BGB.



Umgekehrt ist wohl § 578 BGB eine der Normen, welche die meisten anderen Normen zitiert:

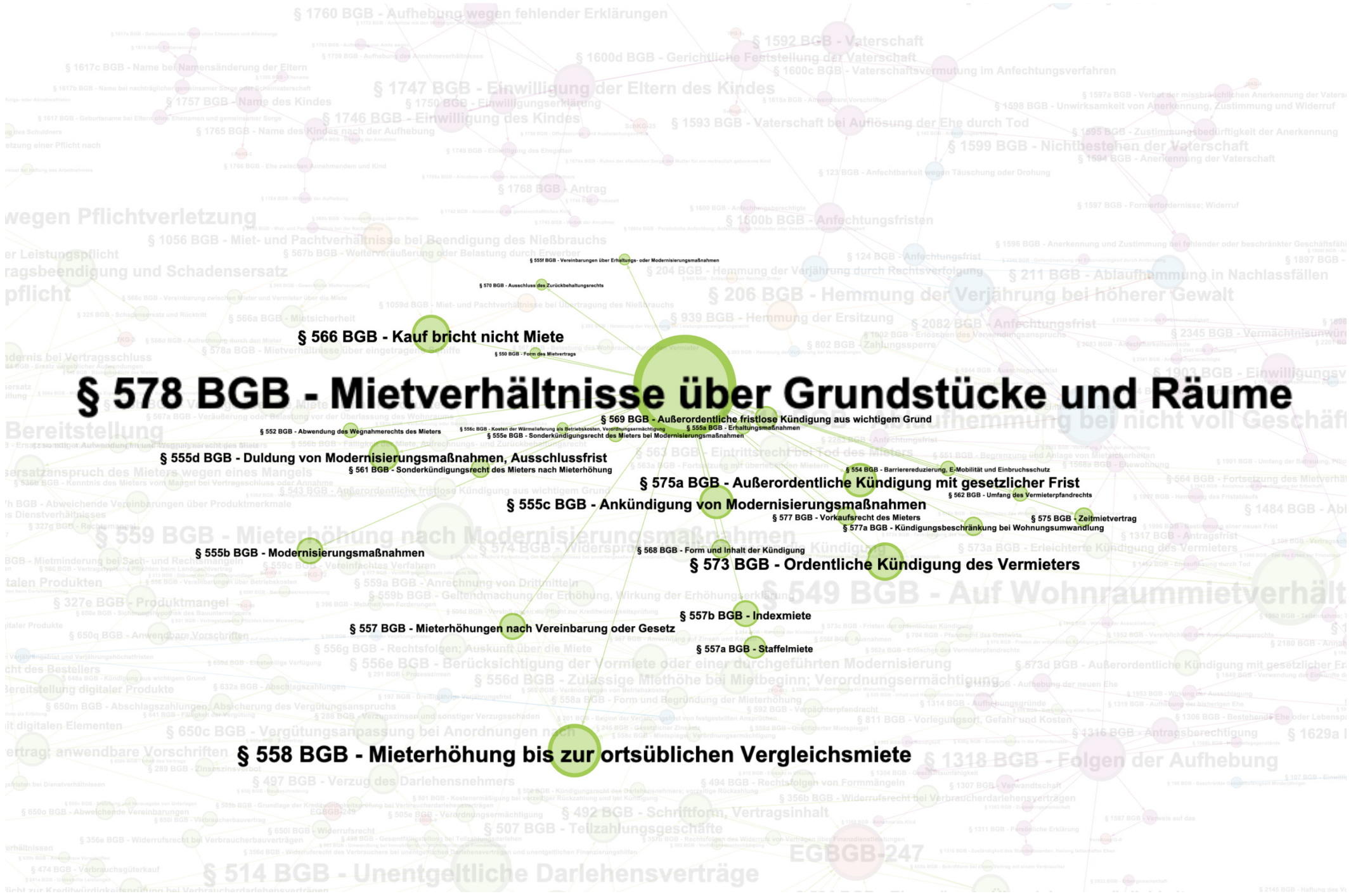


Abb. 14: § 578 BGB referenziert die meisten anderen Normen.



III. Spezifische Fragestellungen

1. Gibt es Normen, die als Zwischenschritt ganze Bücher des BGB verbinden?

Dieser Graph zeigt alle Verbindungsnormen, die mindestens zwei verschiedene Bücher des BGB verbinden. Die Bücher sind jeweils wieder in unterschiedlichen Farben dargestellt. Hier sind verschiedene Möglichkeiten denkbar, wie diese Verbindung von Büchern im BGB zustande kommen: Norm A (aus Buch 1) zitiert Norm B (aus Buch 2), diese wiederum zitiert Norm C (aus Buch 1 oder 3). Norm A (aus Buch 1) zitiert Norm B (aus Buch 2), die auch von Norm C (aus Buch 3) zitiert wird.

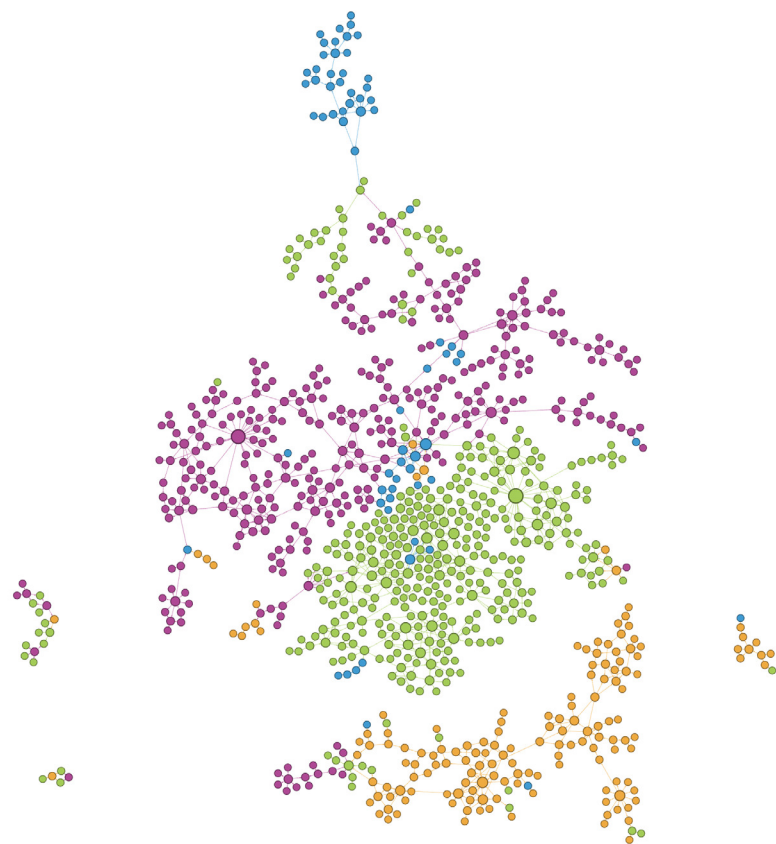


Abb. 15: Normen, die verschiedene Bücher miteinander verknüpfen

Abbildungen 15-17 zeigen hierbei Beispiele für Variante 1:

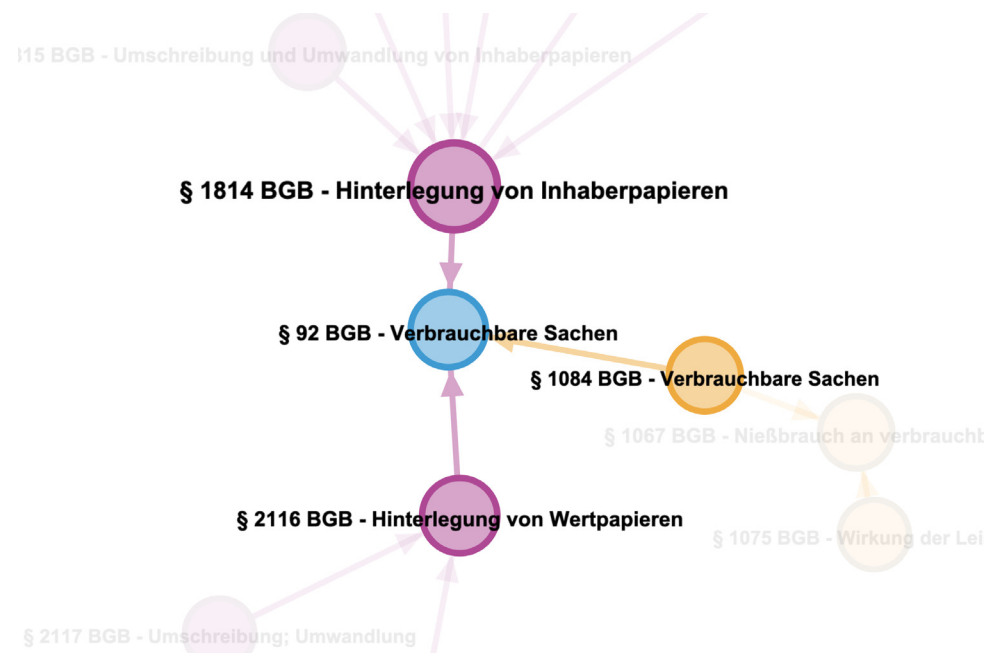


Abb. 16: Beispiel 1 für Variante 1.

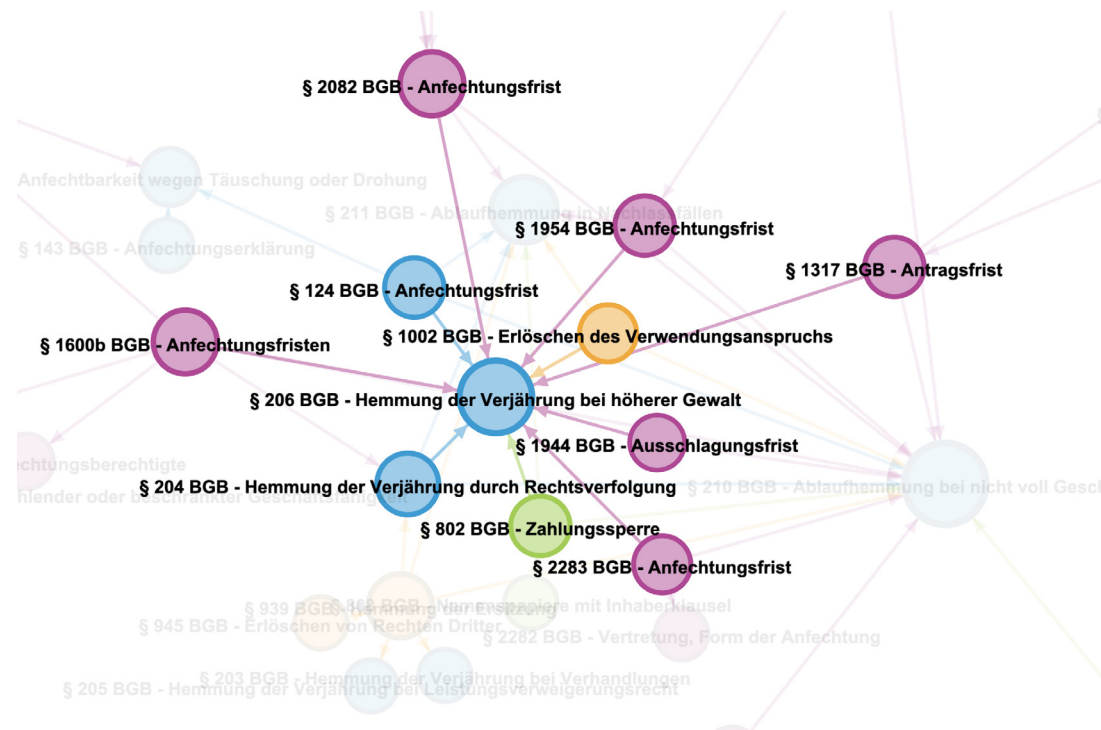


Abb. 17: Beispiel 2 für Variante 1.



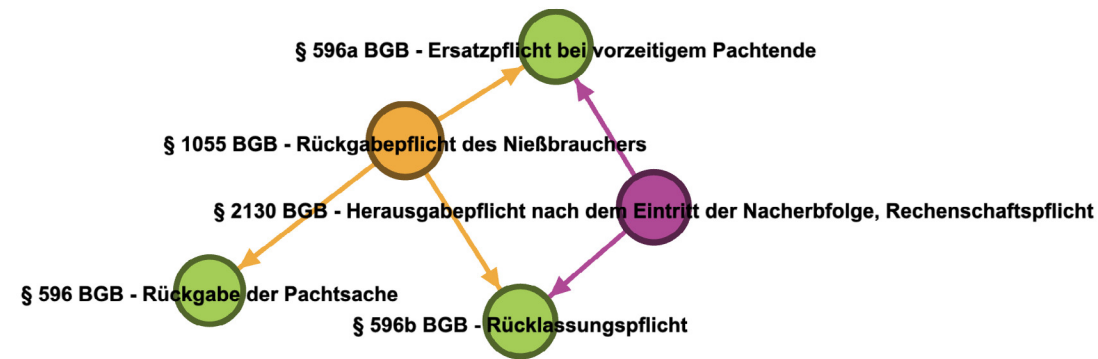


Abb. 18: Beispiel 3 für Variante 1.

Das folgende Bild zeigt ein Beispiel für Variante 2. So lässt sich ein Pfad von §§ 997, 994 BGB zu § 989 BGB erkennen.



Abb. 19: Beispiel für Variante 2.

2. Gibt es Normen, die eine Vielzahl anderer Normen verbinden?

Das Beispiel des § 578 BGB haben wir bereits oben angesprochen. Diese Norm zitiert eine Vielzahl von anderen Normen. Hier noch einmal übersichtlich die Vielzahl solcher Normen, wobei § 578 BGB hier ganz zentral als Knoten (in grüner Darstellung in der Mitte des Netzwerkes) sichtbar ist:

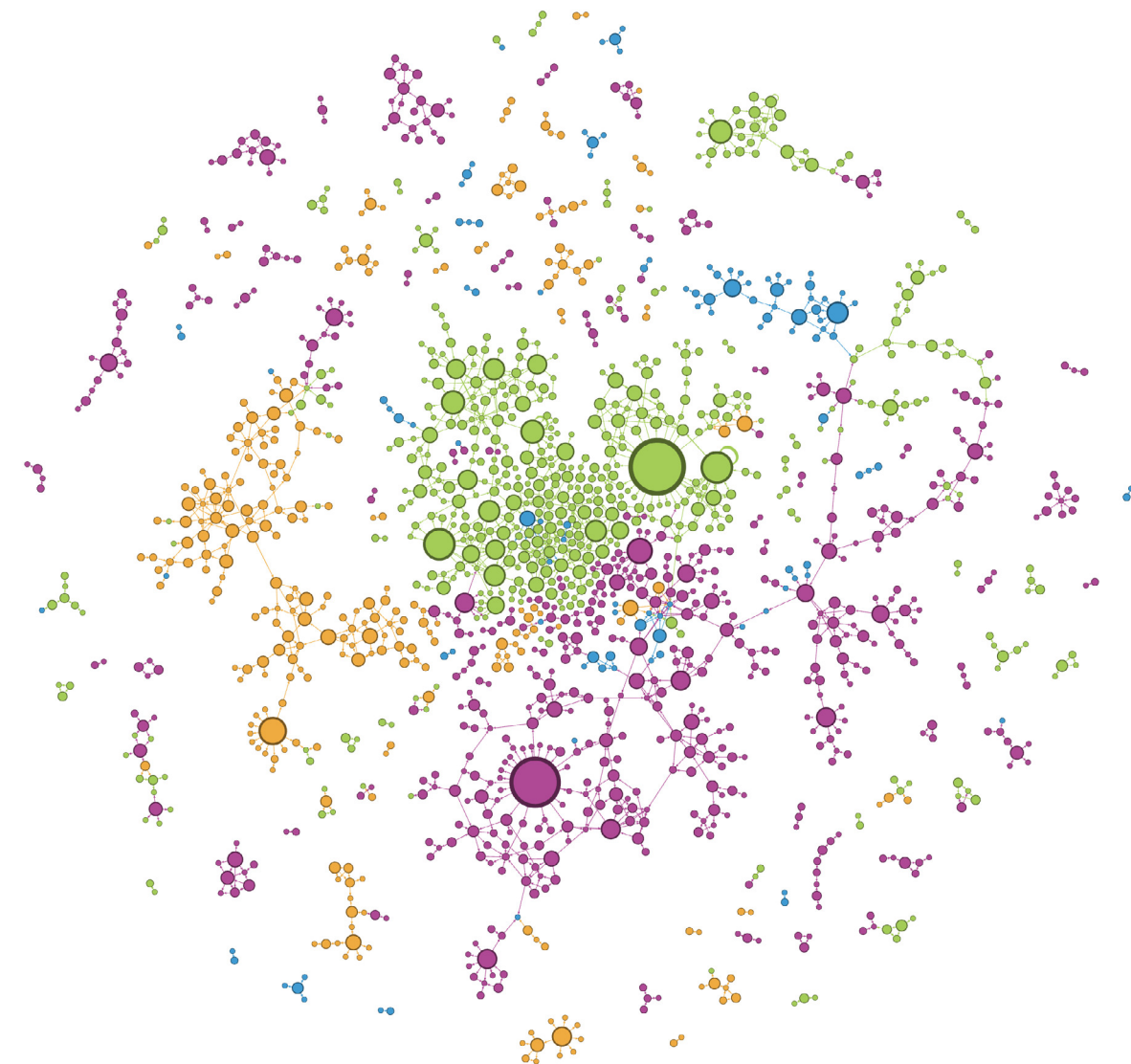


Abb. 20: Visualisierung derjenigen Normen, welche die meisten anderen Normen referenzieren.

3. Gibt es abgeschlossene Bereiche, die nur untereinander referenzieren?

Ja, es gibt mehrere solcher Bereiche. Beispielhaft seien die §§ 1570 ff. BGB genannt. Lediglich als eingehende Zitierung wird § 1570 BGB zusätzlich noch von § 1586a BGB zitiert.

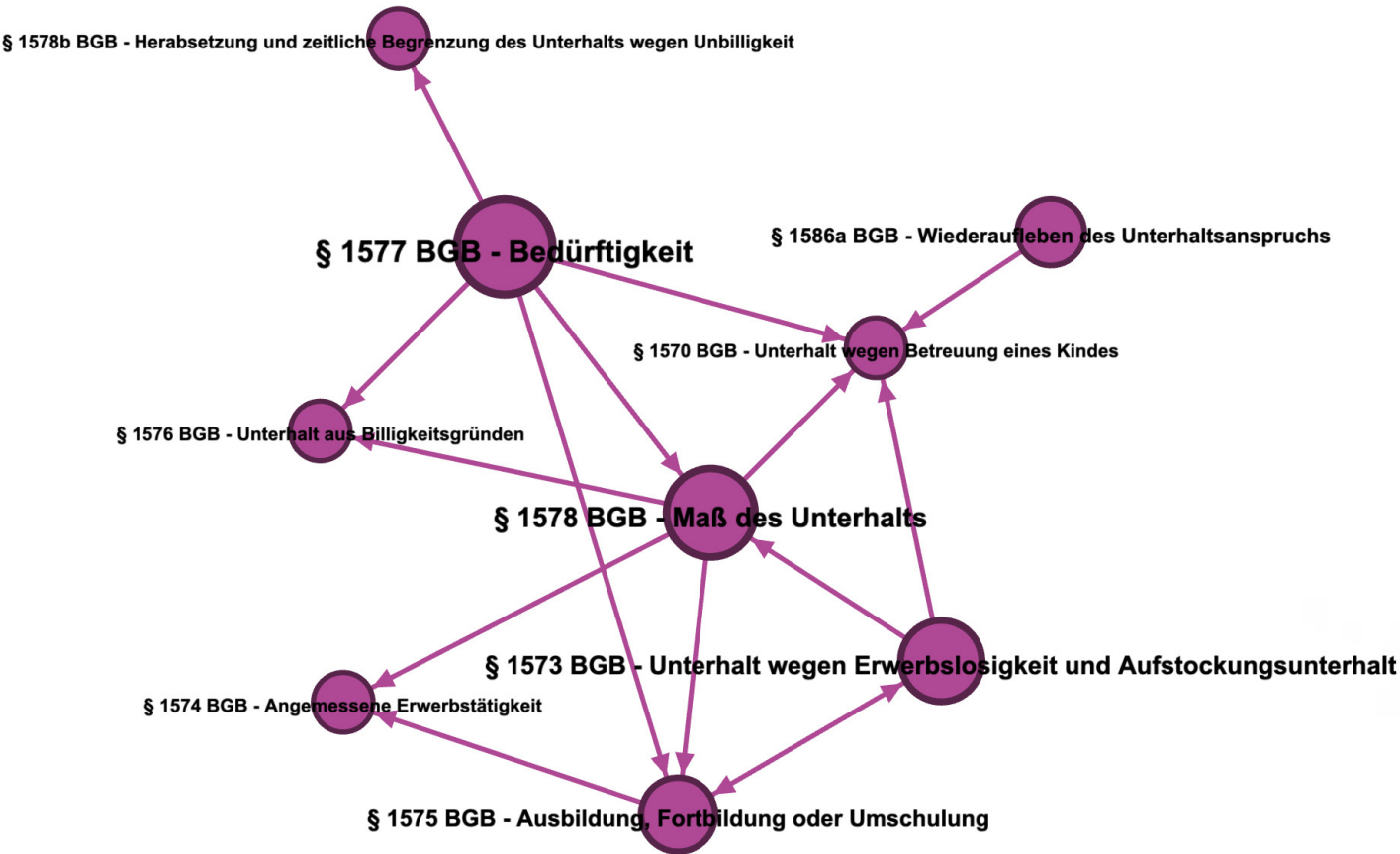


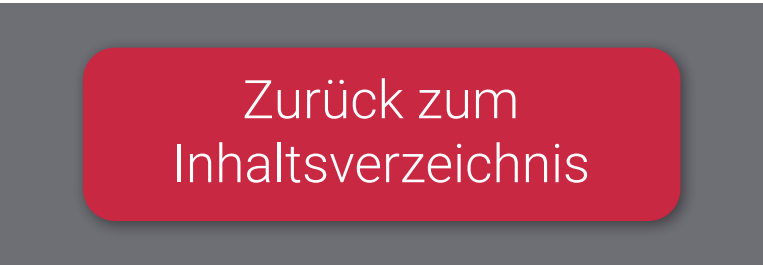
Abb. 21: Beispiel für einen abgeschlossenen Bereich, der sich nur selbst referenziert.

C. Ergebnis

Auch wenn dieser Beitrag lediglich einen kleinen Einstieg in die Datenvisualisierung anhand des BGB geben kann, so gehen wir doch von einem hohen Potenzial für zukünftige Entwicklungen aus.

Anwendungsfälle in der Praxis können nicht nur im Rahmen der Gesetzgebung sein, um etwa fehlerhafte Zitierungen, Zirkelschlüsse oder ähnlich problematische Veränderungen leicht erkennbar zu machen. Eine solche Darstellung kann auch Jurist:innen in der täglichen Arbeit unterstützen. Würde man alle möglichen Anspruchsgrundlagen für einen Fall visualisieren, könnte man hiermit Gemeinsamkeiten und Unterschiede leicht erkennbar machen. Aber auch in der juristischen Ausbildung können solche Visualisierungen helfen. Wie leicht wäre es, zu erklären, wie die Zusammenhänge im BGB nun wirklich sind, wenn man diese einfach visualisieren könnte, statt sie nur im Text markieren zu können.

Wir hoffen, mit diesen Beispielen Anregungen für die Verwendung von Datenvisualisierung auch für Jurist:innen gegeben zu haben. Vielleicht wird sie sich ja auch im juristischen Umfeld etablieren.





Hier geht's zur ganzen Ausgabe!

Was das BGB mit Data Science und das StGB
mit Deepfakes zu tun hat und noch vieles mehr
in 12 spannenden Beiträgen!

