

Untersuchung, Implementierungen und Bewertung von Graph-Metriken

Studienarbeit

im Studiengang Informatik

an der Dualen Hochschule Baden-Württemberg Stuttgart, Campus Horb am Neckar

von

Benedict Weichselbaum

12. Januar 2021

Bearbeitungszeitraum
Matrikelnummer, Kurs
Betreuer & Gutachter

28.09.2020 - 31.05.2021
6275457, TINF2018
Prof. Dr. ing. Olaf Herden

Erklärung

Ich versichere hiermit, dass ich meine Studienarbeit mit dem Thema *Graphen: Metriken und Ähnlichkeit* selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Ich versichere zudem, dass die eingereichte elektronische Fassung mit der gedruckten Fassung übereinstimmt.

Nürnberg, 12. Januar 2021

Benedict Martin Weichselbaum

Abstract

Inhaltsverzeichnis

Abbildungsverzeichnis	I
Tabellenverzeichnis	II
Abkürzungsverzeichnis	III
1 Einleitung	1
1.1 Motivation für die Studienarbeit	1
1.2 Fragestellungen	1
2 Graph-Metriken	3
2.1 Grundlegende Metriken	3
2.2 Distanzmetriken	6
2.3 Kreis-basierte Metriken	8
2.4 Zusammenhangsmetriken (Connectivity)	9
2.5 Zentralitätsmetriken	12
2.6 Arborizität	12
2.7 Chromatische Zahl und chromatischer Index	12
2.8 Weitere Metriken	12
2.9 Übersicht der vorgestellten Graphmetriken	12
3 Ähnlichkeit von Graphen	13
4 Implementierung und Umsetzung der Metriken	14
4.1 Implementierung in verschiedenen Graphdatenbanken	14
4.2 Vergleich der Implementierungen	14
5 Graphmetriken und Ähnlichkeit in Anwendung	15
6 Fazit und Zusammenfassung	16
6.1 Zusammenfassung der Ergebnisse	16
6.2 Fazit	16
Glossar	17
Literatur	18

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

1 Einleitung

1.1 Motivation für die Studienarbeit

Graphen sind einer der wichtigsten Datenstrukturen der Informatik. Warum kann man das sagen? In seinem Buch „Algorithmische Graphentheorie“ nennt Volker Turau, Professor an der Universität Hamburg-Harburg, den Grund dafür:

Graphen sind die in der Informatik am häufigsten verwendete Abstraktion. Jedes System, welches aus diskreten Zuständen oder Objekten und Beziehungen zwischen diesen besteht, kann als Graph modelliert werden.

[Tur04]

Diese netzartigen Strukturen können dabei die verschiedensten Konstrukte repräsentieren. Dazu zählen Straßennetze, Computernetzwerke, elektrische Schaltungen aber auch zum Beispiel chemische Moleküle. [Tit19]

Um Graphen zu beschreiben und zu charakterisieren, haben sich über die Zeit zahlreiche Metriken, bzw. Eigenschaften für diese herausgebildet („graph properties“ [Lov12]). Das heißt, einem Graphen können gewisse Kennzahlen zugeordnet werden, die ihn auszeichnen. Auch diese Metriken sind, wie die Graphen selbst, meist praktisch anwendbar. Zum Beispiel in der Untersuchung von Netzwerken [EK13].

Diese Studienarbeit soll nun diese Metriken genauer untersuchen. Hierbei ist es zunächst wichtig die verschiedenste Metriken vorzustellen und zu erläutern. Dabei ist es auch wichtig herauszufinden, wie verbreitet diese Metriken sind und inwieweit die jeweiligen Kennzahlen zu bewerten sind. Des Weiteren soll auf Basis der Metriken auch der Begriff der Ähnlichkeit von Graphen aufgegriffen werden.

Neben einer theoretischen Betrachtung der Graphmetriken soll auch eine Implementierung stattfinden. Es ist dabei das Ziel, mithilfe von Graphdatenbanken die jeweiligen Metriken umzusetzen und diese miteinander zu Vergleichen.

In einem Weiteren Teil ist außerdem noch darauf einzugehen, welche Anwendung die gezeigten Metriken haben, um den praktischen Nutzen der Thematik aufzuzeigen.

1.2 Fragestellungen

Auf Basis dieser Motivation können nun auch die konkreten Fragestellungen formuliert werden, die diese Arbeit betrachten soll. Insgesamt sollen vier wissenschaftliche Fragen

beantwortet werden.

1. Welche Graph-Metriken gibt es und wie sind diese zu ermitteln und zu kategorisieren?

Hierzu gehört, wie bereits erwähnt die Vorstellung der einzelnen Metriken, aber auch eine Kategorisierung in Rubriken, um Metriken besser voneinander abzugrenzen, da diverse Metriken höchst unterschiedliche Aussagen über einen Graphen treffen. Es wird auch darauf eingegangen welche Motivation hinter den jeweiligen Metriken steht. Bei der Beantwortung dieser Frage soll außerdem auch darauf eingegangen werden, inwieweit die beschriebene Metrik in bestimmten Mathematikbibliotheken wie „Sage Math“ oder „Wolfram“ vorkommen.

2. Wie sind die vorgestellten Metriken zu bewerten?

In diesem Abschnitt soll es vor allem darum gehen, die vorgestellten Metriken dahingehend zu bewerten, wie „schwer“ es ist, sie zu ermitteln. Außerdem soll bei der Bewertung auch auf die Verbreitung eingegangen werden.

3. Was beschreibt der Ähnlichkeitsbegriff bei Graphen?

Basierend auf Graph-Metriken lässt sich auch ermitteln, ob zwei Graphen Ähnlichkeiten aufweisen [WM19]. Auch auf diesen Aspekt soll die Arbeit Bezug nehmen und dabei ein Anwendungs-Beispiel konstruieren.

4. Wie können die vorgestellten Metriken in Graphdatenbanken verwendet werden, bzw. implementiert werden?

Auf die theoretische Betrachtung der Graph-Metriken folgt dann ein praktischer Teil, der behandeln soll, wie sich die Metriken in bekannten Graphdatenbanken umsetzen lassen, bzw. umgesetzt wurden. Dabei ist es wichtig herauszufinden welche Graphkennzahlen bereits teil der Graphdatenbank-Lösungen sind, bzw. welche Metriken selbst umgesetzt werden müssen.

5. Wie sind die jeweiligen Implementierungen zwischen und innerhalb der Graphdatenbanken zu bewerten?

Folgend auf die Implementierung, ist es noch wichtig zu verstehen, wie diese Umsetzungen zu betrachten sind. Dabei wird vor allem ein Fokus auf das Thema Performance und Skalierung gelegt.

6. Welche Anwendungen gibt es für Graph-Metriken und den Vergleich von Graphen (Ähnlichkeit)?

Als letztes soll sich die Studienarbeit mit praktischen Beispielen beschäftigen. Es ist dabei wichtig zu verstehen, welchen konkreten Nutzen die gezeigten Kennzahlen für Graphen in modernen Anwendungsszenarien haben.

2 Graph-Metriken

Dieser erste Teil der Arbeit wird sich nun ausführlich mit einer weiten Reihe an Graph-Metriken beschäftigen. Hierbei sollen die ersten zwei Fragestellungen der Arbeit genau beantwortet werden. Zur jeweiligen Vorstellung einer Graph-Metrik sollen dabei die folgenden Punkte erläutert werden:

- Was drückt die Metrik aus (Definition)?
- Wie ist die Metrik im Bezug auf den Rechenaufwand zu bewerten?
- Inwieweit ist die Metrik verbreitet? Zum Beispiel in der Literatur oder in Mathematikbibliotheken.
- Was ist die konkrete Motivation für die Metrik, falls diese auszumachen ist?

Es ist noch zu erwähnen, dass alle im folgenden vorgestellten Metriken über die einzelnen Sektionen der Arbeit in Kategorien eingeteilt sind.

Darüber hinaus ist noch eine grundsätzliche Notationen während der Arbeit zu klären: Ein **Graph G** ist ein Paar bestehend aus **Knoten V** und **Kanten E**.

$$G = (V, E), \text{ wobei } E \subseteq V \times V$$

Für V können wir auch $V(G)$ schreiben, für E auch $E(G)$. [Die00] V ist dabei Englisch und bedeutet „Vertices“, E steht für „Edges“. Wenn es um die Datenstrukturen von Graphen geht, kommen im Rahmen dieser Arbeit hauptsächlich Adjazenzmatrizen und Adjazenzlisten zum Einsatz. Allerdings können bei Bedarf auch Inzidenzen (Beziehung zwischen Knoten und Kanten) im Graphen eine Rolle spielen, wie Inzidenzmatrizen und Inzidenzlisten. [Kne19; Die00]

2.1 Grundlegende Metriken

Ein einem ersten Teil sollen grundlegende Graph-Kennzahlen vorgestellt werden. Diese beschreiben einen Graphen auf rudimentäre Art und Weise und zeigen die am einfachsten zu berechnenden Eigenschaften des Graphen.

Ordnung und Größe eines Graphen

Die Frage danach, wie viele Knoten ein Graph hat lässt sich mit der „**Ordnung**“ eines Graphen beantworten. Die **Ordnung** beschreibt dabei einfach die Anzahl der Elemente

in der Menge V . Man schreibt: $|V|$ oder $|V(G)|$ oder auch $|G|$. [Die00] Diese Eigenschaft ist essentiell zur allgemeinen Beschreibung und z.B. graphischen Darstellung eines Graphen. Sie lässt sich dabei in sämtlichen mathematischen Bibliotheken finden, wie SageMath, Matlab und Wolfram [Sag20; Mat20a; Wol20a]. Die Komplexität zur Erfassung der Metrik gestaltet sich dabei äußerst einfach. Bei einer Adjazenzmatrix lässt sich die Anzahl der Knoten dadurch herausfinden, wie „lang“ eine Dimension des zweidimensionalen Arrays bzw. der zweidimensionalen Liste. Dies kann man je nach Implementierung der jeweiligen Datenstruktur in einer Komplexität von $O(n)$ oder $O(1)$ herausfinden.

Eine weitere grundlegende Kennzahl von Graphen ist dessen „**Größe**“. Die Größe beschreibt dabei die Anzahl der Kanten, die im Graphen vorkommen, also die Anzahl der Elemente in der Menge E . Man schreibt analog zur Größe des Graphen: $|E|$ oder $|E(G)|$ oder auch $||G||$. [Bal97; Die00] Auch diese Metrik ist weit verbreitet. So lässt sie sich in vielen Büchern zur Graphentheorie finden, aber auch in den genannten Mathematikbibliotheken [Sag20; Mat20a; Wol20a]. Die Anzahl der Kanten innerhalb eines Graphen herauszufinden, erweist sich nicht ganz so trivial wie das Herausfinden der Ordnung. Ist ein Graph nicht gerichtet, d.h. seine Kanten haben keine feste Richtung [Die00] so ist seine Adjazenzmatrix symmetrisch. Man kann also zählen wie viele Einträge es innerhalb der Matrix auf der Hauptdiagonalen und einer der Hälften gibt. Das ergäbe immer $\frac{1}{2}n^2$ Schritte, wenn n die Ordnung des Graphen ist. Die Komplexität betrüge also $O(n^2)$. Bei der Darstellung durch eine Inzidenzliste wäre das anders. Hier könnte einfach die Größe der Liste gesucht werden und man wüsste die Größe des Graphen. Die Komplexität wäre hier, wie schon erwähnt, je nach Implementierung $O(n)$ oder $O(1)$.

Der Grad eines Knotens

Während die zwei ersten vorgestellten Metriken vor allem den Graphen als ganzes beschreiben, ist es auch noch wichtig zu wissen, was einen einzelnen Knoten auszeichnet, um einen Graphen besser zu beschreiben. Hierzu gibt es die grundlegende Metrik des **Grad** eines Knotens. Der Grad eines Knotens beschreibt die Anzahl der mit einem Knoten inzidenten Kanten [Die00]. D.h. es drückt aus, wie viele Kanten mit einem Knoten verbunden sind. Man kann dies z.B. durch eine Funktion ausdrücken, die einen Knoten v auf eine natürliche Zahl abbildet: $d(v)$.

Auf Basis dieser Metrik lässt sich auch andere verwandte Metriken ableiten. Hierzu gehört der „**Minimalgrad**“ und der „**Maximalgrad**“. Der Minimalgrad ist der kleinste Knoten-Grad eines Graphen G : $\delta(G) := \min\{d(v) \mid v \in V(G)\}$. Parallel dazu ist der Maximalgrad der größte Knoten-Grad eines Graphen G : $\Delta(G) := \max\{d(v) \mid v \in V(G)\}$. Darüber hinaus kann man noch den „**Durchschnittsgrad**“ eines Graphen bestimmen. Dieser bildet den Durchschnitt aller Knotengrade ab: $d(G) := \sum_{v \in V(G)} d(v) / |V|$. [Die00]

Des Weiteren gibt es bei der Betrachtung eines gerichteten Graphen zusätzliche Abwandlungen der Metrik. Da hier die Kanten immer zu einem Knoten gerichtet sind unterscheidet man speziell zwischen dem „**Eingangsgrad**“ und dem „**Ausgangsgrad**“. Der Eingangsgrad eines Knoten beschreibt dabei die Anzahl der Kanten, die auf einen Knoten „zeigen“. Der Ausgangsgrad zeigt wie viele Kanten von einem Knoten „weggehen“. [Bal97]

Auch der Grad eines Knotens und die meisten seiner verwandten Metriken sind weit verbreitet. So sind der allgemeine Grad, der Eingangsgrad, der Ausgangsgrad in allen drei betrachteten Mathematikbibliotheken vorhanden. SageMath unterstützt sogar nativ die Metrik „Durchschnittsgrad“. [Sag20; Mat20a; Wol20a]

Die Berechnung eines Grades über eine Adjazenzmatrix oder eine Adjazenzzliste ist in linearer Zeit lösbar ($O(n)$). Bei der Adjazenzmatrix muss einfach nur die jeweilige Reihe des zugehörigen Knotens durchlaufen werden und gezählt werden, wie häufig ein Eintrag für eine Kante enthalten. Mit Hilfe der Adjazenzzliste kann einfach die Größe der Liste als Grad genommen werden, die dem Knoten zugehörig ist.

Anzahl der Zusammenhangskomponenten

Eine weitere Variante einen Graphen grundlegend zu beschreiben, besteht darin seine Zusammenhangskomponenten zu zählen. Hierfür ist es zunächst wichtig zu verstehen, was Zusammenhang bei Graphen bedeutet.

Ein Graph gilt dann als zusammenhängend, wenn gilt: $\forall a, b (a \in V \wedge b \in V \wedge a \neq b \implies \text{Weg_existiert}(a, b))$. Anschaulich bedeutet das, dass es zwischen zwei beliebigen Knoten immer einen Weg geben muss, der die beiden Knoten miteinander verbindet. Graphisch erscheint ein zusammenhängender Graph so, dass sich keine verschiedenen, klar voneinander trennbaren Komponenten erkennen lassen. Diese einzelnen Komponenten oder Partitionen eines nicht-zusammenhängenden Graphen werden „Zusammenhangskomponenten“ genannt. Innerhalb der Zusammenhangskomponenten gilt natürlich wieder die Eigenschaft des Zusammenhangs. [Die00]

Die Anzahl der Zusammenhangskomponenten gibt nun an, wie viele Komponenten innerhalb eines Graphen vorhanden sind. Die Metrik ist sowohl in MatLab als auch in SageMath vertreten [Sag20; Mat20b]. Zur Ermittlung der Anzahl wird sich eines einfachen Algorithmus bedient, der die Tiefen- oder Breitensuche nutzt, die jeweils jeden Knoten als „besucht“ markiert, den sie traversiert. Folgender Pseudocode beschreibt diesen Algorithmus:

```
1 Integer zaehleKomponenten (graph)
2   int komponentenanzahl = 0
3   for (knoten in graph)
4     if (knoten ist nicht besucht)
5       tiefen_oder_breitensuche(graph, knoten)
6     komponentenanzahl++
```

Da der Algorithmus von einem der Suchalgorithmen abhängig ist, bestimmt dieser die Komplexität zur Ermittlung der Kennzahl. Jeder Knoten wird genau einmal in dieser Prozedur besucht. Zudem wird über alle Knoten in der äußeren Zählschleife iteriert. Es ergibt sich dadurch eine Komplexität von $O(|V| + |E|)$.

2.2 Distanzmetriken

Innerhalb der Graphentheorie gibt es den Begriff des Wegs. Ein Weg beschreibt dabei einen Graphen, der Knoten in einer Reihe hintereinander verbindet. Somit hat der Anfangs- und End-Knoten den Grad 1 und alle „mittleren“ Knoten den Grad 2. Meist sucht man aber einen bestimmten Weg innerhalb eines bestehenden Graphen. Der Weg ist hier dann ein Teilgraph des ursprünglichen Graphen. Anschaulicher lässt sich ein Weg also als eine Folge von Kanten beschreiben, in der kein Knoten zweimal besucht wird. Die Länge eines Weges ist dabei die Anzahl der Kanten, die in einem Weg vorhanden sind. [Die00] Auf Basis des Weges und seiner Längen-Eigenschaft lassen sich nun eine Reihe von Metriken definiert werden.

Abstand/Distanz

Der „Abstand“ ist eine Metrik, die auf Basis von zwei Knoten innerhalb eines Graphen definiert wird. Sie beschreibt die Länge des kürzesten Weges zwischen den zwei Knoten, von denen man den Abstand wissen will. Aufgeschrieben werden kann die Metrik mittels einer Funktion, die für den Graph G zwei Knoten x und y auf eine natürliche Zahl abbildet: $d_G(x, y)$ [Die00] Diese Metrik ist wichtig als Basis für andere Metriken. Wie die bisherigen Metriken ist auch diese in den jeweiligen Bibliotheken vertreten [Sag20; Mat20c; Wol20a]. Zur Berechnung der Metrik kann auf verschiedene bekannte, graphtraversierende Algorithmen zurückgegriffen werden. Dazu zählen die Breitensuche, der Dijkstra-Algorithmus oder der Bellman-Ford-Algorithmus [Sag20]. Somit ist auch die Komplexität zum Herausfinden der Metrik gleich mit der der Algorithmen. So wäre bei einer Breitensuche eine Komplexität von $O(|V| + |E|)$ zu erwarten, da jede Kante abgegangen wird. Der Dijkstra-Algorithmus hingegen hat eine Komplexität von $O(|V|^2)$ [Jun13].

Extrenzität eines Knotens

Mit Hilfe des Abstandes lässt sich nun u.a. die „**Extrenzität**“ eines Knotens bestimmen. Die „Extrenzität“ ist dabei der maximale Abstand den ein Knoten von einem anderen Knoten in einem Graphen G haben kann. Eine einfache formale Notierung für den Knoten x wäre: $ecc(x, G) = \max_{x,y} \{d_G(x, y)\}$, wobei x der gegebene Knoten ist. [Har01]

Herauszufinden ist diese Kennzahl beispielsweise über den Dijkstra-Algorithmus, der den kürzesten Abstand zu jedem anderen Knoten sucht und dann einfach der größte zu wählen ist. D.h. das die Metrik auch die Komplexität hat wie bei der Metrik „Abstand“ erläutert wurde. Die Extrenzität eines Knotens ist anschließend noch für andere Metriken eine wichtige Basis und allgemein weit verbreitet in den genannten Bibliotheken [Sag20; Mat20c; Wol20a].

Durchmesser

Wie schon erwähnt, ist es nun möglich auf Basis des Abstands weitere Metriken zu definieren. Hierzu zählt unter anderem der „**Durchmesser**“ eines Graphen. Der Durchmesser beschreibt dabei den größten Abstand zweier Knoten im Graphen G . [Die00] D.h. es ist der Abstand von allen Knoten zu allen Knoten zu berechnen und davon die größte Zahl auszuwählen. Formal notiert lässt sich die Metrik folgendermaßen beschreiben: $\text{Durchmesser}(G) = \max_{x,y} \{d_G(x,y)\}$. Nimmt man zur Ermittlung der Abstände dabei den Dijkstra-Algorithmus und wendet diesen dann jeweils auf jeden einzelnen Knoten an, kann eine Komplexität von $O(|V|^3)$ angenommen werden, um die Metrik zu extrahieren. Auch diese Metrik lässt sich z.B. in SageMath oder Wolfram finden. In Matlab kann der Durchmesser über die Distanzmatrix ermittelt werden, die Matlab erstellen kann. [Sag20; Mat20c; Wol20a] Die Anwendung für diese Metrik kann z.B. sein, rein topologische Eigenschaften des Graphen herausfinden zu wollen. Allerdings kann auch in realen Problemen der Durchmesser als Metrik auftauchen. So ist z.B. der Abstand und damit der Durchmesser auch mit gewichteten Kanten berechenbar. [Sag20; GIT14] In einem Navigationssystem wäre der Durchmesser dann die längste fahrbare Strecke.

Radius

Parallel zum Durchmesser eines Graphen kann man auch dessen „**Radius**“ bestimmen. Hierfür wird die Metrik der Extrenzität wichtig und der Begriff der Zentralität. Eine Kante ist dann *zentral* bzw. im Zentrum eines Graphen, wenn dessen Extrenzität minimal ist. Dies kann nur einen Knoten, aber auch mehrere Knoten betreffen. Die minimale Extrenzität in einem Graphen, die die Knoten des Zentrums haben, nennt man dann auch den „**Radius**“ des Graphen. Geschrieben wird $\text{rad } G = \min_{x \in V(G)} \max_{y \in V(G)} d_G(x,y)$. [Die00] Der Radius lässt sich grundsätzlich auf die gleiche Weise herausfinden, wie der Durchmesser und hat dementsprechend die gleiche Komplexität. Des Weiteren ist diese Metrik auch weit verbreitet und lässt sich in allen untersuchten Bibliotheken finden [Sag20; Wol20a; Mat20c]

2.3 Kreis-basierte Metriken

Ausgehend von einem Weg innerhalb eines Graphen können wir auch den Begriff des Kreises definieren. Ein Kreis ist dabei einfach ein Weg, der eine zyklische Eckenfolge hat. Sei x_i ein Knoten, so hat ein Kreis folgende typische Eckenfolge: $x_0 \dots x_{k-1}, x_0$. [Die00] Auch auf Basis solcher Kreise lassen sich verschiedene Graph-Metriken definieren.

Länge eines Kreises

Ist ein solcher Kreis, wie vorhin beschrieben, gegeben, kann unter anderen an diesem seine Länge abgelesen werden. Die Länge beschreibt dabei die Anzahl der Kanten, die ein Kreis enthält. Ist erst einmal ein Kreis gegeben lässt sich die Länge leicht berechnen, denn die Länge eines Kreises ist gleich mit der Anzahl der Knoten innerhalb eines Kreises. Somit lässt sich die Länge in konstanter Zeit berechnen. [Die00] Diese Metrik ist allerdings nicht direkt in referenzierten Mathematikbibliotheken vertreten, da grundsätzlich nur Graphen allgemein verarbeitet werden. Allerdings wäre es möglich, auf Basis des Wissens einen zyklischen Graphen zu erstellen und beispielsweise sich die Kantenanzahl zurückgeben zu lassen.

Tailenweite und Umfang

Nimmt man die Länge eines Kreises als Basis, können weitere Kennzahlen für den gesamten Graphen ermittelt werden. Eine davon ist die „Tailenweite“ des Graphen. Die „Tailenweite“ ist so definiert, dass sie den Wert der Länge des kürzesten Kreises innerhalb des Graphen annimmt. Umgekehrt lässt sich auch der „Umfang“ des Graphen bestimmen. Der Umfang ist dabei so groß wie die Länge des größtmöglichen Kreises innerhalb eines Graphen. Hat ein Graph keinen Kreis, so ist es nicht möglich für beide Kennzahlen einen Wert zu ermitteln. Allerdings haben diese dann einen festen Wert. So beträgt die „Tailenweite“ in diesem Fall ∞ und der Umfang null. [Die00]

Um die Tailenweite und den Umfang eines Graphen herauszufinden, ist es grundsätzlich notwendig die jeweiligen Zyklen innerhalb des Graphen herauszufinden. Hierfür lässt sich unter anderem eine modifizierte Tiefensuche nutzen, die mittels Markierung der Knoten erkennt, ob sie bereits schon einmal bei einem Knoten war und infolgedessen einen Zyklus erkennt. Hierbei wird unterschieden, ob ein Knoten noch nicht bearbeitet wurde, in Bearbeitung ist oder bereits der Algorithmus auf ihm vollständig abgeschlossen wurde. Wird der Algorithmus auf einem Knoten aufgerufen, der sich noch in Bearbeitung befindet, ist ein Zyklus gefunden. Mit einer richtigen Ausgabe kann dieser dann auch benannt werden. Daraus folgt auch, dass die Ermittlung der Tailenweite und des Umfangs entspricht, denn nach der Tiefensuche, muss einfach der größte bzw. der kleinste Zyklus gewählt werden, um die Metriken zu ermitteln. Die Komplexität beträgt damit $O(|V| + |E|)$. [Kne19; Vöc+08]

Zur Metrik „Umfang“ lassen sich in den genannten Bibliotheken, außer bei Wolfram, keine direkten Implementierungen finden. Allerdings ist es möglich in SageMath und in Wolfram die „Tailenweite“ direkt zu evaluieren. [Sag20; Wol20b]

2.4 Zusammenhangsmetriken (Connectivity)

Neben der Definition von Metriken auf Basis von Distanzen ist es auch möglich, Kennzahlen zu ermitteln, die den Zusammenhang eines Graphen betrachten. Hierfür ist es wichtig zu verstehen, wann ein Graph als zusammenhängend gilt und was eine Zusammenhangskomponente bzw. Partition eines Graphen ist. Dies wurde bei der Metrik „Anzahl der Zusammenhangskomponenten“ erläutert.

Dichte

Bei der Vorstellung grundlegender Graphmetriken wurden u.a. die Größe und die Ordnung eines Graphen erklärt. Darauf aufbauend kann eine Kennzahl ermittelt werden, die aussagt, wie stark vernetzt ein Graph ist. Die „Dichte“ eines Graphen gibt an, inwiefern der Graph so viele Kanten hat wie es ihm theoretisch möglich ist. Die „Dichte“ setzt also tatsächliche Kantenanzahl (Größe) und die mögliche Kantenanzahl in ein Verhältnis. Die Metrik kann darauffolgend einen Wert zwischen 0 und 1 annehmen. Ist der Wert 0 hat der Graph keine Kanten. Ist der Wert hingegen 1 so hat man einen vollständigen Graphen vor sich liegen. Möchte man die Dichte eines Graphen ermitteln ist es nötig die Größe des Graphen durch die potenzielle Größe zu teilen. Dies ist mit folgender Gleichung möglich: $\frac{|E|}{\binom{|V|}{2}}$. Hat man statt einem ungerichteten Graphen einen gerichteten muss die Formel leicht abgewandelt werden, da für einen vollständigen Graphen nun doppelt so viele Kanten nötig sind: $\frac{|E|}{2 \cdot \binom{|V|}{2}}$. [Sag20]

Zur Implementierung der Metrik ist es infolgedessen nur nötig die Größe und die Ordnung des Graphen herauszufinden. Die Komplexität zur Berechnung der „Dichte“ ist deshalb gleich der Komplexität zur Berechnung von Größe und Ordnung addiert. Die Berechnung der „Dichte“ selbst erfolgt in konstanter Zeit. Durch die Ableitung der Metrik aus zwei grundlegenden Kennzahlen ist die Berechnung auch problemlos möglich, ohne dass es eine explizite Implementierung in einer Bibliothek gibt. Allerdings bieten Wolfram und SageMath spezielle Funktionen für die „Dichte“ eines Graphen. [Sag20; Wol20b; Mat20a]

Stärke

Oft repräsentieren Graphen ein Netzwerk. Im Rahmen von Netzwerken wird unter anderem von deren „Stärke“ gesprochen. Die „Stärke“ gibt dabei das minimale Verhältnis zwischen entfernten Kanten und dadurch erstellter Zusammenhangskomponenten an.

Es muss dabei allerdings die Anzahl der Zusammenhangskomponenten insgesamt erhöht werden. Ist die „Stärke“ eines Netzwerks, bzw. eines Graphen, hoch, ist es u.a. schwieriger für einen Angreifer das Netzwerk stark zu beschädigen, greift dieser die Verbindungen, bzw. Kanten, des Netzwerks an. Zur Berechnung der Stärke $\sigma(G)$ seien folgende Annahmen gegeben: Sei Π die Menge aller möglichen Partitionierungen der Knoten V und $\partial\pi$ die Menge an Kanten, die entfernt werden müssten, um die Partitionierung π zu erreichen, gilt folgende Formel zur Errechnung der Stärke:

$$\sigma(G) = \min_{\pi \in \Pi} \frac{|\partial\pi|}{|\pi| - 1}$$

Es wird also jede mögliche Partitionierung der Knoten V betrachtet und ermittelt welche Kanten man entfernen müsste, um diese Partitionierung der Knoten zu erhalten. Die Anzahl der Elemente in der jeweiligen Menge werden dann in ein Verhältnis gesetzt. Dabei wird eine Zusammenhangskomponenten aus π subtrahiert, da ein Graph immer zumindest aus einer Komponente besteht. Aus all diesen erstellten Verhältnissen ist nun das Minimum das Ergebnis. [Tru93; Cun85] Es kann auch anders gesagt werden, dass ein Graph bei dem die Entfernung weniger Kanten zu vergleichsweise vielen Zusammenhangskomponenten führt, ein sehr „schwacher“ Graph ist. Umgekehrt ist es bei einem „starken“ Graphen nicht möglich, selbst durch die Entfernung vieler Kanten (Zähler), eine vergleichsweise hohe Anzahl an Zusammenhangskomponenten (Nenner) zu erreichen.

Die Berechnung der „Stärke“ und die Verbesserung der Komplexität des Algorithmus war Thema mehrerer wissenschaftlicher Arbeiten. Die beste erreichte Komplexität erzielte dabei V. A. Trubin mit einer Komplexität von $O(\min(\sqrt{m}, n^{2/3})mn \log(n^2/m + 2))$. m ist hierbei die Anzahl an Kanten im Graphen, n die Anzahl an Knoten. [Tru93]

In SageMath, Wolfram oder MatLab ist die „Stärke“ von Graphen nicht implementiert. Darüber hinaus ist es auch möglich statt über die Entfernung von Kanten die Metrik über die Entfernung von Knoten definieren. In diesem Fall spricht man über die „Härte“ oder „Zähigkeit“ (engl. „Toughness“) des Graphen. [Chv06]

„Vertex Connectivity“/Zusammenhang

Die Stärke eines Graphen ist eine nicht ganzzahlige Metrik zur Beschreibung des allgemeinen Zusammenhangs innerhalb eines Graphen. Wie bei der Stärke schon erwähnt, ist diese Metrik in den einschlägigen Bibliotheken nicht zu finden. Die nächsten zwei Metriken sind sowohl in SageMath als auch in Wolfram zu finden und beschreiben die Stärke des Zusammenhangs des Graphen mittels einer ganzen Zahl [Sag20; Wol20a]

Die erste dieser Metriken ist die „Vertex Connectivity“ oder „Zusammenhang“. Hierbei wird ein zusammenhängender oder auch nicht zusammenhängender Graph betrachtet und ermittelt wie viele Knoten aus dem Graphen mindestens entfernt werden müssen,

sodass der Graph nicht mehr zusammenhängend ist. k entspricht dieser minimalen Anzahl an Knoten. Formal lässt sich sagen, dass ein Graph k -zusammenhängend ist, wenn $|G| > k$ und der Graph für jede mögliche Knotenmenge X mit der Mächtigkeit $< k$ zusammenhängend bleibt, sobald man alle Knoten $X \subseteq V$ aus V entfernt ($G - X$). Da ein Graph der 3-zusammenhängend ist auch 4/5/...-zusammenhängend ist, ist die absolute „Vertex Connectivity“ bzw. der „Zusammenhang“ das k , das für den Graph minimal ist. Der Zusammenhang ist dann 0, wenn der Graph von Anfang an nicht zusammenhängend ist oder der Graph nur aus einem Knoten besteht. [Die00] Auch bei dieser Metrik gilt wie bei der Stärke, dass der Graph schwerer zu „trennen“ ist, je höher die jeweilige Kennzahl ist. Daraus ist auch zu folgen, dass bei hohem Zusammenhang beispielsweise ein Netzwerk weniger Anfällig für Angriffe ist.

Um den Zusammenhang eines Graphen algorithmisch herauszufinden, können zunächst zwei triviale Fälle abgedeckt werden. Ist ein Graph leer oder trivial (nur ein Knoten) ist der Zusammenhang, wie bereits erwähnt, 0. Ist hingegen der Graph vollständig, beträgt der Zusammenhang $|V|$. Allerdings lässt sich auch ein allgemeiner Algorithmus definieren, der die „Vertex Connectivity“ berechnet. Für den Algorithmus wird eine zusätzliche Funktion benötigt. $N(a, b)$ nimmt zwei Knoten entgegen. Eine Menge an Knoten, die bei Entfernung dafür sorgt, dass zwischen a und b kein Weg mehr existiert, wird „Knoten-Separator“ genannt. N gibt nun die Kardinalität des minimalen „Knoten-Separators“ von a und b zurück. Sind a und b direkt mit einer Kante verbunden, gibt es keinen „Knoten-Separator“. Sind diese Knotenpaare bei der Berechnung der „Vertex Connectivity“ anzusehen, ist unnötig und muss nicht betrachtet werden. Schlussendlich ist nämlich die Kardinalität des kleinsten minimalen „Knoten-Separators“ die gesuchte Kennzahl. Zur Berechnung gibt Shimon Even in seinem Buch „Graph Algorithms“ folgenden Algorithmus an [Eve12]:

```

1  Vertex-Connectivity(V, E)
2      Sortiere Knoten  $v_1, v_2, \dots, v_{|V|}$  so, dass es von  $v_1$  keine Kante
   zu irgendeinem  $v$  gibt
3       $\gamma = \text{unendlich}$ 
4       $i = 1$ 
5      while  $i \leq \gamma$ 
6          for each  $v$ , sodass  $v_i$  keine Kante zu irgendeinem  $v$  gibt
7               $\gamma = \min\{\gamma, N(v_i, v)\}$ 
8      return  $\gamma$ 

```

Der gezeigte Algorithmus terminiert mit γ gleich dem Zusammenhang. In seinen Ausführungen erläutert Even zudem, dass der Algorithmus eine Zeitkomplexität von $O(|V|^{1/2} \cdot |E|^2)$ aufweist.

„Edge Connectivity“

2.5 Zentralitätsmetriken

2.6 Arborizität

2.7 Chromatische Zahl und chromatischer Index

2.8 Weitere Metriken

2.9 Übersicht der vorgestellten Graphmetriken

Nachdem nun eine weite Reihe an Graph-Metriken vorgestellt wurden, sollen diese mit Hilfe einer Aufzählung zusammengefasst werden. Aufgelistet werden jeweils der Name, bzw die Bezeichnung, der Metrik, ihre Definition und die Komplexität zur Berechnung der Kennzahl. Bei der Definition wird dabei eine Mathematische oder eine wörtliche Beschreibung angegeben. Gegebenenfalls auch beides, falls es für das Verständnis förderlich ist.

Ordnung

Definition: Anzahl der Knoten eines Graphen

Komplexität: $O(1)$ oder $O(n)$

Größe

Definition: Anzahl der Kanten eines Graphen

Komplexität: Adjazenzmatrix/liste: $O(n^2)$; Inzidenzmatrix/liste: $O(1)$ oder $O(n)$

3 Ähnlichkeit von Graphen

4 Implementierung und Umsetzung der Metriken

4.1 Implementierung in verschiedenen Graphdatenbanken

4.2 Vergleich der Implementierungen

5 Graphmetriken und Ähnlichkeit in Anwendung

6 Fazit und Zusammenfassung

6.1 Zusammenfassung der Ergebnisse

6.2 Fazit

Glossar

Literatur

- [Bal97] V. K. Balakrishnan. *Schaum's outline of theory and problems of graph theory*. Schaum's outline series. New York: McGraw-Hill, 1997. 293 S. ISBN: 978-0-07-005489-9.
- [Chv06] V. Chvátal. "Tough graphs and hamiltonian circuits". In: *Discrete Mathematics* 306.10 (Mai 2006), S. 910–917. ISSN: 0012365X. DOI: 10.1016/j.disc.2006.03.011. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0012365X06001397> (besucht am 09.12.2020).
- [Cun85] William H. Cunningham. "Optimal Attack and Reinforcement of a Network". In: *J. ACM* 32.3 (Juli 1985). Place: New York, NY, USA Publisher: Association for Computing Machinery, S. 549–561. ISSN: 0004-5411. DOI: 10.1145/3828.3829. URL: <https://doi.org/10.1145/3828.3829>.
- [Die00] Reinhard Diestel. *Graphentheorie*. 2., neu bearb. und erw. Aufl. Springer-Lehrbuch. OCLC: 247312585. Berlin: Springer, 2000. 314 S. ISBN: 978-3-540-67656-0.
- [EK13] W. Ellens und R. E. Kooij. *Graph measures and network robustness*. eprint: 1311.5064. 2013.
- [Eve12] Shimon Even. *Graph algorithms*. Unter Mitarb. von Guy Even. 2nd ed. Cambridge, NY: Cambridge University Press, 2012. 189 S. ISBN: 978-0-521-51718-8 978-0-521-73653-4.
- [GIT14] GITTA. *Durchmesser eines Graphen*. Durchmesser eines Graphen. 2014. URL: https://www.gitta.info/Accessibiliti/de/html/StructPropNetw_learningObject2.html (besucht am 18.11.2020).
- [Har01] Frank Harary. *Graph theory*. 15. print. OCLC: 248770458. Cambridge, Mass: Perseus Books, 2001. 274 S. ISBN: 978-0-201-41033-4.
- [Jun13] Dieter Jungnickel. *Graphs, networks and algorithms*. 4th ed. Algorithms and computation in mathematics 5. OCLC: 821566132. Berlin: Springer, 2013. 675 S. ISBN: 978-3-642-32278-5 978-3-642-32277-8.
- [Kne19] Helmut Knebl. *Algorithmen und Datenstrukturen: Grundlagen und probabilistische Methoden für den Entwurf und die Analyse*. OCLC: 1123167896. 2019. ISBN: 978-3-658-26511-3.

- [Lov12] László Lovász. *Large networks and graph limits*. American Mathematical Society colloquium publications volume 60. Providence, Rhode Island: American Mathematical Society, 2012. 475 S. ISBN: 978-0-8218-9085-1.
- [Mat20a] Matlab. *Directed and Undirected Graphs - MATLAB & Simulink - MathWorks Deutschland*. Directed and Undirected Graphs. 2020. URL: <https://de.mathworks.com/help/matlab/math/directed-and-undirected-graphs.html> (besucht am 10.11.2020).
- [Mat20b] Matlab. *Graph and Network Algorithms*. Graph and Network Algorithms. 2020. URL: https://de.mathworks.com/help/matlab/graph-and-network-algorithms.html?searchHighlight=graph&s_tid=srchtitle (besucht am 12.01.2021).
- [Mat20c] Matlab. *Shortest path distances of all node pairs*. Distances. 2020. URL: <https://de.mathworks.com/help/matlab/ref/graph.distances.html> (besucht am 18.11.2020).
- [Sag20] SageMath. *Graph Theory*. Sage Math Reference Manual. 2020. URL: <https://doc.sagemath.org/html/en/reference/graphs/index.html> (besucht am 25.10.2020).
- [Tit19] Peter Tittmann. *Graphentheorie: Eine anwendungsorientierte Einführung*. 3., aktualisierte Auflage. München: Hanser, Carl, 2019. 168 S. ISBN: 978-3-446-46052-2 978-3-446-46503-9.
- [Tru93] V. A. Trubin. "Strenght of a graph and packing of trees and branchings". In: *Cybernetics and Systems Analysis* (1993).
- [Tur04] Volker Turau. *Algorithmische Graphentheorie*. Oldenbourg Wissenschaftsverlag, 1. Jan. 2004. ISBN: 978-3-486-59377-8. DOI: 10.1524/9783486593778. URL: <https://www.degruyter.com/view/title/310250> (besucht am 24.10.2020).
- [Vöc+08] Berthold Vöcking u. a., Hrsg. *Taschenbuch der Algorithmen*. eXamen.press. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. ISBN: 978-3-540-76393-2 978-3-540-76394-9. DOI: 10.1007/978-3-540-76394-9. URL: <http://link.springer.com/10.1007/978-3-540-76394-9> (besucht am 06.12.2020).
- [WM19] Peter Wills und Francois G. Meyer. *Metrics for Graph Comparison: A Practitioner's Guide*. eprint: 1904.07414. 2019.
- [Wol20a] Wolfram. *Graph Measures & Metrics*. Wolfram Language Documentation. 2020. URL: <https://reference.wolfram.com/language/guide/GraphMeasures.html> (besucht am 25.10.2020).

[Wol20b] Wolfram. *Wolfram Language & System Documentation Center*. Documentation Center. 2020. URL: <https://reference.wolfram.com/language/> (besucht am 06.12.2020).