

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

8. November 2020

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, 8. November 2020

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 Distanz-Metriken	4
2.3 Zusammenhangsmetriken (Connectivity)	4
2.4 Zentralitätsmetriken	4
2.5 Chromatische Zahl und chromatischer Index	4
2.6 Weitere Metriken	4
2.7 Übersicht der vorgestellten Graphmetriken	4
3 Ähnlichkeit von Graphen	5
4 Implementierung und Umsetzung der Metriken	6
4.1 Implementierung in verschiedenen Graphdatenbanken	6
4.2 Vergleich der Implementierungen	6
5 Graphmetriken und Ähnlichkeit in Anwendung	7
6 Fazit und Zusammenfassung	8
6.1 Zusammenfassung der Ergebnisse	8
6.2 Fazit	8
Glossar	9
Literatur	10

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)?
- Welche Motivation hat die Metrik?
- In welchen bekannten Mathematikbibliotheken oder Graph-Datenbanken lässt sich die Metrik finden?
- Wie ist die Metrik im Bezug auf den Rechenaufwand zu bewerten?
- Welche Verbreitung hat die Metrik?

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

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

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

[Die00]

2.1 Grundlegende Metriken

Test

Ordnung eines Graphen

2.2 Distanz-Metriken

2.3 Zusammenhangsmetriken (Connectivity)

2.4 Zentralitätsmetriken

2.5 Chromatische Zahl und chromatischer Index

2.6 Weitere Metriken

2.7 Übersicht der vorgestellten Graphmetriken

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

- [Aig15] Martin Aigner. *Graphentheorie: eine Einführung aus dem 4-Farben Problem*. 2., überarbeitete Auflage. Springer Studium Mathematik Bachelor. OCLC: 927721160. Wiesbaden: Springer Spektrum, 2015. 196 S. ISBN: 978-3-658-10322-4 978-3-658-10323-1.
- [Alo88] N. Alon. "The linear arboricity of graphs". In: *Israel Journal of Mathematics* 62.3 (Okt. 1988), S. 311–325. ISSN: 0021-2172, 1565-8511. DOI: 10.1007/BF02783300. URL: <http://link.springer.com/10.1007/BF02783300> (besucht am 24. 10. 2020).
- [And77] Lars Dovling Andersen. "On edge-colorings of graphs." In: *MATHEMATICA SCANDINAVICA* 40 (1. Dez. 1977), S. 161. ISSN: 1903-1807, 0025-5521. DOI: 10.7146/math.scand.a-11685. URL: <http://www.mscaand.dk/article/view/11685> (besucht am 24. 10. 2020).
- [Ber01] Claude Berge. *The theory of graphs*. Dover books on mathematics. Mineola, N.Y: Dover, 2001. 247 S. ISBN: 978-0-486-41975-6.
- [BI19] José Bento und Stratis Ioannidis. "A family of tractable graph metrics". In: *Applied Network Science* 4.1 (Dez. 2019), S. 107. ISSN: 2364-8228. DOI: 10.1007/s41109-019-0219-z. URL: <https://appliednetsci.springeropen.com/articles/10.1007/s41109-019-0219-z> (besucht am 22. 10. 2020).
- [BK79] Frank Bernhart und Paul C Kainen. "The book thickness of a graph". In: *Journal of Combinatorial Theory, Series B* 27.3 (Dez. 1979), S. 320–331. ISSN: 00958956. DOI: 10.1016/0095-8956(79)90021-2. URL: <https://linkinghub.elsevier.com/retrieve/pii/0095895679900212> (besucht am 24. 10. 2020).
- [BM08] Raymond Bisdorf und Jean-Luc Marichal. *Counting non-isomorphic maximal independent setsof then-cycle graph*. Nov. 2008. URL: <https://arxiv.org/abs/math/0701647v2> (besucht am 24. 10. 2020).
- [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.

- [Gus83] Dan Gusfield. "Connectivity and edge-disjoint spanning trees". In: *Information Processing Letters* 16.2 (Feb. 1983), S. 87–89. ISSN: 00200190. DOI: 10.1016/0020-0190(83)90031-5. URL: <https://linkinghub.elsevier.com/retrieve/pii/0020019083900315> (besucht am 24. 10. 2020).
- [HM11] Javier Martín Hernandez und Piet Van Mieghem. *Classification of graph metrics*. 2011. URL: https://www.nas.ewi.tudelft.nl/people/Piet/papers/TUDreport20111111_MetricList.pdf (besucht am 22. 10. 2020).
- [Hun14] Michael Hunger. *Neo4j 2.0 Eine Graphdatenbank für alle*. OCLC: 875609599. 2014. ISBN: 978-3-86802-315-2 978-3-86802-654-2. URL: <https://nbn-resolving.org/urn:nbn:de:101:1-2014032620976> (besucht am 24. 10. 2020).
- [Jan20] JanusGraph. *JanusGraph Dokumentation*. 2020. URL: <https://docs.janusgraph.org/> (besucht am 25. 10. 2020).
- [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.
- [Moh89] Bojan Mohar. "Isoperimetric numbers of graphs". In: *Journal of Combinatorial Theory, Series B* 47.3 (Dez. 1989), S. 274–291. ISSN: 00958956. DOI: 10.1016/0095-8956(89)90029-4. URL: <https://linkinghub.elsevier.com/retrieve/pii/0095895689900294> (besucht am 24. 10. 2020).
- [Öst02] Patric R.J. Östergård. "A fast algorithm for the maximum clique problem". In: *Discrete Applied Mathematics* 120.1 (Aug. 2002), S. 197–207. ISSN: 0166218X. DOI: 10.1016/S0166-218X(01)00290-6. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0166218X01002906> (besucht am 24. 10. 2020).
- [RWE15] Ian Robinson, Jim Webber und Emil Eifrem. *Graph databases [new opportunities for connected data]*. OCLC: 1028626678. Sebastopol, CA: O'Reilly, 2015. ISBN: 978-1-4919-3200-1. URL: <https://neo4j.com/graph-databases-book/?ref=home> (besucht am 24. 10. 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).
- [SU11] Edward R. Scheinerman und Daniel H. Ullman. *Fractional graph theory: a rational approach to the theory of graphs*. Dover books on mathematics. OCLC: ocn721885660. Minola, N.Y: Dover Publications, 2011. 211 S. ISBN: 978-0-486-48593-5.
- [Tig20] TigerGraph. *TigerGraph Dokumentation*. 2020. URL: <https://docs.tigergraph.com/> (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.
- [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).
- [WM19] Peter Wills und Francois G. Meyer. *Metrics for Graph Comparison: A Practitioner's Guide*. eprint: 1904.07414. 2019.
- [Wol20] Wolfram. *Graph Measures & Metrics*. Wolfram Language Documentation. 2020. URL: <https://reference.wolfram.com/language/guide/GraphMeasures.html> (besucht am 25.10.2020).