

Volker Ahlers

Visualisierung komplexer Netzwerke

Abstract

In vielen Anwendungskontexten spielen Netzwerke aus Entitäten (Personen, Orte, abstrakte Objekte) und Beziehungen zwischen Entitäten eine wichtige Rolle, die mathematisch als Graphen betrachtet werden können. Die Visualisierung solcher Netzwerke – engl. *Graph Drawing* – erleichtert es dem Menschen, diese zu verstehen und beispielsweise Muster zu erkennen. Im folgenden Beitrag werden Anwendungen aus der Bioinformatik und der IT-Sicherheit mit ihren jeweiligen Herausforderungen vorgestellt. Insbesondere wird darauf eingegangen, wie diese Herausforderungen, etwa die Dynamik eines Netzwerkes, die Wahl der visuellen Mittel bestimmen.

Networks of entities (persons, locations, abstract objects) as well as relations between these entities play a crucial role in many application contexts. Mathematically speaking networks can be regarded as graphs. Network visualization – or *graph drawing* – helps human beings to understand their information content and detect, for example, inherent patterns. In this contribution applications from bioinformatics and IT security are presented together with their respective challenges. A special focus will be laid on the question how these challenges, such as the network dynamics, influence the choice of visual means.

Warum Visualisierung?

Der Mathematiker Richard Hamming stellt sein erstmals 1962 erschienenes Lehrbuch *Numerical Methods for Scientists and Engineers* unter das Motto:

„The Purpose of Computing Is Insight, Not Numbers.”

(HAMMING 1986: 3)

Das primäre Ziel von Computersimulationen und – im übertragenen Sinne – auch Datenanalysen ist es nicht, Zahlen zu produzieren, sondern Erkenntnisse zu gewinnen. Ein wichtiges Hilfsmittel für die Interpretation von Zahlen und Informationen durch den Menschen stellt die Visualisierung dar. Wir Menschen können große Mengen visueller Informationen schnell erfassen und verarbeiten, lange Zahlenkolonnen dagegen nur langsam und mit Mühe. Der Begriff *Scientific Visualization* (Wissenschaftliche Visualisierung) wurde 1987 durch eine von Bruce McCormick, Thomas DeFanti und Maxine Brown geleitete Arbeitsgruppe eingeführt und einleitend wie folgt beschrieben bzw. beworben:

„Visualization offers a method for seeing the unseen. It enriches the process of scientific discovery and fosters profound and unexpected insights. In many fields it is already revolutionizing the way scientists do science.” (MCCORMICK/DEFANTI/BROWN 1987: 3)

Netzwerke stellen spezielle Daten dar, die aus Entitäten und Verbindungen zwischen diesen Entitäten bestehen. Als Beispiele seien Beziehungen zwischen Personen in sozialen Netzen oder Verkehrsverbindungen zwischen Orten genannt; Netzwerke können aber selbstverständlich auch abstrakterer Natur sein. Mathematisch werden derartige Zusammenhänge durch Graphen beschrieben, die aus Knoten (Entitäten) und Kanten (Verbindungen) bestehen. Sowohl Knoten als auch Kanten können mit weiteren Metadaten angereichert sein. Insbesondere können Kanten numerische Werte (Gewichte/Bewertungen) zugewiesen werden, die etwa die Länge oder die Kosten eines Weges darstellen. Es sei betont, dass ein mathematischer Graph nur die abstrakten Verbindungen repräsentiert und keine geometrischen Informationen enthält, also keine Knotenpositionen, Kantenverläufe und ähnliches.

Das Beispiel in Abbildung 1 (angelehnt an SPENCE 2007) zeigt, wie Visualisierung helfen kann, die Struktur von Netzwerken zu verstehen. Links (a) ist eine Liste von Telefonaten zwischen Angestellten einer Firma zu sehen. Die erste Visualisierung (b) lässt bereits erkennen, dass Bob viel telefoniert und dass es Dreierbeziehungen gibt, etwa aus Alice, Bob und Grace. Die zweite Visualisierung (c) schließlich zeigt, dass es drei vollständig getrennte Teilnetze gibt, die untereinander keinen Telefonkontakt haben. Auf Algorithmen zum Erstellen von Netzwerkvisualisierungen wird im nächsten Abschnitt eingegangen.

Ein historisches Beispiel für die Visualisierung eines Netzwerkes ist die 1864 von Charles Minard erstellte Darstellung des französischen Weinexports, siehe Abbildung 2 (vgl. TUFTE 2001). Die Menge des in ein bestimmtes Land exportierten Weins ist über die Breite der Kanten codiert, welche sich nach und nach aufspalten – eine Technik, die als Flowmap auch heute noch gebräuchlich ist, beispielsweise in Sankey-Diagrammen für Energieflüsse. Interessant sind grafische Details der historischen Karte: So wurden beispielsweise der

Ärmelkanal und die Straße von Gibraltar verbreitert, um ausreichend Platz für die Kanten zu haben.

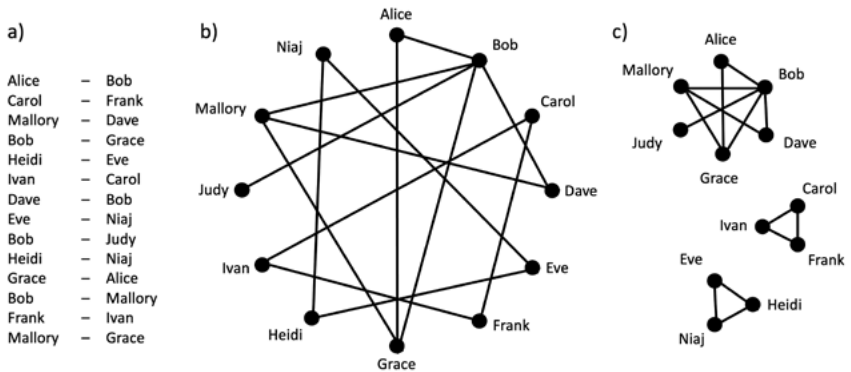


Abbildung 1: Zwei verschiedene Visualisierungen (b, c) desselben Netzwerkes (a) von Telefonkontakten (angelehnt an Spence 2007: 73), Volker Ahlers

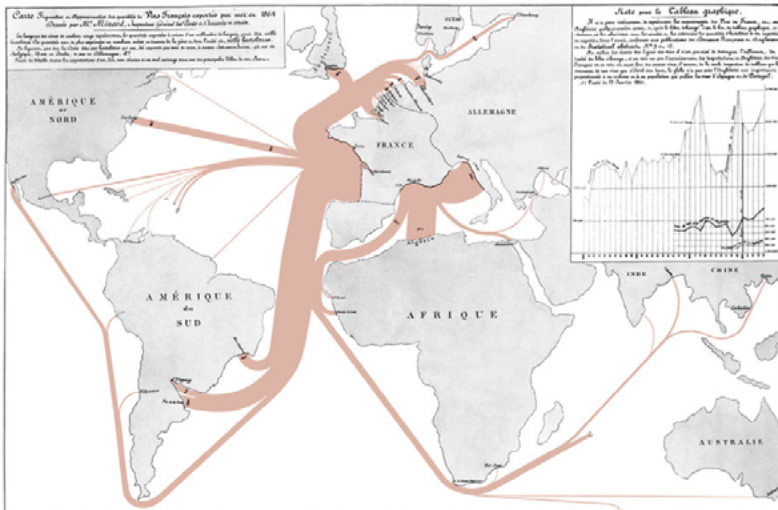


Abbildung 2: Visualisierung des französischen Weinexports im Jahr 1864 durch Charles Joseph Minard, Public Domain, https://commons.wikimedia.org/wiki/File:Minard%20%80%99s_map_of_French_wine_exports_for_1864.jpg

Ein weiteres bekanntes Beispiel ist der 1931 von Harry Beck modernisierte Londoner U-Bahn-Plan (vgl. SPENCE 2007).¹⁾ Die Erkenntnis, dass die genaue geografische Linienführung für Nutzerinnen und Nutzer einer U-Bahn irrelevant ist, führte zu der heute weltweit gebräuchlichen diagrammatischen Darstellung mittels gerader Kanten, die horizontal, vertikal oder diagonal verlaufen.

Graph Drawing

Wie lassen sich aus mathematischen Graphen, die nur Verbindungsinformationen enthalten, grafische Darstellungen erstellen? Wie werden insbesondere die Knotenpositionen und Kantenverläufe, sogenannte Graph Layouts, bestimmt? Die Fachdisziplin, die sich der Entwicklung entsprechender Algorithmen widmet, heißt Graph Drawing.

Typische Anforderungen an ein Graph Layout lauten wie folgt (vgl. BATTISTA et al. 1999):

- Kompakte Fläche des gesamten Graphen
- Vermeidung von Kantenkreuzungen
- Vermeidung von Kantenbiegungen und -knicken
- Minimierung der Kantenlängen
- Maximierung des Winkels zwischen benachbarten Kanten
- Darstellung von im Graphen enthaltenen Symmetrien
- Clustern eng zusammenhängender Knoten

Die Anforderungen widersprechen sich teilweise und lassen sich in der Regel nicht gleichzeitig optimal erfüllen. So werden sich zum Beispiel Kantenkreuzungen oftmals nur mit Hilfe längerer, gebogener oder geknickter Kanten vermeiden lassen, was zudem die benötigte Zeichenfläche vergrößern kann.

Zur grafischen Gestaltung der Knoten und Kanten stehen im nächsten Schritt verschiedene Mittel zur Verfügung, die auch als Kanäle bezeichnet werden (vgl. MUNZNER 2015):

- Knoten: Form, Größe, Farbe, Textlabel, ...
- Kanten: Linienart, Linienbreite, Farbe, Pfeilspitzen (bei gerichteten Kanten), Textlabel, ...

Ein einfacher Layout-Algorithmus für Graphen ist das Circle Layout, welches in Abbildung 1 b) verwendet wird. Die Knoten werden gleichmäßig auf einem Kreis angeordnet und durch gerade oder (besser) gebündelte Kanten verbunden. Für eine erste Übersicht ist dieses Layout gut geeignet, ein Beispiel aus der Software-Visualisierung findet sich auf der Plattform Observable.²⁾ Das Beispiel zu

1 <https://www.ltmuseum.co.uk/collections/stories/design/mapping-london-ionic-tube-map>

2 <https://observablehq.com/@d3/hierarchical-edge-bundling/2>

Abbildung 1 b) hat jedoch bereits verdeutlicht, dass manche Informationen darin nicht unmittelbar erkennbar sind.

Gute Ergebnisse liefert häufig das Force-directed Layout (kräftegetriebenes Layout), wohinter sich eine ganze Klasse von Layout-Algorithmen verbirgt. Die Grundidee basiert auf einem physikalischen Modell: Die Kanten werden als Sprungfedern mit bevorzugter Ruhelänge angesehen, die Knoten als elektrisch geladene Körper, die sich gegenseitig abstoßen. Eine zufällige Anordnung der Knoten besitzt in aller Regel zahlreiche gedehnte oder gestauchte Sprungfedern, so dass das System instabil ist. Es wird dann schrittweise berechnet, wie die wirkenden Federkräfte das System in einen stabilen Zustand bringen; in Abbildung 3 ist dies veranschaulicht (vgl. BATTISTA et al. 1999).

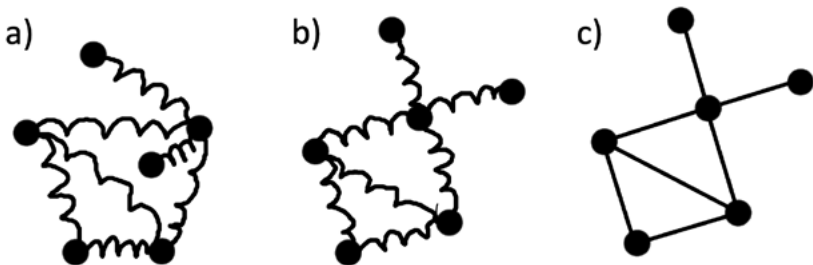


Abbildung 3: Grundidee des Force-directed Layouts (angelehnt an Battista 1999: 303), Volker Ahlers

Um das Prinzip des Force-directed Layouts genauer zu erläutern, brauchen wir einige mathematische Definitionen. Ein Graph $G(V,E)$ besteht es einer Menge $V = \{v_i\}$ von Knoten (engl. *vertices*) und einer Menge $E = \{e_i\}$ von Kanten (engl. *edges*). Die Kanten werden durch Knotenpaare $[u,v]$ definiert mit $u,v \in V$. Wir beschränken uns auf ungerichtete Graphen, in denen die Reihenfolge von u und v unerheblich ist. Die auf den Knoten v an der Position $rv = (rv,x, rv,y)$ wirkenden Kräfte können dann wie folgt berechnet werden.

Anziehende Federkräfte:

mit Federkonstante $c_{att} > 0$, Ruhelänge d_0 und euklidischem Abstand

Abstoßende elektrische Kräfte:

mit Abstoßungskonstante $c_{rep} < 0$.

Der grundlegende Algorithmus lautet dann wie folgt.

1. Starte mit zufälligen Knotenpositionen.
2. Berechne anziehende und abstoßende Kräfte, ändere Knotenpositionen.
3. Wiederhole Schritt 2, bis Kräftegleichgewicht herrscht.

$$\mathbf{F}_{\text{att}}(v) = \sum_{u:|u,v| \in E} c_{\text{att}}(d(\mathbf{r}_u, \mathbf{r}_v) - d_0) \frac{\mathbf{r}_v - \mathbf{r}_u}{d(\mathbf{r}_u, \mathbf{r}_v)},$$

$$d(\mathbf{r}_u, \mathbf{r}_v) = \sqrt{(r_{u,x} - r_{v,x})^2 + (r_{u,y} - r_{v,y})^2}.$$

$$\mathbf{F}_{\text{rep}}(v) = \sum_{u \in V, u \neq v} \frac{c_{\text{rep}}}{(d(\mathbf{r}_u, \mathbf{r}_v))^2} \frac{\mathbf{r}_v - \mathbf{r}_u}{d(\mathbf{r}_u, \mathbf{r}_v)},$$

Um zu verhindern, dass der Algorithmus frühzeitig in suboptimalen Zustand verharrt, werden in jedem Iterationsschritt zufällige Störungen zu den Kräften addiert, deren Stärke mit fortschreitender Zeit verringert wird. Um ein Überschwingen der Knoten zu vermeiden, kann eine mit einem Luftwiderstand vergleichbare Dämpfung ergänzt werden.

Beispiele für Force-directed Layouts finden sich auf der Plattform Observable^[3] sowie in den folgenden Abschnitten. Es gibt viele Varianten des vorgestellten Grundprinzips; insbesondere kann es vorteilhaft sein, abstraktere Kräfte zu definieren, die weniger nah an der realen Physik sind.

Anwendungen in der Bioinformatik

In der Bioinformatik werden unter anderem Sequenzierungsdaten untersucht, aus denen Wechselwirkungen zwischen Genen und Proteinen hervorgehen. Zur Analyse der resultierenden Gen-Protein- oder Protein-Protein-Netzwerke ist eine Visualisierung hilfreich. Eine Herausforderung ist hier, dass die Netzwerke in der Regel sehr groß sind. In dem Projekt BioGranat^[4] wurde in Kooperation zwischen der Hochschule Hannover und dem King's College London eine Java-basierte Software-Plattform entwickelt, die auf einfache Weise durch Analyse- und Visualisierungskomponenten ergänzt werden kann. Abbildung 4 zeigt die vergleichende 3D-Visualisierung zweier Netzwerke desselben Organismus in verschiedenen Zuständen. Die beiden Graphen können fächerartig gegeneinander gedreht werden, so dass gemeinsame und unterschiedliche angeregte Proteine sichtbar werden (vgl. DAND et al. 2013).

3 <https://observablehq.com/@d3/force-directed-graph>

4 <https://sourceforge.net/projects/biogranat/>

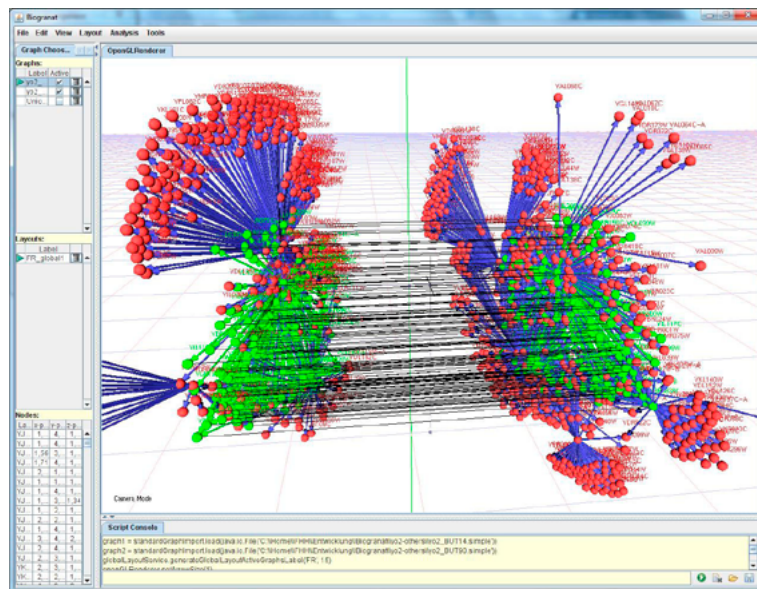


Abbildung 4: Vergleichende 3D-Visualisierung zweier Protein-Protein-Netzwerke desselben Organismus in verschiedenen Zuständen; Projekt BioGranat (vgl. Dand et al. 2013), Volker Ahlers

In Abbildung 5 ist ein Ausschnitt des Komponenten- und Klassendiagramms der BioGranat-Plattform zu sehen. In der Software-Entwicklung werden verschiedene Netzwerkvisualisierungen verwendet (z. B. UML- und Flussdiagramme), um die Struktur eines großen Software-Systems auf einer abstrakten Ebene darzustellen.

Anwendungen in der Netzwerksicherheit

Ein Gebiet, auf dem auf natürliche Weise Netzwerkdaten anfallen, ist die IT-Netzwerksicherheit. In dem Projekt VisITMeta^[5] wurde das IF-MAP-Protokoll verwendet, um mögliche Angriffe oder Bedrohungen im aktuellen Zustand eines IT-Netzes zu erkennen. Das IF-MAP-Protokoll definiert Identifier (z. B. Geräte, IP-Adressen, User-IDs) und Metadata (Beziehungen zwischen Identifier und Metadaten). In der gewählten Visualisierung spiegelt sich dies in zwei

5 <https://github.com/trustathsh/visitmeta>

Eine besondere Herausforderung stellt die dynamische Natur eines IT-Netzes dar, da zum Beispiel mobile Geräte an- und abgemeldet werden oder Nutzer und Nutzerinnen sich ein- und ausloggen. In der Visualisierung werden Änderungen innerhalb eines Zeitfensters durch grüne und rote Leuchteffekte dargestellt, siehe Abbildung 6 (vgl. AHLERS et al. 2016).

Eine Möglichkeit, Angriffe oder Bedrohungen in IT-Netzen zu erkennen, besteht darin, Regeln zu definieren und zu warnen, wenn diese erfüllt sind. Zur Feinabstimmung dieser Regeln und zur Entscheidung, ob es sich um einen falschen Alarm handelt, ist es hilfreich, zu erkennen durch welche Ereignisse oder Netzwerkzustände eine Regel ausgelöst wurde. Zu diesem Zweck sind in Abbildung 7 links der aktuelle Netzwerkzustand und rechts die eine Warnung auslösende Regel gemeinsam dargestellt. In der Software-Anwendung VisIT-Meta ist es möglich, interaktiv die Regel zu ändern und im Sinne einer What-if-Simulation auf den historischen Daten zu testen (vgl. AHLERS et al. 2019).

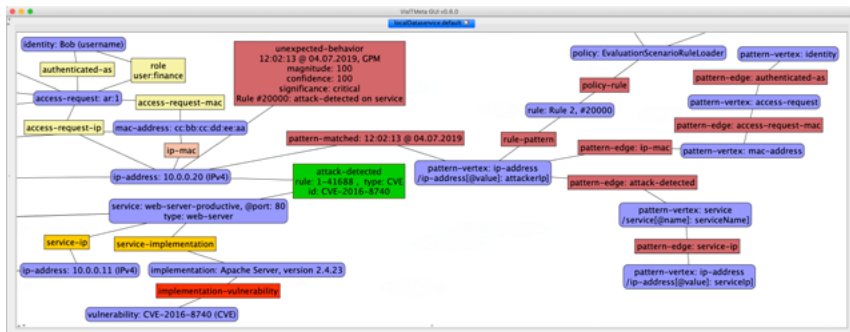


Abbildung 7: Gemeinsame Visualisierung des aktuellen Zustands eines IT-Netzes (links), einer Angriffswarnung (Mitte) und der die Warnung auslösenden Regel (rechts); Projekt VisITMeta (vgl. Ahlers et al. 2019), Volker Ahlers

Ausblick

Besondere Herausforderungen im Hinblick auf die Visualisierung stellen sehr große sowie dynamische Netzwerke dar. Ein Lösungsansatz für sehr große Netzwerke besteht darin, zunächst nur die grobe Struktur zu zeigen, indem eng verbundene Knoten zu Metaknoten zusammengefasst werden, die bei Bedarf interaktiv feiner aufgelöst werden können – falls erforderlich in mehreren Hierarchiestufen. Ein Ansatz für dynamische Netzwerke wurde oben gezeigt, der jedoch nur funktioniert, wenn es in dem betrachteten Zeitfenster vergleichsweise wenige Änderungen gibt. Eine Alternative sind animierte Darstellungen.

Bei der Wahl der grafischen Gestaltungsmittel werden Erkenntnisse der Medizin (z. B. Aufbau der Netzhaut), der Physik (z. B. Natur des Lichts) und der Wahrnehmungspsychologie (z. B. Farbwahrnehmung, Gestaltgesetze) genutzt (vgl. WARE 2013). Eine interdisziplinäre Zusammenarbeit zwischen Fachleuten aus den Bereichen Informatik und Design ist wünschenswert, findet in Forschungsprojekten aber nur selten statt.

Die vorgestellten Forschungsprojekte wurden gefördert durch Mittel des Deutschen Akademischen Austauschdienstes und des British Council (BioGranat: Förderkennzeichen D/07/09921, ARC 1297) sowie des Bundesministeriums für Bildung und Forschung (VisITMeta: Förderkennzeichen 17PNT032).

Literaturverzeichnis

- AHLERS VOLKER; FELIX HEINE; BASTIAN HELLMANN; CARSTEN KLEINER; LEONARD RENNERS, THOMAS ROSSOW; RALF STEUERWALD: Integrated visualization of network security metadata from heterogeneous data sources. In: *Graphical Models for Security (GraMSec 2015)*, Revised Selected Papers. LNCS 9390, S. 18-34. Springer, Cham 2016
- AHLERS, VOLKER; BASTIAN HELLMANN; GABI DREO RODOSEK: A user study of the visualization-assisted evaluation and management of network security detection events and policies. In: *Proceedings of the 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS 2019)*, S. 668-673. IEEE 2019
- DAND, NICK; FRAUKE SPRENGEL; VOLKER AHLERS, THOMAS SCHLITT: BioGranat-IG: A network analysis tool to suggest mechanisms of genetic heterogeneity from exome sequencing data. In: *Bioinformatics* 29(6), 2013, S. 733-741
- DI BATTISTA, GIUSEPPE; PETER EADES; ROBERTO TAMASSIA, IOANNIS G.: *Tollis: Graph Drawing*. Prentice-Hall, Upper Saddle River 1999
- HAMMING, RICHARD W.: *Numerical Methods for Scientists and Engineers*. New York [Dover Publications] 1986
- MCCORMICK, BRUCE H; THOMAS A. DEFANTI; MAXINE D. BROWN (Hrsg.): Visualization in Scientific Computing. In: *Computer Graphics*, 21 (6). New York (USA) [ACM SIGGRAPH] 1987
- MUNZNER, TAMARA: *Visualization Analysis & Design*. Boca Raton (USA) [CRC Press] 2015
- SPENCE, ROBERT: *Information Visualization*. [Pearson Education] Harlow (USA) 2007
- TUFTE, EDWARD R.: *The Visual Display of Quantitative Information*. Cheshire (UK) [Graphics Press] 2001
- WARE, COLLIN: *Information Visualization*. Waltham MA. (USA) [Morgan Kaufmann] 2013