

Zugverbindungen in Frankreich
 4551 Knoten, 7793 Kanten
 (davon 2408 transitiv)

Visualisierung von Verkehrsdaten

von Ulrik Brandes und Dorothea Wagner

Die Visualisierung netzwerkbasierter Informationen ist ein sehr interessantes und gleichzeitig sehr ansprechendes Thema. Das Ziel der angemessenen graphischen Wiedergabe jeglicher Information führt auf eine Vielzahl spannender Fragestellungen, die von ästhetischen Aspekten über die kognitive Wahrnehmung bis hin zur tatsächlichen Erzeugung der Graphiken reichen. Darüber hinaus scheinen „Netzwerke“ in der heutigen Zeit in nahezu allen Bereichen und in den verschiedensten Zusammenhängen von Interesse zu sein. Hauptziel der Visualisierung ist es, die Struktur der Netzwerke zu verstehen, insbesondere wenn sie sehr groß und komplex sind. Der Entwurf eines geeigneten Layouts, d.h. der räumlichen Anordnung in der graphischen Darstellung, ist aber schon für kleine Netzwerke eine mühsame und zeitraubende Arbeit, die nach Automatisierung geradezu ruft. Am Beispiel der Visualisierung von Bahnverkehrsdaten erläutern wir die Formalisierung eines Layoutproblems und stellen daraus resultierende Ergebnisse vor.

Die Visualisierung von Netzwerken wird in der Mathematik/Informatik oft mit dem etwas naiv klingenden Begriff „Graphenzeichnen“ umschrieben. Vermutlich, weil sich eines der frühen bedeutenden Resultate

des Gebietes in einem Aufsatz von William T. Tutte findet, den er 1963 mit dem Titel „How to Draw a Graph“ überschrieb [5]. Interessanterweise veröffentlichte Donald E. Knuth – unabhängig davon –

im gleichen Jahr „Computer Drawn Flowcharts“ [4] und zeigte so eine erste Anwendung in der Informatik auf. Das Feld der für Mathematiker und Informatiker interessanten Aspekte des Graphenzeichnens reicht von Graphen- und Ordnungstheorie über Modellierungsfragen und den Entwurf von Algorithmen bis hin zu Design und Implementation unterstützender Softwaresysteme. Fachübergreifende Anwendungen liegen heute in der visuellen Datenanalyse und graphischen Präsentation von Information beispielsweise in Biologie und Chemie, aber auch in den Ingenieurs- und Sozialwissenschaften. Die zu zeichnenden Graphen repräsentieren so unterschiedliche Dinge wie z.B. Evolutionshierarchien, biochemische Reaktionsketten, Schaltpläne, Streckennetze oder Soziogramme. Entsprechend ist die automatisierte Visualisierung von Graphen ein Thema, das sich wachsender Beliebtheit erfreut. Seit 1993 findet jährlich eine internationale Tagung zu *Graph Drawing* statt, und erste Bücher zu den algorithmischen Aspekten des Graphenzeichnens erscheinen [2, 3].

Das algorithmische Kernproblem liegt in der Bestimmung einer geeigneten räumlichen Anordnung innerhalb des Diagramms. Das *Layout* legt ausgehend von einer vorgegebenen Form der graphischen Repräsentation die Positionen für die sich aus ihr ergebenden *Layoutelemente* fest. Für die naheliegende *geradlinige Repräsentation*, in der Knoten durch Punkte und Kanten durch gerade Linien dargestellt werden, ergeben sich als *Layoutelemente* die Knoten des Graphen, da deren Positionen die gesamte räumliche Anordnung festlegen. Abhängig von der Information, die in der Visualisierung eines Graphen sichtbar werden soll, sind gewisse Eigenschaften eines Layouts besonders wünschenswert; bei der geradlinigen Repräsentation etwa die weitgehende Einhaltung einer vorgegebenen Kantenlänge und die Vermeidung von sich kreuzenden Kanten oder sehr kleinen Winkeln zwischen Kanten. Um die jeweilige Qualität verschiedener Layouts zu vergleichen, wird daher häufig eine *Zielfunktion* definiert, die gewissermaßen die „Häßlichkeit“ eines gegebenen Layouts quantifiziert und daher über alle zulässigen Layouts zu minimieren ist.

Im allgemeinen ist es schwierig, eine globale Zielfunktion anzugeben, die ein Layout als Ganzes bewertet. Einfacher ist es, lokale Kriterien über Konfigurationen kleiner Teilmengen der *Layoutelemente* zu formulieren und diese dann zusammenzufassen. In dem oben bereits erwähnten Artikel [5] bemüht Tutte eine physikalische Analogie, um benachbarte Knoten eines Graphen $G = (V, E)$ in einer geradlinigen Repräsentation nah beieinander zu plazieren. Ist $(x_v)_{v \in V}$ ein Vektor von Knotenpositionen in der Ebene, dann wird der Abstand zwischen benachbarten Knoten $u, v \in V$ mit der Potentialenergie einer Fe-

der der Ideallänge Null bewertet und über alle Kanten des Graphen aufsummiert. Die Zielfunktion lautet demnach

$$\sum_{\{u,v\} \in E} \text{Anziehung}(x_u, x_v),$$

wobei $\text{Anziehung}(x_u, x_v) = d(x_u, x_v)^2$ und $d(x_u, x_v)$ den Euklidischen Abstand der beiden Knotenpositionen bezeichne. Offensichtlich sind optimale Layouts in diesem Modell wenig interessant, da alle Knoten dieselbe Position erhalten. Die Situation ändert sich jedoch, wenn die Positionen einiger Knoten im vorhinein fixiert werden. Tutte zeigt zunächst, daß es ausreicht, in jeder Zusammenhangskomponente des Graphen die Position eines Knotens vorzugeben, um die Existenz und Eindeutigkeit einer Optimallösung zu garantieren, die dann durch Lösen des sich aus den partiellen Ableitungen ergebenden Gleichungssystems effizient bestimmt werden kann. Das optimale Layout legt jeden nicht zuvor fixierten Knoten in den Schwerpunkt seiner Nachbarn, doch besonders interessant ist, daß dieser Ansatz bei dreifach zusammenhängenden planaren Graphen, für die eine beliebige Facette auf einem konvexen Polygon fixiert wird, zu kreuzungsfreien Zeichnungen führt, in denen alle Facetten konvex sind. Ein Beispiel ist in Abbildung 1 gezeigt.

Aus der reichhaltigen Fülle von Anwendungen des Graphenzeichnens wollen wir hier nun ein Problem aus der *Verkehrsanalyse und -planung* herausgreifen, in dem sich eine zwar wesentlich komplexere, jedoch auf ganz ähnlichen Analogien beruhende Zielfunktion als adäquat erwiesen hat. Das Problem entstammt einer Kooperation mit der *TLC/EVA*, der Tochter der *Deutsche Bahn AG*, die alle Informatik- und Logistikprobleme, insbesondere die Aufbereitung der Fahrplandaten und die Fahrplanauskunft, bearbeitet. Das grundsätzliche Ziel der Kooperation ist die Entwicklung von Methoden der Datenreduktion und -visualisierung für die explorative Analyse großer Mengen von Fahrplandaten der europäischen öffentlichen Verkehrsbetriebe. Da die Datensätze von beachtlicher Größe sind, werden sie in einem ersten Schritt zu einem Graphen, dem sogenannten *Fahrplangraphen*, reduziert. Die Knoten dieses Graphen entsprechen allen regulären Halten der Transportmittel (meistens handelt es sich dabei um Bahnhöfe). Da die Knoten folglich bereits mit Koordinaten entsprechend der geographischen Lage der Halte gegeben sind, ist für die Visualisierung zunächst gar kein Layoutproblem zu erkennen. Im Fahrplangraphen sind zwei Knoten durch eine Kante verbunden, wenn es eine Fahrverbindung (z.B. einen Zug) gibt, die von einem der zugehörigen Halte ohne Zwischenhalt zum anderen verkehrt.

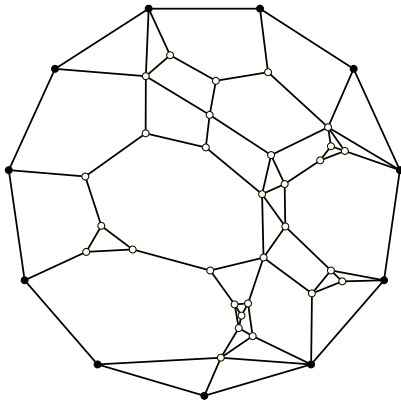


Abb.1. Optimales Layout eines dreifach zusammenhängenden planaren Graphen in Tuttes Modell [5] (die dunklen Knoten wurden fixiert)

Ein wichtiges Element der Datenanalyse ist die Klassifikation der Kanten des Fahrplangraphen in zwei Kategorien: *minimale* und *transitive* Kanten. Minimale Kanten sind solche, die zu einer direkten Verbindung ohne durchfahrenen Zwischenhalt gehören. Typischerweise werden sie durch Regionalzüge induziert, die „an jedem Baum“ halten. Transitive Kanten resultieren aus Verbindungen, die wenigsten einen Halt passieren und werden daher in der Regel durch Schnellzüge induziert. Diese Klassifikation geschieht automatisch und kann leicht durch verschiedene Kantenfarben angezeigt werden. Doch schon bei kleinen Beispielen wird deutlich, daß die naheliegende Visualisierung durch eine geradlinige Repräsentation, in der die Knoten entsprechend ihrer geographischen Koordinaten plazierte werden, zu schwer lesbaren Zeichnungen führt, in denen zahlreiche Kanten in spitzen Winkeln zusammenlaufen oder sich gar überlappen. Da minimale Kanten gewöhnlich Schienenstrecken entsprechen, kann ihre geradlinige Repräsentation problemlos beibehalten werden. Für transitive Kanten erscheint jedoch, entsprechend ihrer Bedeutung als Zwischenhalte überfahrende Verbindungen, die Darstellung durch gekrümmte Kurven sinnvoller. Wir haben uns aus verschiedenen Gründen für kubische Bézier-Kurven entschieden. Abbildung 2 macht den Vorteil gegenüber einer geradlinigen Repräsentation deutlich. Allerdings haben wir uns dadurch dann doch ein Layoutproblem eingehandelt, nämlich die Platzierung der Stützpunkte.

Zu einem gegebenen Fahrplangraphen $G = (V, E)$ wollen wir wünschenswerte Eigenschaften für ein Layout der Stützpunkte beschreiben. Die Knoten von G werden im folgenden mit den Positionen entsprechend ihrer geographischen Lage identifiziert, und

minimale Kanten und sehr lange transitive Kanten als gerade Linien repräsentiert.¹ Um nun den Verlauf der Bézier-Kurven zur Darstellung transitiver Kanten zu bestimmen, müssen nur deren Stützpunkte positioniert werden. Ist $\check{E}_{\tau_1} \subseteq E$ die Menge der transitiven Kanten, deren Länge unter einem vorgegebenen Schwellwert τ_1 liegt, dann besteht die Menge der Layoutelemente, $L = \{b_u(e), b_v(e) : e = \{u, v\} \in \check{E}_{\tau_1}\}$ aus den beiden inneren Stützpunkten für die Kanten in \check{E}_{τ_1} , und es sind Positionen $x_b, b \in L$, zu bestimmen. Wir nennen Stützpunkte, die zu derselben Kante gehören, *Partner* und bezeichnen die *Verankerung* v eines Stützpunkts $b_v(e) \in L$ mit $a_{b_v(e)}$. Die *Initialpositionen* der Stützpunkte seien so festgelegt, daß sich jeder Stützpunkt auf der geraden Linie durch die Endpunkte seiner Kante in gleichem Abstand von Partner und Verankerung befindet. Das Layout soll nun für jeden Stützpunkt die Positionen von Partner, Verankerung, allen Stützpunkten, welche die gleiche Verankerung oder eine nahe gelegene Initialposition haben, und die Knoten von G in der Nähe der Initialposition berücksichtigen. Nähe zu einer Initialposition sei dabei über eine parametrisierte Ellipse um die (gedachte) geradlinige Repräsentation der zugehörigen Kante definiert. Die Menge aller Halte und anderen Stützpunkte, die in der zu einem Stützpunkt b gehörenden Ellipse liegen, werde mit \mathcal{E}_b bezeichnet. Wie oben erläutert, konstruieren wir unsere Zielfunktion aus intuitiven lokalen Kriterien, die auf physikalischen Analogien beruhen.

Abstand zu Knoten. Stützpunkte sollten nicht zu nah bei den (fix positionierten) Knoten des Graphen liegen. Ähnlich der Anziehung zwischen benachbarten Knoten in Tuttes Modell wird dazu eine parametrisierte Abstoßungsfunktion

$$\text{Abstoßung}(x_b, x_{b'} | \varrho) = \frac{\varrho^4}{d(x_b, x_{b'})^2}$$

definiert. Um ausreichenden Abstand zu den umliegenden Knoten zu gewährleisten, enthält die Zielfunktion für jeden Stützpunkt $b \in L$ zu einer Kante $\{u, v\} \in \check{E}_{\tau_1}$ den additiven Beitrag

$$\sum_{s \in \mathcal{E}_b \cap V} \text{Abstoßung}(x_b, s | \varrho_1 \cdot \lambda_b),$$

wobei $\lambda_b = \frac{d(u,v)}{3}$ die Initiallänge und ϱ_1 ein Dehnungsfaktor ist. Damit die Kurven sich nicht zu stark von der geradlinigen Repräsentation unterscheiden, wird dieses Kriterium für die Verankerung leicht modifiziert. Anziehung und Abstoßung lassen sich zu einem Abstandskriterium $\text{Abstand}(x_b, x_{b'} | \lambda) =$

¹ Die Abbildung auf Seite 11 macht deutlich, warum es nicht sinnvoll ist, alle transitiven Kanten als Bézier-Kurven zu repräsentieren.

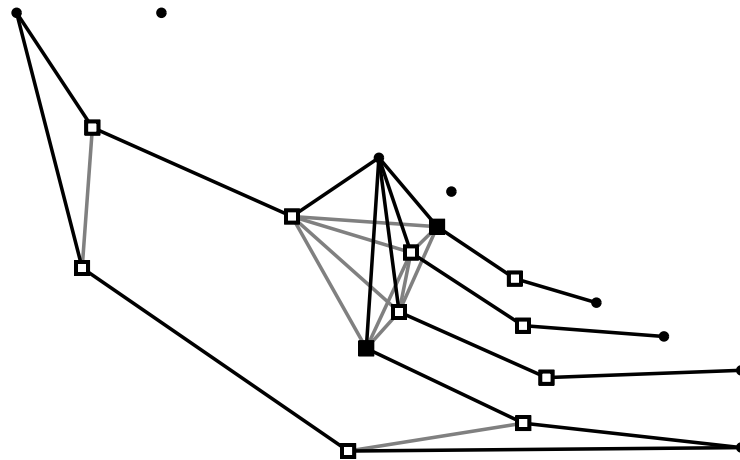
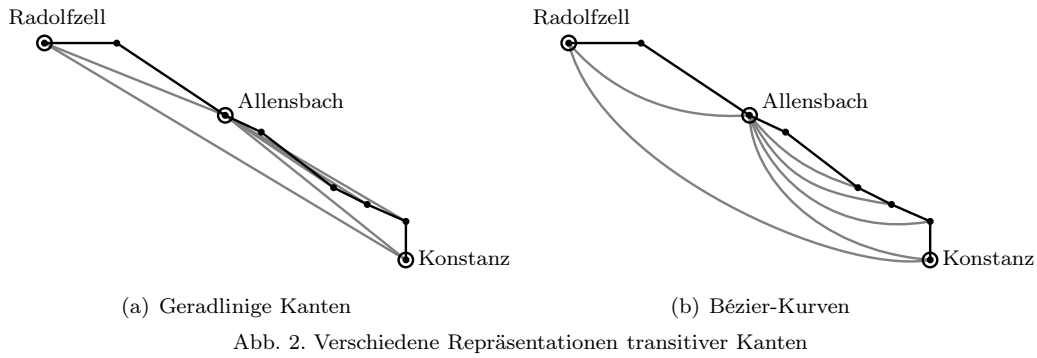


Abb. 3. Imaginärer Hilfsgraph für das Beispiel aus Abb. 2. Die durch schwarze Vierecke gekennzeichneten Layoutelemente sind nicht gebündelt, da sich die Ziellängen ihrer zugehörigen Segmente zu sehr unterscheiden (Schwellwert τ_2).

Abstoßung $(x_b, x_{b'} \mid \lambda) + \text{Anziehung}(x_b, x_{b'})$ kombinieren, das für $d(x_b, x_{b'}) = \lambda$ minimal wird. Durch Beiträge

$$\text{Abstand}(x_b, a_b \mid \lambda_1 \cdot \lambda_b)$$

mit einem Dehnungsfaktor λ_1 werden Stützpunkte genügend nahe an ihren Verankerungen gehalten.

Abstand von Stützpunkten. Wie bei nahegelegenen Knoten sollte ein Stützpunkt $b_1 \in L$ auch nicht zu nahe bei einem anderen Stützpunkt b_2 mit nahe gelegener Initialposition enden. Falls b_1 weder Partner von b_2 ist, noch an b_2 gebunden (s.u.) ist, ergänzen wir die Zielfunktion daher um

$$\text{Abstoßung}(x_{b_1}, x_{b_2} \mid \varrho_2 \cdot \min\{\lambda_{b_1}, \lambda_{b_2}\}).$$

Der gewünschte Abstand zwischen Partnern b_1 und b_2 ist identisch mit dem gewünschten Abstand von der jeweiligen Verankerung,

$$\text{Abstand}(x_{b_1}, x_{b_2} \mid \lambda_1 \cdot \lambda_{b_1}).$$

Bündelung. Um den visuellen Eindruck des Auslassens kleinerer Halte zu verstärken, sollen transitive

Kanten mit einem gemeinsamen Endknoten bevorzugt auf der gleichen Seite einer Schienenstrecke platziert werden. Stützpunkte $b_1, b_2 \in L$, die dieselbe Verankerung haben, werden dazu durch imaginäre Kanten *gebündelt*, falls λ_{b_1} nicht zu sehr von λ_{b_2} abweicht, d.h. falls $\frac{1}{\tau_2} < \frac{\lambda_{b_1}}{\lambda_{b_2}} < \tau_2$ für einen Schwellwert $\tau_2 \geq 1$ (andernfalls würden kurze transitive Kanten von langen transitiven Kanten verzerrt). Die Formalisierung dieses Kriteriums lautet

$$\beta \cdot \text{Abstand}\left(x_{b_1}, x_{b_2} \mid \lambda_2 \cdot \frac{\lambda_{b_1} + \lambda_{b_2}}{2}\right),$$

wobei λ_2 der Dehnungsfaktor für die Länge der Bündelungskanten und β die Wichtigkeit der Bündelung im Verhältnis zu den anderen Beiträgen ist.

Letztlich modelliert die resultierende Zielfunktion das Layout einer geradlinigen Repräsentation für des in Abbildung 3 gezeigten imaginären Hilfsgraphen. Seine Knoten sind die (bereits positionierten) Knoten des Fahrplangraphen und die (noch zu positionierenden) Stützpunkte. Die Kanten des Hilfsgraphen entsprechen den Segmenten der Bézier-Kurven und

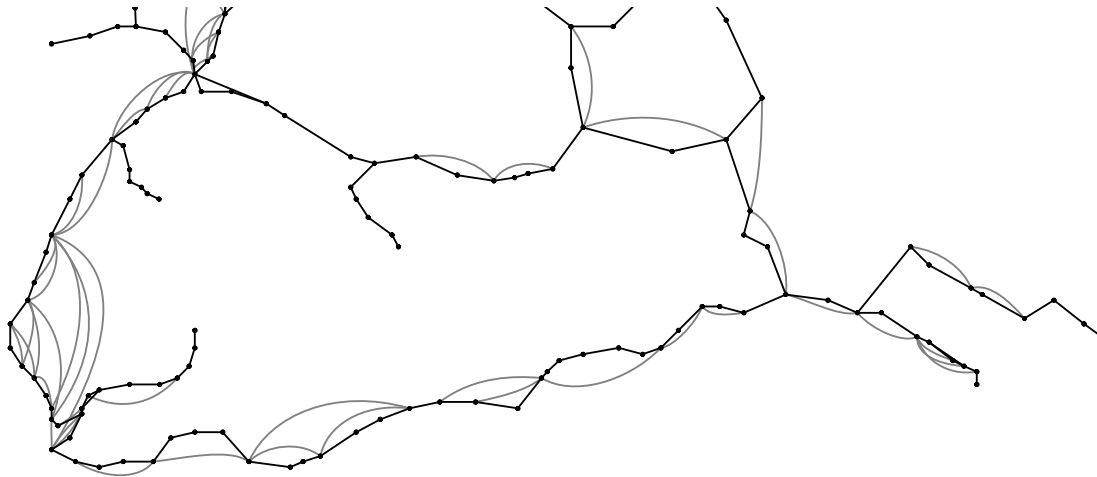


Abb. 4. Regionalzüge entlang der Rhein-Strecke von Konstanz über Basel nach Freiburg

den Bündelungskanten, jeweils mit einer vorgegebenen Ziellänge. Anders als in Tutes Modell ist die Minimierung der Zielfunktion \mathcal{NP} -schwer, so daß wir ein iteratives Verfahren zur näherungsweise Lösung verwenden. Durch entsprechende Wahl der zahlreichen Parameter lassen sich unterschiedliche Charakteristiken der Kurvenverläufe erzielen.

Die hier beschriebene Zielfunktion ist das Ergebnis zahlreicher experimenteller Studien mit verschiedenen lokalen Kriterien. Unter anderem hat sich dabei die Konvexitätseigenschaft von Bézier-Kurven (die Kurve verläßt die konvexe Hülle ihrer Stützpunkte nicht) als deutlicher Vorteil gegenüber anderen Splines erwiesen. Zur Zeit untersuchen wir andere Ansätze, um Layouts von zumindest vergleichbarer Qualität so schnell zu bestimmen, daß sie in einem interaktiven Tool zur Analyse von Fahrplangraphen berechnet werden können.

Literaturverzeichnis

- [1] U. Brandes und D. Wagner. Using graph layout to visualize train interconnection data. *Proc. Graph Drawing '98*, Lecture Notes in Computer Science 1547, pp. 44–56. Springer, 1998.
- [2] G. Di Battista, P. Eades, R. Tamassia und I.G. Tollis. *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice Hall, 1999.
- [3] M. Kaufmann, D. Wagner (Hrsg.). *Drawing Graphs: Methods and Models*. Teubner, 2000.
- [4] Donald E. Knuth. Computer Drawn Flowcharts. *Communications of the ACM*, 6(9):555–563, 1963.
- [5] W.T. Tutte. How to draw a graph. *Proceedings of the London Mathematical Society, Third Series*, 13:743–768, 1963.

Adresse der Autoren

Ulrik Brandes
 Dorothea Wagner
 Universität Konstanz
 Fakultät für Mathematik und Informatik
 78457 Konstanz
 ulrik.brandes@uni-konstanz.de
 dorothea.wagner@uni-konstanz.de