

平成 30 年 6 月 21 日現在

機関番号：32604

研究種目：基盤研究(C) (一般)

研究期間：2014～2017

課題番号：26330348

研究課題名(和文) 属性付きグラフのレスポンス可視化の研究

研究課題名(英文) Responsive Visualization of Property Graphs

研究代表者

藤村 考 (Fujimura, Ko)

大妻女子大学・社会情報学部・教授

研究者番号：90634642

交付決定額(研究期間全体)：(直接経費) 2,400,000円

研究成果の概要(和文)：オープンソース開発コミュニティやソーシャルメディアなどでは、大規模なネットワーク構造のデータがよく出現し蓄積される。本研究では、ユーザがこのようなネットワークデータの構造特性を容易に発見できるようにする新しい可視化手法を提案する。また、このようなネットワークの成長の過程を3Dアニメーションで簡単かつ直観的に表現する可視化ツールを開発した。そのツールはthree.jsおよびNeo4jグラフデータベースを使用して実装され、Webブラウザ上で簡単に操作可能である。

研究成果の概要(英文)：In an open source development community, or social networks, a large-scale network-structured data is often appeared and accumulated. This research proposes a new visualization method that enables users to easily discover the structural characteristics of such network-structured data. We also developed a visualization tool that shows the evolution process of such networks in three-dimensional animation intuitively. Our visualization tools are implemented using three.js library and the graph database Neo4j and it can be operated easily on a Web browser.

研究分野：データマイニング, 情報可視化

キーワード：情報可視化 グラフ可視化 ストリーミンググラフ グラフデータベース アニメーション git 3DCG

1. 研究開始当初の背景

ソーシャルメディアやスマホアプリの普及により、個人や企業を問わず、クラウド上に大量のデータが蓄積されるようになった。このような大量のデータをマーケティング等に活用しようとするニーズが高まっている。しかし、大量のデータから有用なデータを分析し、発見するためのデータマイニングのスキルを有する分析者（いわゆるデータサイエンティスト）は不足しており、データの活用が進んでいないという状況がある。

データマイニングは一般に、分析する観点（切り口）が良ければうまくいくが、その切り口を見つけることが難しく、分析者の経験やスキルに頼っている部分が多い。

そこで、本研究テーマでは、分析者が多様な属性をインタラクティブに選択し、選択された属性によって分析結果がどう違うかをアニメーションによりわかりやく分析者に認知させるレスポンス可視化(responsive visualization)と呼ぶアプローチにより、分析の切り口が発見しやすくなることを期待できると考えた。これにより、現在、産業界で人材不足が深刻化しているデータサイエンティストに依存することなく、将来的に営業活動等に従事している現場の担当者が容易に大量のデータを分析できるようになることを目指す。

2. 研究の目的

本研究の目的は、購買履歴やソーシャルメディア等にみられるネットワーク型データの新しい可視化手法を確立することである。特に多数の属性が付与されているグラフにおいて、属性毎にどのような構造上の違いがあるかどうかを人間が容易に認知可能にするための情報可視化技術を確立することにある。本研究では、多数の属性の中からデータ分析者がインタラクティブに選択した属性で絞り込んだマイニング結果を瞬時に反映するレスポンス可視化というアプローチにより、この課題に取り組み、実装と評価を行う。

3. 研究の方法

その要素技術の開発を進める上でのキーポイントは、近年の Chrome 等の高性能な JavaScript 実行環境の活用と、HTML5 の中に取り入れられた WebGL の 3D モデルをブラウザ上の JavaScript から容易に操作可能にする three.js 等のオープンソースのライブラリの活用である。これらを駆使することで、WebGL による美しい 3D 表現と、JavaScript による高いインタラクティブ性を両立させた新しい属性付きグラフの可視化が可能になると考えた。具体的には、多数の属性の中からデータ分析者がインタラクティブに選択した属性がネットワーク構造の中でどのような特徴があるかをリアルタイムに分析し、可視化結果に瞬時に反映するシ

ステムを開発してその有用性を評価する。

本研究で取り組む属性付きグラフのレスポンス可視化の主な研究課題は次の3項目である。

(課題1) 人間が認知しやすい可視化表現をデザインすること

(課題2) 汎用性が高く、高効率の実装方式を明らかにすること

(課題3) レスポンス可視化の評価指標を明らかにすること

課題1は、系統的なアプローチで取り組むことは困難であるため、アイデア→実装→評価の繰返しでデザインを改善することで、よりよいデザインを生み出す。課題2では、汎用性の高い属性付きグラフのデータ交換形式(データをシステムにインポートするためのデータ形式)と、Webブラウザとサーバとの間の役割分担とデータ通信の実装方式(リアルタイムな可視化を実現するために、WebSocket 上の独自プロトコルを検討する。課題3では、レスポンス可視化という新しい可視化手法の効果をどのような評価指標で評価するべきかという根本的な問題から検討し、その結果を踏まえて評価実験を実施する。

4. 研究成果

本研究では、最初に上記課題2を達成するため、様々なアプローチを検討した結果、グラフデータベース Neo4j を活用することにより主要な問題を解決することができた。

Neo4j は属性付きグラフをそのままデータベースに格納することができるだけでなく、Cypher と呼ぶクエリ言語を備え、この Cypher によって柔軟な条件でサブグラフを抽出できるからである。これにより、多様なデータを扱える汎用的なグラフ可視化システム CypherVis3D を開発した。

従来のグラフ可視化システムでは、ユーザをノード、エッジを友人関係とした可視化のように、一つの可視化空間に配置されるノードについてはユーザ(人)という均質の種類のノードを扱うのが一般的であった。一方、CypherVis3D では一つの可視化空間において、「消費者」、「カート」、「商品」といった全く異なる種類のノードが混在することを前提とする。このような多様な種類のノードを分かりやすく表現するため、ノードを 3D モデルで表現した。さらに、Cypher によって抽出された多様なノードやエッジにおける複数の属性は、従来、大きさ、色、濃淡、形状等の様々な視覚属性を使うことにより、同時に視覚的な表現が可能であるが、人間はそのような多数の視覚属性を同時に把握することはできない。そこで CypherVis3D では、データ分析者がインタラクティブに選択した属性を瞬時に可視化結果に反映するインタフェースを導入した。具体的には、分析者のインタラクティブな属性の選択によって、特定

の属性を有するノードを一定の高さに持ち上げて表示する 3D 表現を使ったインタフェースである。

また、CypherVis3D では、Cypher クエリによって抽出されたエッジを指定した順序でソートすることができるので、ストリーミンググラフの生成が容易となるというメリットがある。これにより例えば、時系列でソートしたグラフを可視化の入力することにより、グラフの生成過程をアニメーションとして表現可能になった。つまり、課題 2 については、独自プロトコルではなく、WebSocket 上に Cypher を送信するという方式により目的を達成することができた。

その後、CypherVis3D を、様々なデータに対して適用する実験を実施した。適用したデータはパナマ文書とオープンソースのソフトウェアリポジトリである GitHub のコミット履歴データである。パナマ文書は膨大な数のオフショア法人とそれらの役員や株主の関係が含まれるデータであり、それらのネットワーク成長過程を可視化することで新しい発見ができるのではないかと考えた。パナマ文書を CypherVis3D で可視化することは容易であり、本システムの汎用性は実証できたが、残念ながら、パナマ文書のデータセットは疎結合のネットワークであったため、連続したデータが比較的少なく、CypherVis3D の時系列の分析のメリットは限定的なものであることが分かった。

一方、GitHub のコミット履歴は、中心的なコントリビュータが継続してソースファイルを更新するため、時系列分析に向いていることが分かった。

しかし、CypherVis3D では、時間推移に伴い、単純にノードとエッジが追加される様子をアニメーションとして表現するものであったが、その結果は常に変化しつづけるため、一定期間経過した後、いつ誰がどのファイルを頻繁に更新したかといった全体像を把握することは難しい。

2017 年度では、このような背景からレスポンス可視化の中で時系列に変化する動的グラフの可視化にフォーカスして検討を進めた。そこで 2017 年度では、3D 空間の z 軸を時間軸として表す新しいグラフ可視化ツール gitVis3D を開発した。これにより、x-y 平面はその時のノード間の関係を表すスナップショットとみなすことが可能となり、当時どのオブジェクトが存在し、それらを誰が作成または変更したのかを直観的に把握することが可能になった。この成果は課題 1 の人間が認知しやすい可視化表現をデザインするという課題に関する一つの成果である。

また、2016 年度に開発した GitHub のコミット履歴を Neo4j にインポートするためのツールは、バッチで動作するものであったが、2017 年度は GitHub API を使ってリアルタイムに GitHub からコミット履歴を Neo4j にインポートする機能を開発した。これらにより、

GitHub リポジトリにある任意のプロジェクトの可視化が可能となり、実用性の高いツールを作ることができた。

一方、課題 3 の評価指標に関しては、従来から、タスクの設定とその達成時間の比較による評価が用いられてきたが、このような指標はタスクの設定が恣意的になる可能性があり、新たな評価指標を模索した。しかし、本研究では、結果として客観性の高い評価指標を提案することができなかった。これらに関しては今後の課題である。

WebGL の普及により、Web ブラウザ上で 3D グラフィックが高速にレンダリング可能となった。しかし、3D 空間の特徴を活かした実用的なグラフ可視化ツールは、まだほとんど存在していない。本研究により、3D 空間における時間軸表現の有用性及び three.js/HTML5 上で十分な処理性能が達成できることを確認することができた。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表] (計 4 件)

1. Ko Fujimura, 3D Visualization of Versioning Graph, The 11th IEEE Pacific Visualization Symposium (PacificVis), 2018.
2. 藤村 考, GitVis3D - バージョニング・グラフの 3D 可視化ツール, 電子情報通信学会 ライフインテリジェンスとオフィス情報システム研究会, 2018.
3. 藤村 考, Neo4j を用いたストリーミンググラフの 3D 可視化, 電子情報通信学会 ライフインテリジェンスとオフィス情報システム研究会, 2016.
4. 藤村 考, Neo4j に格納された属性付きグラフの 3D レスポンシブ可視化, 電子情報通信学会 ライフインテリジェンスとオフィス情報システム研究会, 2016.

[その他]

1. Ko Fujimura, gitVis3D - 3D Visualization of Git Repository, <https://github.com/kofujimura/gitVis3D>
2. Ko Fujimura, CypherVis3D - 3D Visualization of streaming data retrieved by Cypher Query Language provided by Neo4j, <https://github.com/kofujimura/cypherVis3D>

6. 研究組織

(1) 研究代表者

藤村 考 (KO FUJIMURA)
大妻女子大学・社会情報学部
情報デザイン専攻・教授
研究者番号：90634642

(2) 研究分担者
なし

(3) 連携研究者
なし

(4) 研究協力者
なし