

# 大規模Cソースコードの ヘッダ依存関係を可視化

宮下 修人

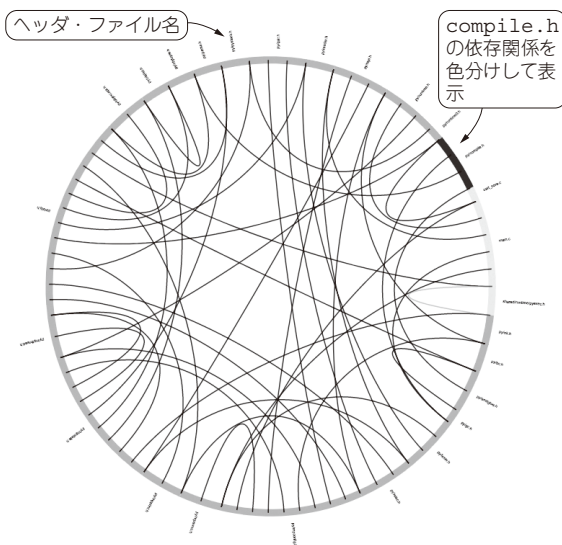


図1 本稿でやること…大規模なCソースコードのヘッダ・ファイルの依存関係をデータ化して可視化する  
MicroPython(minimal版)のヘッダ・ファイルの依存関係を表現するコード・ダイアグラムの表示例

データ・サイエンスは、データを扱う分野全般に適用可能な技術です。それであれば、大規模なソース・ファイルを管理することの多い組み込み開発の現場でもデータ・サイエンスの考え方を適用できるはずです。(編集部)

本稿では、オープンソースの組み込みソフトウェアであるMicroPythonを題材にして、C言語のソースコードにおけるヘッダ・ファイルの依存関係(図1)を表現するデータを使ったネットワーク分析を例題として紹介します。ネットワークの可視化と簡単なネットワーク指標を使った分析について解説します。

## ネットワーク分析の基礎知識

### ● ネットワーク分析とは何か？

本稿で解説するネットワーク分析を理解するには、そもそもネットワークとは何なのかについて確認する

必要があります。

#### ▶物事の関係構造をネットワークでモデル化する

ネットワークとは、「A君とBさんは友達である」という友人関係や、「ライブラリCはライブラリDを読み込んでいる」という依存関係といった物事の間を表現する構造(関係構造)を点(ノード)、線(エッジ)で抽象化したものです。

ネットワーク分析は、このような物事の間を表現する構造をモデル化して、ネットワーク全体や、ネットワークの構成要素であるノードやエッジの性質について調べる分析の方法です。

#### ▶関係構造に着目してシステム全体を探る

ネットワーク分析は、データ・サイエンスにおいて主流派である、回帰分析のような統計分析とはやや異なる特徴を持つ分析方法です。

データ・サイエンスの主流派である統計分析は基本的に、個々のデータが有する属性の情報に基づいて統計モデルを構成する要素について分析を行います。それに対しネットワーク分析は、個々の要素が持つ属性ではなくその関係構造に着目し、システム全体の特徴や個々の要素の役割について探ります。ただし、属性と関係構造は必ずしも相反する概念ではなく、あくまで相補的に運用することが肝要です。

### ● 有向グラフと無向グラフ

ネットワークはノードとエッジによって構成されます。ノードの間の関係を表すエッジに向きが有るか無いかで有向グラフと無向グラフの2種類に分類できます。

#### ▶分類

有向グラフとはノードの間の関係を表すエッジに向きが有るネットワークであり、例えば本稿で対象としているようなヘッダ・ファイルやライブラリの依存関係のネットワークは図2(a)に示す有向グラフで表現されます。それに対し、対等な友人関係や協力関係といった、関係を表すエッジが特定の方向を持たないネットワークのことを無向グラフと呼びます[図2(b)]。

#### ▶ネットワーク分析では区別が重要

ネットワーク分析を行うとき、対象とするネット