Network Analysis in Python

A Brief Introduction

Oliver Posegga

University of Bamberg,
An der Weberei 5, 96047 Bamberg, Germany
http://www.uni-bamberg.de/sna/

Abstract. This tutorial provides an entry-level introduction to social network analysis using Python and NetworkX. We discuss Python as a platform for network analysis and provide a systematic overview on the available libraries that elevate Python to a powerful toolbox for network research. Further, we introduce the fundamental concepts of network analysis and visualization, which we illustrate with practical examples based on a freely available dataset that we analyze with the software introduced in the beginning of the tutorial.

1 Goals

The goal of this tutorial is to provide a brief but comprehensive introduction to social network analysis using Python. After attending the tutorial, participants will be familiar with the basic concepts of network analysis, know how to analyze networks from a multi-level multi-theory perspective, understand the fundamentals of the Python ecosystem and know how to quickly setup their own Python environment. Further, they will be able to model, analyze, and visualize simple networks from freely available datasets using contemporary Python libraries.

2 Requirements

This course is designed to be an entry-level tutorial for individuals interested in social network analysis and serves as a starting point and overview on the topic. Participants are not required to be familiar with social network analysis and related software in general. Previous knowledge, however, will be beneficial to understand the topics of the tutorial. Participants who are already familiar with social network

analysis and Python are not the primary target audience of the tutorial, but might be interested in some of the advanced topics (e.g. interactive network visualization using Python and d3.js [2]), which are briefly discussed and demonstrated as an outlook at the end of the tutorial.

3 Structure and Content

The tutorial covers three major topics, i.e. (1) the Python ecosystem, including network analysis and visualization libraries, (2) the fundamentals of network analysis, and (3) the fundamentals of network visualization. The contents discussed for each topic are briefly outlined in the following.

3.1 Python and Social Network Analysis

There are multiple readily available software solutions that come with many of the different methods and techniques falling into the domain of social network analysis, some of which have been around for decades (e.g. UCINET [8], Pajek [7], Gephi [4]). Recently, however, Python and the ecosystem evolving around it have gained popularity in the network community. As a programming language, Python is known for its intuitive syntax, readability, extensibility, versatility, cross-platform availability, high degree of customizability, and the large community that has emerged from it. As a platform for social network analysis, it primarily benefits from the extensive number of libraries and extensions contributed by this community, which elevate Python from a simple programming language to a flexible and powerful ecosystem with a wide range of scientific applications.

Over the course of the tutorial, we provide a brief overview of the various libraries that are necessary to use Python for social network analysis. For this tutorial, we use a freely available Python distribution, i.e. Anaconda [1], which provides a comprehensive scientific Python environment, including Jupyter [5], a web-based, interactive development environment, formerly known as iPython.

We briefly discuss the scientific Python environment and its setup, before we demonstrate how Anaconda [1] can be used in conjunction with Docker [3] to provide a flexible cross-platform environment for network analysis. Further, we provide an overview of the most commonly used network analysis libraries for Python and continue the tutorial with a practical introduction to the NetworkX [6] library using practical examples based on a freely available dataset.

3.2 The Fundamentals of Social Network Analysis

We discuss the basic concepts of network analysis using the previously introduced dataset based on simple analyses conducted with NetworkX [6]. Over the course of this introduction, we follow an exploratory analysis pattern [9], which comprises the following steps: *Definition* of nodes and edges, *manipulation* of the network, *computation* of network measures, *visualization* of the network.

In the definition step, we discuss the implicit and explicit assumptions that have to be made when modeling network structures from different types of network data. With regard to the manipulation step, we explain several approaches to querying and manipulating network structures using NetworkX [6]. After covering those steps, we proceed with the introduction of well-known network measures along a multi-level multi-theory framework [10], which systematically captures different units of analysis, ranging from the level of individual actors to the network level. Finally, we cover the visualization of networks in general, and in Python in particular, in a dedicated section of this tutorial.

3.3 The Fundamentals of Network Visualization

One of the profound strengths of network analysis lies within the beauty and intuitive nature of network visualizations. While it has become deceptively easy to visualize networks using tools like Gephi [4], which provide easy access to a variety of sophisticated layout algorithms and a plethora of useful visualization features, creating meaningful visualizations requires a systematic understanding of their building blocks.

We briefly discuss those building blocks and provide an overview of the most common layout algorithms used in practice. Using the dataset analyzed in the first part of this tutorial, we demonstrate how to create simple network visualizations using NetworkX [6].

The last part of this tutorial is dedicated to the discussion of the interactive visualization of networks. We demonstrate two approaches to create such visualizations: The first approach is based on exporting network data from Python and importing them into Gephi [4]. The second approach utilizes the d3.js [2] framework in conjunction with Python and NetworkX [6].

References

1. Anaconda. https://anaconda.org/
2. d3js. https://d3js.org/
3. Docker. https://www.docker.com/
4. Gephi. https://gephi.org/
5. Jupyter. http://jupyter.org/
6. Networkx. https://networkx.github.io/
7. Pajek. http://mrvar.fdv.uni-lj.si/pajek/
8. Ucinet. https://sites.google.com/site/ucinetsoftware/home/
9. De Nooy, W., Mrvar, A., Batagelj, V.: Exploratory social network analysis with Pajek, vol. 27. Cambridge University Press (2011)
10. Monge, P.R., Contractor, N.S.: Theories of communication networks. Oxford University Press, USA (2003)