

Netzwerkanalyse SS16

Projektbeschreibung

Viktor Dinkel

Paper zum Nachbauen/Imitieren:

„Network Analysis of the Stock Market“

http://web.stanford.edu/class/cs224w/projects_2015/Network_Analysis_of_the_Stock_Market.pdf

1. Welches Problem wird gelöst?
Generierung, Visualisierung und Analyse eines Netzwerks aus Börsendaten mit der Anleitung gegeben im Paper „Network Analysis of the Stock Market“. Mit dieser Netzwerkanalyse wird ein Portfolio von verschiedenen Aktien zusammengestellt, deren Kurse korrelieren und sie mit einer Zentralitätsfunktion eine optimierte Auswahl darstellen, so dass eine gemeinsame Investition sinnvoll wäre. Es wird keine Vorhersage über den Aktienkurs getroffen, ob gekauft/verkauft werden soll. Im Paper wird eine Visualisierung der Finanzkrise im Jahr 2007 vorgenommen, wo gezeigt wird, dass sie von wenigen (wichtigen) Finanzgrößen im Aktienmarkt ausging. Das, oder etwas Ähnliches, wird in diesem Projekt nicht nachgebildet.
2. Welche Daten werden genutzt?
Es werden historische Börsendaten von Plattformen wie Yahoo Finance oder dem Python Paket Quandl genutzt, voraussichtlich für das Jahr 2015. Die Daten werden also mittels einer API der entsprechenden Anbieter heruntergeladen. Ob es nur eine Auswahl von Börsenkursen, oder ob es alle Kurse vom gesamten Jahr sein werden, ist abhängig von der Verfügbarkeit der Daten.
3. Wie werde ich das Projekt bearbeiten?
 - a) Datenbesorgung (inkl. Programmierung der Schnittstellen)
 - b) Generierung des Netzwerks mittels Python (Networkx, Kurskorrelation mit Verzögerung).
 - c) Clustering des Netzwerks (im Paper ist von community detection mittels Fruchterman Reingold Algorithmus in Gephi die Rede, ich werde voraussichtlich nicht Gephi nutzen)
 - d) Analyse des Netzwerks, speziell bezüglich Zentralitäten: degree, betweenness & closeness
 - e) Bestimmung der Kandidaten für ein Portfolio mittels optimierender Zentralitätsfunktion
4. Welche Algorithmen/Techniken/Modelle werde ich nutzen?
 - Umsetzung in Python mit Networkx
 - Korrelationsfunktion angewandt auf Kursmatrix
 - Netzwerkgenerierung aus Adjazenzmatrix
 - Clustering/Community-Erkennung (genaues Verfahren noch unbekannt)
 - Analyse: degree-, betweenness-, closeness-centrality, degree-distribution
 - Kandidatenbestimmung mittels Zentralitätsfunktion
5. Wie bestimme ich, ob das Projekt erfolgreich abgeschlossen wurde?
Das resultierende Portfolio wird bewertet:
 - a) Auswertung der Aktienkurse des Portfolios hinsichtlich deren „training-set“
 - b) Leistungsauswertung des Portfolios hinsichtlich Daten außerhalb des „training-set“