

Zusätzliche Datensätze für heute: <http://www.kai-arzheimer.com/Netzwerkanalyse/datensaetze2.zip>

## 1 Pajek kennenlernen

### 1.1 Datentypen, Dateien und Menüs

Pajek wurde von Mathematikern entwickelt und liest eine Vielzahl von Formaten aus dem naturwissenschaftlich-technischen Bereich direkt ein. Pajek selbst kennt eine Reihe von Datentypen (vs. -formate). Zu den wichtigsten gehören Netzwerke (\*.net), Partitionen (\*.clu) und Vektoren (\*.vec). Netzwerke speichern die Beziehungen zwischen den Knoten (Vertices), Partitionen speichern die Zugehörigkeit eines Knotens zu einer Gruppe (kategoriale Variable), Vektoren speichern kontinuierliche Werte für ein Knoten. Alle drei Datentypen können unabhängig voneinander gespeichert und geladen werden. Graphische Darstellungen und Berechnungen sind aber nur möglich, wenn die Informationen kompatibel sind. Neben weiteren Datentypen kennt Pajek außerdem ein Projekt-Format (\*.paj), in dem mehrere Variablen/Netzwerke zusammengefaßt werden.

Daten werden entweder über das File-Menü oder über die entsprechenden Buttons am linken Fensterrand geladen/gespeichert. Anders als in ucinet muß eine Datei nur einmal geöffnet werden. Anschließend können Sie aus den Ausklappmenüs auswählen, auf welche Datei(en) sich ein Befehl beziehen soll.

Die Organisation der Menüs orientiert sich an den Datentypen: Unter Net finden Sie Prozeduren, die sich auf ein einzelnes Netzwerk beziehen, Networks enthält Befehle, die auf zwei und mehr Netzwerke angewendet werden usw. Soweit dies sinnvoll ist, können Sie Datentypen auch ineinander umwandeln (z. B. eine kategoriale Partition als Vektor speichern oder einen Vektor in Kategorien zerlegen und als Partition speichern). Besonders wichtig ist das Draw-Menü. Anders als bei ucinet sind mathematische und graphische Analyse in Pajek in einem Programm zusammengefaßt. Deshalb sind in Pajek typischerweise neben dem Hauptfenster zwei weitere Fenster geöffnet: eine graphische Darstellung des Netzwerkes und ein Ausgabefenster. Wie alle Befehle beziehen sich auch die Kommandos im Draw-Menü stets auf die ausgewählten Dateien.

### 1.2 Zentralität und Zentralisation

Öffnen Sie über den Button oder über das Menü (File - Network - Read) die Datei kurs.net. Zeichnen Sie das Netzwerk mit Draw - Draw (oder strg-g). Ist dieses Netzwerk dicht? Wieviele Verbindungen könnten theoretisch bestehen?

Das Netzwerk ist gerichtet und gewichtet. Für die folgenden Analysen wollen wir die Richtung (und die Gewichte der Linien) zunächst ignorieren. Wählen Sie Net - Transform - Arcs -> Edges - All. Warum werden Sie gefragt, was mit den „multiple lines“ passieren soll? Wählen Sie die Option „Summe“. Zeichnen Sie das Netzwerk erneut (mit strg-g).

Lassen Sie sich mit `Info - Network - General` einige allgemeine Informationen über das Netzwerk zeigen. Welches sind die zentralen Akteure in diesem Netzwerk?

Überprüfen Sie Ihre Vermutung, indem Sie die drei bekannten Maße für Zentralität bestimmen: `Net - Partition - Degree - All` erzeugt eine Partition mit aus dem Degree sowie einen Vektor mit dem normalisierten Degree. `Draw - Partition` zeichnet das Netzwerk neu, jetzt mit verschiedenen Farben für die Degrees. Das ist noch nicht sehr hilfreich. `Draw - Partition - Vector` ist schon besser, aber die Unterschiede sind immer noch nicht gut zu erkennen. Wählen Sie im Grafikfenster `Options - Size - of Vertices - 15`.

Berechnen Sie als nächstes die Closeness Centrality: `Net - Vector - Centrality - Closeness - All`. Warum scheitert die Berechnung?

Finden Sie die im Datensatz enthaltenen Komponente(n): `Net - Components - Weak - 2`. Sie sehen jetzt in den Auswahlmenüs oben das symmetrisierte Netz und darunter eine Partition. Lassen Sie sich mit `Info - Partition` (Voreinstellungen akzeptieren) Näheres zu dieser Partition anzeigen. Wählen Sie jetzt `Operations - Extract from Network - Partition` um die Knoten, die zur Komponente gehören zu extrahieren. Akzeptieren Sie die Voreinstellung (1-\*). Was bedeutet dies?

Im Ergebnis erhalten Sie ein weiteres Netzwerk mit nur noch 12 Knoten. Wiederholen Sie jetzt die Berechnung der Closeness Centrality. Den Grad der Zentralisierung finden Sie im Ausgabefenster. Mit `Info - Vector` können Sie sich die Werte anzeigen lassen. Alternativ zeichnen Sie das Netzwerk mit `Draw - Draw Vector` neu. Mit `Options - Mark Vertices Using - Vector Values` können Sie die Abbildung um die Zentralitätswerte ergänzen.

Bestimmen Sie jetzt bitte die Betweenness Centrality/Centralisation. Warum liegt diese relativ niedrig? Lassen Sie wiederum das Netzwerk zeichnen.

### 1.3 K-Cores und Blockmodelling

Auch hier interessieren wir uns wieder für k-Cores. Wählen Sie `Net - Partitions - Core - All`. Lassen Sie das Netz zeichnen (farbige Markierung der Partition). Wie interpretieren Sie das Ergebnis?

Zum Schluß wollen wir ein einfaches (und sehr restriktives) Zentrum-Peripherie Blockmodel auf diesen Datensatz anwenden. Wählen Sie `Operations - Blockmodeling - Random Start`. Akzeptieren Sie die Vorgaben (strukturell Äquivalenz). Wie interpretieren Sie die Ausgabe? Die Zuordnung zu einem der Cluster (Zentrum oder Peripherie) wird als Partition gespeichert. Sie können sich die Partition wieder visualisieren lassen. Finden Sie die Einteilung plausibel?

## 2 Große Datensätze

Pajek ist vor allem zur Analyse großer Datensätze geeignet. Öffnen Sie das Pajek-Projekt `world_trade`. Dies enthält das Netzwerk von 80 Staaten, zwischen denen Industriegüter gehandelt werden (Stunde 1) sowie zwei Partitionen: den Kontinent und die Stellung im Welt-system. Lassen Sie sich die Einteilung nach Kontinenten visualisieren. Mit Hilfe der Partition können Sie beispielsweise die europäischen Länder aus dem Netzwerk extrahieren: Wählen Sie `Operations - Extract From Network - Partition - 3` (das ist die Ordnungsnummer für Europa). Wie groß ist das extrahierte Netzwerk? Lassen Sie auch dieses Netzwerk zeichnen.

Gehen Sie zurück zu dem ursprünglichen Netzwerk und wählen Sie `Operations - Shrink Network - Partition - 2 - 3`. Lassen Sie auch dieses Netzwerk visualisieren. Was passiert hier?

Die Weltsystem-Partition wurde aufgrund theoretischer Überlegungen festgelegt. Läßt sich eine solche Einteilung auch empirisch reproduzieren? Wählen Sie noch einmal `Operations - Blockmodeling - Random Start`, aber geben Sie drei Cluster vor. Visualisieren Sie das Ergebnis mit der neuen Partition. Vergleichen Sie das Ergebnis mit der theoretisch vorgegebenen Partition.