

# Visualisierung und Analyse der Kennwerte zu den Hochschulpartnerschaften der HTW Berlin

## 1. Aufgabenstellung

Zur Verfügung gestellt wurde uns eine Excel-Datei in der die Partneruniversitäten der HTW Berlin eingetragen sind. Die Partnerschaftsbeziehungen der Universitäten sind nach Fachbereichen und Studiengängen aufgeschlüsselt. Es stehen ebenso diverse Kennwerte wie z.B. die Teilnahme am Erasmus-Programm, gesprochene Sprachen, geschlossene Verträge, sowie In- und Outgoings für jede Partnerschaft zur Verfügung.

Unsere Aufgabe besteht darin diese Beziehungen in einem Tool namens Gephi als Graph zu Visualisieren und anhand der Kennwerte Rückschlüsse auf die Intensität der Partnerschaft ziehen zu können.

## 2. Herangehensweise

Anders als in der Aufgabenstellung vorgegeben haben wir uns zu Aufgabe gemacht alle zur Verfügung stehende Datensätze zu analysieren. Bei der Erstellung des Graphen sind wir schrittweise vorgegangen:

### 2.1. Strukturanalyse

Wie bereits eingangs erwähnt sind die Partnerschaften nach Fachbereichen und Studiengängen aufgeschlüsselt. Wir haben uns entschieden diesen Sachverhalt als eigenständige Knoten darzustellen. Zusätzlich zu diesen Knoten bilden natürlich auch die Universitäten eigene Entitäten die letztendlich über die Studiengänge mit den Fachbereichen über Kanten verbunden sind. Durch diese Aufteilung ist prinzipiell eine Clusterbildung zu erwarten, da im Regelfall jedem Studiengang genau ein Fachbereich zugeordnet werden kann und die Fachbereiche keine direkte Verbindung untereinander haben. Zusammengefasst wird unser geplantes Netz also über drei Typen von Knoten verfügen: **Fachbereiche**, **Studiengänge** und **Universitäten**.

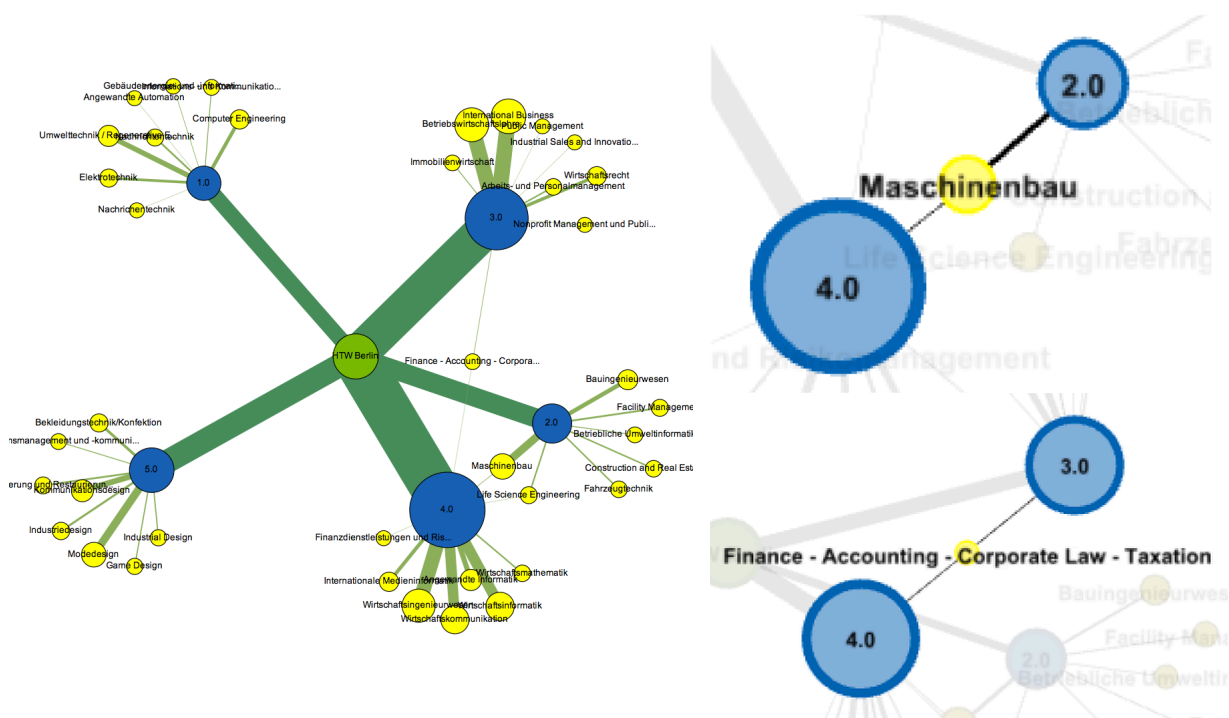
Aufgrund der Menge der Datensätze (478) haben wir uns dazu entschieden die Daten automatisiert in Gephi zu importieren. Das erfolgt über ein Plugin.

### 2.2. Visualisierung Fachbereiche und Studiengänge

Da Aufgrund der Struktur der Daten nicht alle Abhängigkeiten automatisiert in einem Schritt importiert werden können haben wir uns dazu entschlossen den Graphen schrittweise aufzubauen.

Als erster erfolgt der Import der Fachbereiche und Studiengänge, die um unseren Root-Knoten (die HTW-Berlin) angeordnet werden.

Dabei ergibt sich folgender Graph:

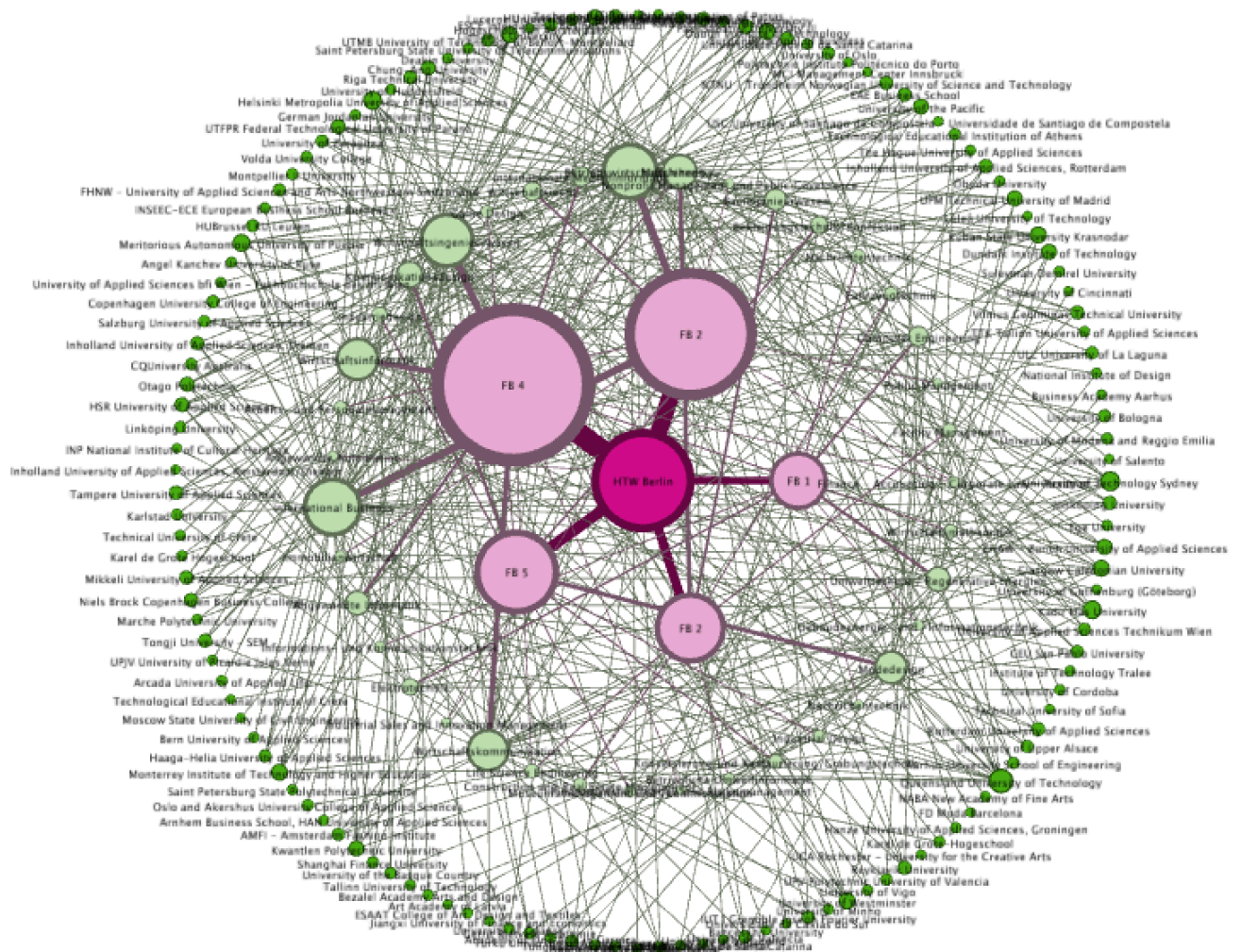


Wie zu erwarten zeigt der Graph eine typische Clusterbildung, da im Normalfall ein Studiengang (gelbe Knoten), immer genau einem Fachbereich (blaue Knoten) zugeordnet werden kann. Hier sind bereits die ersten Unregelmäßigkeiten zu erkennen. Durch den automatischen Import der Daten konnten wir feststellen, dass der Studiengang Maschinenbau und Finance – Accounting – Corporate - Law – Taxation in den Quelldaten zwei verschiedenen Fachbereichen zugeordnet wurde. Diese Fehler haben wir in den Quelldaten anschließend manuell korrigiert. Auf die Art der Gewichtung von Knoten und Kanten gehen wir in Punkt 3 gesondert ein.

### 2.3. Kopplung Partneruniversitäten an Studiengänge

Durch einen weiteren Importvorgang ist es uns nun möglich die entsprechenden Universitätsknoten an die Studiengänge anzuschließen. Die zugehörigen Attribute werden anschließend für die Universitätsknoten (ERASMUS, Sprache, Verträge) und Kanten (In- und Outgoings) nachgetragen. Dieser Vorgang erfolgt semiautomatisch in dem die Excel-Daten auf das Gephi-Datalabformat konvertiert werden. Der dabei entstandene Graph beschreiben wir in Punkt 3 näher.

## 3. Ergebnis



Für eine bessere Sichtbarkeit ist die Quelldatei des entstandenen Graphen auch auf Github (<https://github.com/chillyistkult/Gephi-Testcase>) verfügbar.

## Formale Beschreibung des Netzaufbaus

Die aus 2.2. sichtbaren Cluster sind verschwunden, das liegt in erster Linie an dem gewählten Graph-Layout (nicht etwa weil sich die Struktur der Knoten geändert hätte). Wir haben uns für ein Concentric Layout entschieden, da sich so die Knotenhierarchien am besten darstellen lassen. Im Zentrum des Graphen steht die HTW-Berlin als Root-Knoten (magenta), darum herum ordnen sich die aus 2.2. bekannten Fachbereiche (rosa) und darum die Studiengänge (hellgrün) an. Die in 2.3. beschriebenen Universitätsknoten, bilden den äußersten Ring (dunkelgrün). Die gezeichneten Kanten verbinden die einzelnen Hierarchieebenen miteinander.

Knotentyp	Beschreibung	Anzahl
1	Universitäten	137
2	Studiengänge	39
3	Fachbereiche	4

Unschwer zu erkennen ist ebenfalls die Gewichtung von Knoten und Kanten. Die Universitätsentitäten sind nach der Anzahl an Partnerschaftsbeziehungen zu den jeweiligen Studiengängen gewichtet.

Ein Beispiel zur Verdeutlichung: Die **Queensland University of Technology** unterhält Partnerschaftsbeziehungen in 15 verschiedenen Studiengängen, deshalb ist dieser Knoten mit einer Gewichtung von 15 im unteren rechten Teil des äußersten Rings des Graphen recht deutlich auszumachen. Knoten von Universitäten mit weniger Partnerschaftsbeziehungen werden dementsprechend im Graph auch niedriger gewichtet und daraus resultierend kleiner dargestellt.

Ähnlich verhält es sich mit den Studiengängen. Die Gewichtung ist hier die Summe aller Universitäten die Partnerschaften zu einem entsprechenden Studiengang unterhalten. Selbiges gilt auch für die Fachbereiche. Zusammengefasst lässt sich also feststellen, dass sich die Gewichtungen hin zum Zentrum des Netzes aufaddieren.

Die Kanten sind nach dem selben Prinzip gewichtet. Die Häufigkeit der Partnerschaften mit einem Studiengang wird also sowohl durch die Gewichtung der Knoten als auch Kanten ausgedrückt.

Eine Ausnahme sollen die Kanten zwischen Partneruniversitäten und Studiengängen darstellen. An diesen Kanten sind Out- und Incomings als Attribute vermerkt. Ursprünglich haben wir geplant die Summe dieser Attribute als Gewichtungsgrundlage für diese Kanten zu benutzen. Leider verhindert ein Bug in Gephi genau diese Attribute als Grundlage zu verwenden, daher sind diese Kanten NICHT gewichtet.

Der erstellte Graph bildet ein heterogenes multiplexes Netzwerk ab. Als Sonderfall handelt es sich hier auch um einen multimodalen bipartiten Graph. Die eingangs erwähnten Knoten vom selben Typ (Universitäten, Fachbereiche und Studiengänge), haben untereinander keine Verbindungen.

Die gezeichneten Kanten sind ungerichtet da Aufgrund der Art der Partnerschaft (Hochschulpartnerschaft) stets von einer bilateralen Beziehung auszugehen ist.

#### 4. Analyse

Bei der Analyse des Graphen möchten wir zuallererst auf die direkt sichtbaren Zusammenhänge eingehen. Unschwer zu erkennen ist der Knoten des **Fachbereichs 4**, dieser wird mit einer Gewichtung von 159 sehr groß dargestellt. Daraus lässt sich aufgrund der in 3. beschriebenen Sachverhalte ableiten, dass dieser Fachbereich sehr viele Partnerschaften unterhält (nämlich genau 159).

Wir können hier allerdings nur Aussage über die Anzahl der Partnerschaften nicht aber über deren Intensität treffen. Gleiche Rückschlüsse lassen sich auch auf Studiengangs-Ebene ziehen. Hier stechen die Studiengänge **International Business** und **Betriebswirtschaftslehre** mit 104 bzw. 100 eingetragenen Partnerschaften hervor. Auf Universitätsebene lässt sich – wenn auch nicht so deutlich – ebenso die Anzahl an Partnerschaften direkt an der Größe der Knoten erkennen.

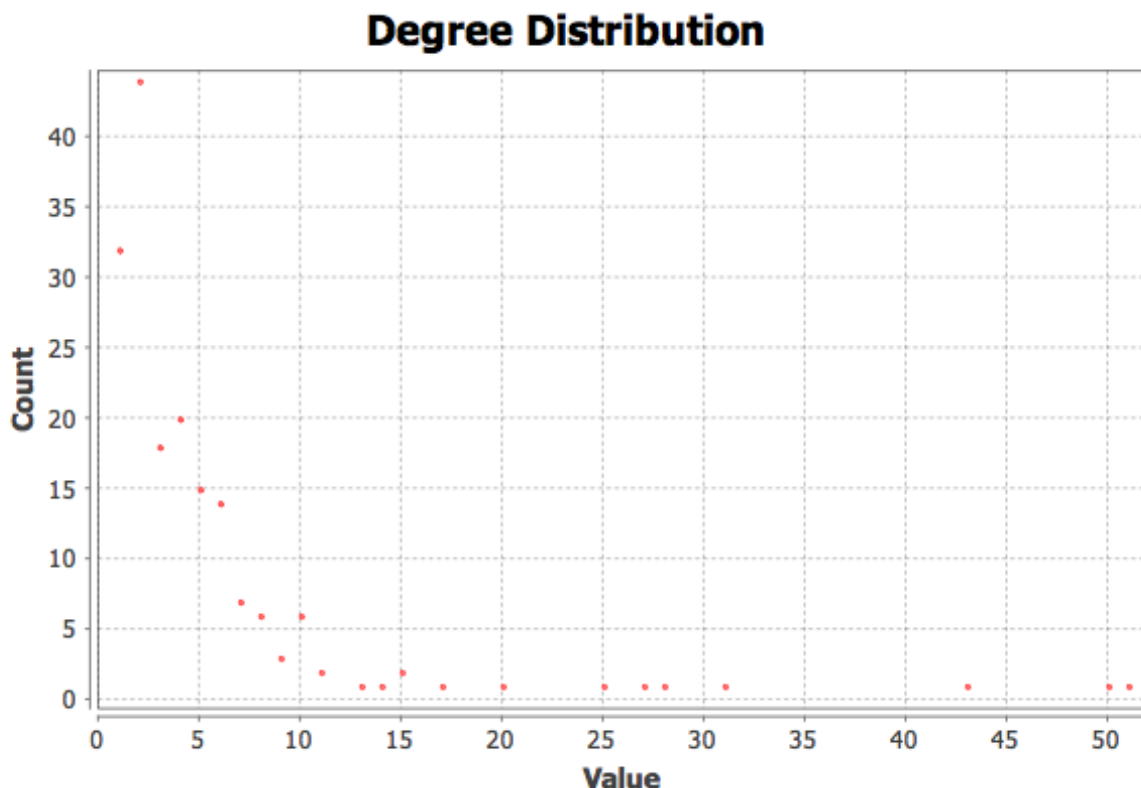
Die Intensität dieser Partnerschaften wollten wir über die In- und Outcomings verdeutlichen. Leider ist diese Eigenschaft zwischen Universitäten und Studiengänge nicht visuell im Graph zu erkennen, da aufgrund eines Bugs in Gephi der Gewichtungsalgorithmus der Kanten für diese Attribute nicht ordnungsgemäß funktioniert.

**Hypothese: Die im Graphen hoch gewichteten Knoten deuten auf intensive Partnerschaften mit der HTW Berlin hin.**

Die Dichte des Netzes beträgt 0,011. Dieser niedrige Wert ist zu erwarten, da die einzelnen Knotentypen:

1. nicht untereinander vernetzt sind
2. jeder Universitätsentität nur über eine (sehr) begrenzte Anzahl von Partnerschaften verfügt

Da die Dichte alleine keine Aussage über die Zentralität einzelner Knoten zulässt haben wir dies im Folgenden gesondert analysiert.



Anhand der Gradverteilung lässt sich eine Aussage über die Zentralität unseres Netzes treffen. Klar zu erkennen ist, dass einige wenige Knoten einen sehr hohen Grad aufweisen, was auf ein (eher hoch) zentralisiertes Netzwerk hindeutet. Das lässt aber aufgrund der Art unseres Netzes (bipartit, multimodal) immer noch keine direkten Rückschlüsse auf die Intensität der einzelnen Partnerschaften zu, unterstützt aber bereits unsere Hypothese.



Daraus folgend haben wir versucht die In- und Outcoming zu analysieren. Hohe In- und Outcoming zwischen Universitäten deuten stark auf eine intensive Partnerschaft hin. Genau diese Aussage brauchen wir für die Validierung unserer Hypothese. Da uns die visuelle Darstellung dieses Sachverhaltes wie erwähnt leider nicht gelungen ist gehen wir hier den Weg über Statistiken.

### Top 5 Incomings

Universität	Studiengang	Incoming
Dublin Institute of Technology	International Buisness	25
Tongji University – CDHAW	Fahrzeugtechnik	22
Dublin Institute of Technology	BWL	20
Glasgow Caledonian University	International Buisness	18
University of Technology Sydney	International Buisness	15

### Top 5 Outgoings

Universität	Studiengang	Outgoings
Helsinki Metropolia University of Applied Sciences	Construction and Real Estate Managment	95
University of Westminster	BWL	19
University of Technology Sydney	BWL	18
Tongji University – CSHAW	Fahrzeugtechnik	16
University of Queensland	BWL	16

Was hier sofort ins Auge fällt sind unsere Eingangs erwähnten Studiengänge (**International Business** und **Betriebswirtschaftslehre**) mit hoher Gewichtung. Diese sind in beiden Rankings häufig vertreten, was letztendlich auch unsere Hypothese bestätigt, dass die Häufigkeit der Partnerschaften in einem Studiengang auch Rückschlüsse auf deren Intensität zulässt.

**Je häufiger also eine Partnerschaft in einem Studiengang auftritt, desto größer sind wahrscheinlich auch deren Intensitäten.**

Natürlich lässt sich die Intensität noch anhand einiger anderer Kennwerte beurteilen sodass wir hier letztendlich keine wissenschaftlich belastbare Aussage treffen aber sehr wohl eine Tendenz feststellen können.

## 5. Fazit

Auch wenn einige offensichtliche Sachverhalte und auch Fehler in den Daten direkt im Graphen sichtbar gemacht werden konnten, ist es fraglich ob der gewählte Aufbau des Graphen für die wissenschaftliche Untersuchung des Kontextes der Aufgabe geeignet ist, da die verschiedenen Knotentypen die korrekte Interpretation verschiedener Kennwerte erschwert. Außerdem ist der Graph nur schwer erweiterbar. Angenommen man möchte das egozentrische Netzwerk der Grades 1,5 einer der Partneruniversitäten ergänzen so müsste man auch für diese den Weg über extra Fachbereichs- und Studiengangsknoten gehen um eine gewisse Konsistenz im Aufbau zu wahren, was letztendlich zu Redundanzen führt und schnell sehr komplex werden kann.

Wir haben gelernt, dass die Auswertung und Visualisierung einer so großen Datenmenge nicht trivial ist und auch auf Grundlage intensiver Vorüberlegungen Fehler in der Darstellung nicht ausgeschlossen werden können. Uns viel es schwer eigene Hypothesen zu entwickeln und anschließend zu validieren, da wir stets das Gefühl hatten, das die gegebene Excel-Tabelle als Grundlage der Analyse besser geeignet gewesen wäre, als das daraus visualisierte Netz. Hier spielt sicherlich auch eine Rolle das Gephi für die automatisierte Verarbeitung solcher strukturierten Daten nicht unbedingt geeignet ist, da viele Limitationen durch Plugins und Oberfläche eine detaillierte Analyse erschweren.