



DOKUMENTATION

Soziale Netzwerkanalyse

NHL Spielerverbindungen

FHNW – Modul SNA
Joël Winter, Manuel Riedi, Janick Hürzeler

1 Anpassungen

Für einen möglichst einfachen Import in Gephi, möglichst ohne manuelle Anpassungen, wurden noch einige Spalten im KNIME-Workflow umbenannt (z.B. Source und Target bei der Edge-Table) oder ergänzt (neue Spalte «Label» mit Namen der Spieler). Zudem wurden einige Spalten noch weggefiltert, die wir für die Analyse als nicht benötigt eingestuft hatten, jedoch beim «Clean up» nicht alle Spalten mit diesen Informationen entfernt hatten.

Da eine Analyse nicht auf dem gesamten Netzwerk aufbaut, sondern gefiltert nach Team und Zeitraum (mehrere Saisons), wurde der Edge-Table Teil des KNIME-Workflows um eine weitere, gefilterte Ausgabe als CSV ergänzt. Dies war nötig, da die Kanten in Gephi in gewichtete Kanten aufsummiert werden, wir aber die einzelnen Kanten zuvor noch einschränken wollten.

2 Soziale Netzwerkanalyse

2.1 Gegenüberstellung von «guten» und «schlechten» Teams

Bei dieser Analyse sollen eher bessere und eher schlechtere Teams aus einem gewissen Zeitraum gegenübergestellt werden. Dabei wurde anhand von Zentralitätswerten untersucht, ob treue Spieler einen Einfluss auf den Erfolg eines Teams haben. Ob eher gleich bleibende Kader ohne viele Spielerwechsel zum Erfolg beitragen können, wurde über die Dichte des Netzwerks versucht herauszufinden.

Als Zeitraum wurden die Saisons von 2010 bis 2015 gewählt, da dort die Auswahl der guten Teams einfach gefallen ist. Ohne Filter auf Saisons hätten kaum Aussagen gemacht werden können, da die NHL so organisiert ist, dass die Teams möglichst ausgeglichen sind («Salary Cap», «Entry Draft») und es über eine längere Zeit keine dominierenden Teams gibt.

2.1.1 Datenaufbereitung

Bei der Datenbeschaffung wurden die Node- und Edge-Table so gefüllt, dass sie sämtliche Daten repräsentieren. Beim Import in Gephi wurden dann die Kanten automatisch aufsummiert und gewichtet. Möchte man die Kanten nicht weiter auf Attribute filtern, so ist dieses Verhalten erwünscht. Da für diese Analyse jedoch die Teams und Saisons einzeln betrachtet und somit gefiltert werden müssen, wurde der KNIME-Workflow noch um solche Filter ergänzt, der dann spezifische Edge-Table-Dateien exportiert. Diese können dann wiederum in den entsprechenden Workspace in Gephi geladen werden.

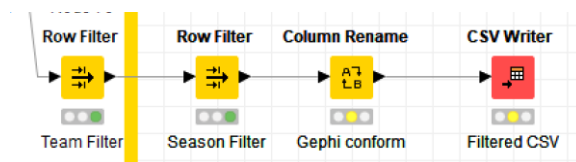


Abbildung 1: Zusätzliche Nodes im KNIME-Workspace für gefilterte Edge-Tabellen

2.1.2 «Gute» Teams

Die Chicago Blackhawks wurden als «gutes» Team ausgewählt, da sie im betrachteten Zeitraum den Stanley Cup (Trophäe für den Sieger der NHL-Playoffs) drei Mal gewonnen haben. Den Los Angeles Kings ist dies zwei Mal gelungen.

2.1.2.1 Chicago Blackhawks

Nodes: 102 (= Anzahl verschiedene Spieler)
Edges: 2'457

Graph Density: 0.477

Kanten vorgefiltert in KNIME auf team_id (16) und seasonYear (2010 – 2015), Topology-Filter für «Degree Range» ab 1, um nur Knoten mit entsprechenden Kanten darzustellen.

Bei der Betrachtung des kompletten Graphen, wobei die Knotengrösse hier der Betweenness-Zentralität entspricht, wird ersichtlich, dass hier mehrere Spieler eine zentrale Rolle tragen. Durch die Darstellung der Kanten nach Gewicht wird zudem noch ein «Ring» von Spielern ersichtlich, der sich um die grossen Knoten in der Mitte ansammelt.

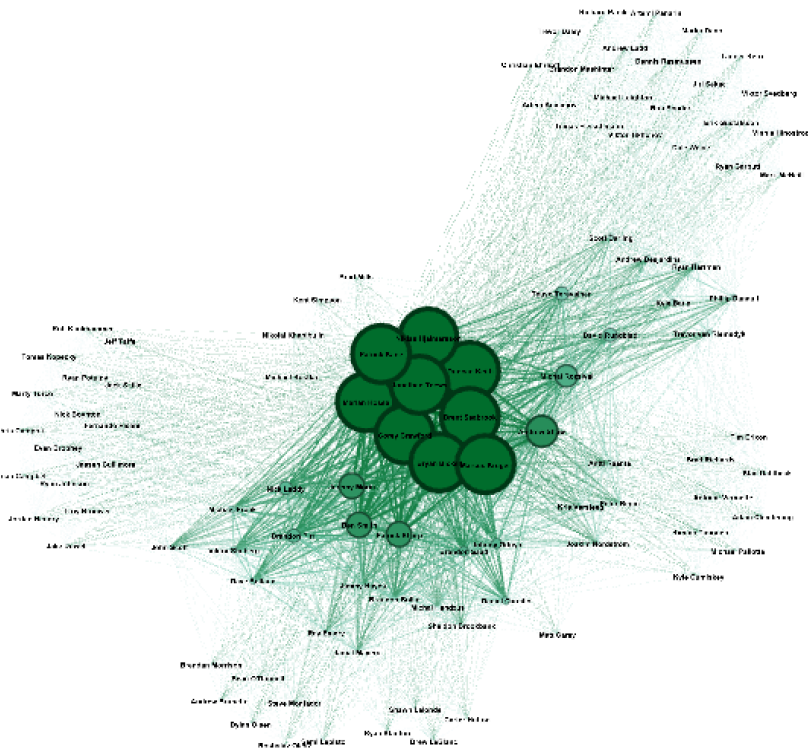


Abbildung 2: Gesamtes Spieler-Netzwerk der Chicago Blackhawks für 2010 – 2015

Fokussiert man sich nun auf die grösseren Knoten, und zieht auch deren Daten in Betracht, so wird schnell ersichtlich, dass es sich dabei um langjährige und vor allem gute Spieler von den Chicago Blackhawks handelt. Diese neun Spieler haben im gesamten Zeitraum von 2010 bis 2015 bei den Blackhawks zusammengespielt und sind sehr wichtige Spieler innerhalb der Organisation. So ist Jonathan Toews seit 2008 ununterbrochen der Captain dieses Teams. Zusammen mit Duncan Keith und Patrick Kane hat er auch sehr viele Skorerpunkte auf dem Konto, wie zum Beispiel in den Playoffs 2015, wo diese drei Spieler die Rangierungen 2, 4 und 5 der besten Playoffs-Skorer belegten.

2.1.2.2 Los Angeles Kings

Nodes: 69 (= Anzahl verschiedene Spieler)

Edges: 1'435

Graph Density: 0.612

Kanten vorgefiltert in KNIME auf team_id (26) und seasonYear (2010 – 2015), Topology-Filter für «Degree Range» ab 1, um nur Knoten mit entsprechenden Kanten darzustellen.

Bei den Los Angeles Kings fällt auf, dass sie eine sehr kleine Anzahl an verschiedenen Spielern hatten (nur 69). Somit haben sie über einen Kader verfügt, der nur sehr wenige Wechsel hatte. Im Graphen (Auszug in Abbildung 3) wird dann auch ersichtlich, dass das Netzwerk trotz gesamthaft kleiner Knotenanzahl über eine Verbindung von Spielern verfügt, die das Zentrum bilden. Diese sind nicht mehr so ausgeprägt, wie beim vorherigen Team, da

es eben wenig Anzahl Knoten sind und es auch um diese zentralen Spieler nochmals eine Reihe von Spielern mit vielen und auch hoch gewichteten Kanten gibt.

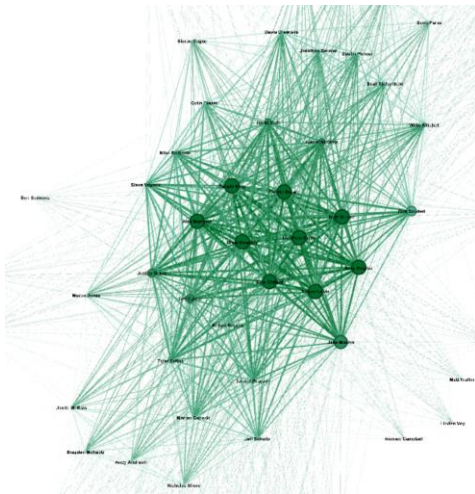


Abbildung 3: Gesamtes Spieler-Netzwerk der Los Angeles Kings für 2010 - 2015

2.1.3 «Schlechte» Teams

Für das schlechte Team wurden jeweils die Plätze in der Abschlusstabelle betrachtet, da die schlechten Teams sich gar nicht für die Playoffs qualifiziert haben. Dabei waren die Edmonton Oilers sehr oft im hintersten Teil der Tabelle zu finden, wie auch die Toronto Maple Leafs. Mehr Informationen zu dieser Auswahl im Anhang (Kapitel 4.1).

2.1.3.1 Edmonton Oilers

Nodes: 116 (= Anzahl verschiedene Spieler)

Edges: 3'049

Graph Density: 0.457

Kanten vorgefiltert in KNIME auf team_id (22) und seasonYear (2010 – 2015), Topology-Filter für «Degree Range» ab 1, um nur Knoten mit entsprechenden Kanten darzustellen.

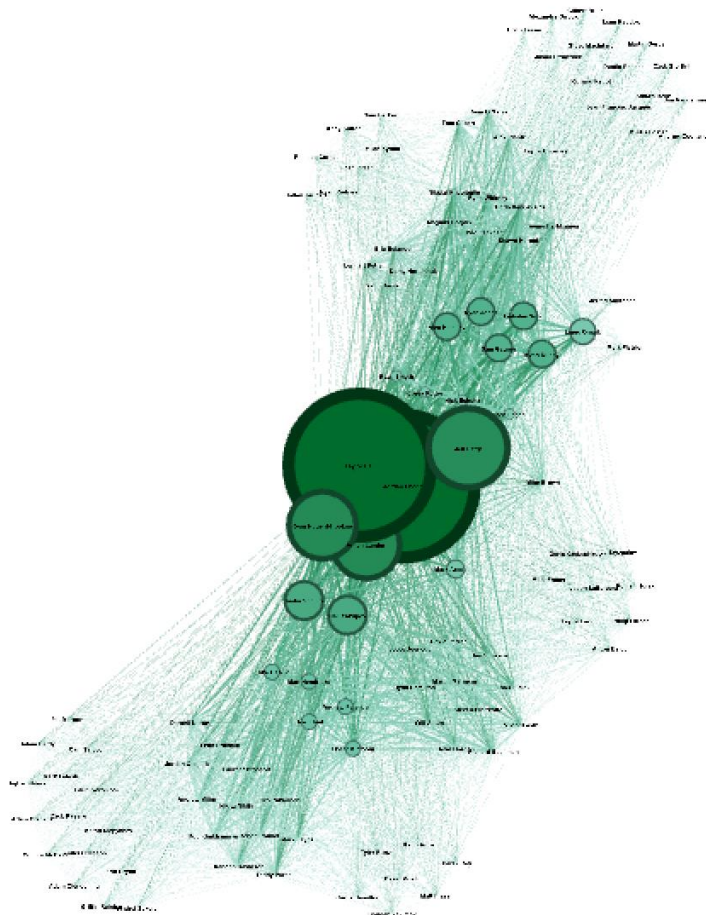


Abbildung 4: Gesamtes Spieler-Netzwerk der Edmonton Oilers für 2010 – 2015

Das Netzwerk der Edmonton Oilers ist sehr zerstreut. Es gibt lediglich zwei Spieler, die über einen sehr hohen Betweenness-Zentralitätswert verfügen. Namentlich sind dies die Spieler Taylor Hall und Jordan Eberle, die während dem gesamten Zeitraum für Edmonton gespielt haben. Denen gegenüber stehen aber sehr viele Spieler, die nur eine Saison in dieser Organisation gespielt haben. Diese konnten über die selbst berechnete Spalte «Avg Weight per Degree» sehr schnell ausfindig gemacht werden (Wert = 1). Mit eine Filter auf diese Spalte wurde ersichtlich, dass es sich dabei um 54 Spieler und somit fast die Hälfte handelt (siehe Abbildung 5).

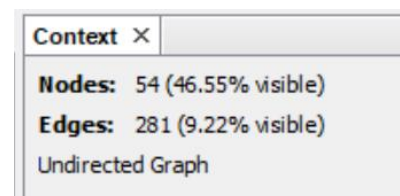


Abbildung 5: Resultat nach Filter mit Wert "1" auf Spalte "Avg Weight per Degree"

2.1.3.2 Toronto Maple Leafs (10)

Nodes: 119 (= Anzahl verschiedene Spieler)

Edges: 3'069

Graph Density: 0.437

Kanten vorgefiltert in KNIME auf team_id (10) und seasonYear (2010 – 2015), Topology-Filter für «Degree Range» ab 1, um nur Knoten mit entsprechenden Kanten darzustellen.

Betrachtet man den Graphen (Abbildung 6) stechen auf den ersten Blick die grossen Knoten im Zentrum heraus. Hierbei handelt es sich um fünf Spieler, die über den gesamten Zeitraum zusammengespelt haben. Es ist aber auch deutlich zu sehen, dass es sehr viele

Knoten gibt, die sehr klein (und kaum ersichtlich) sind. Dies sind Spieler, die nur eine Saison in Toronto gespielt haben.

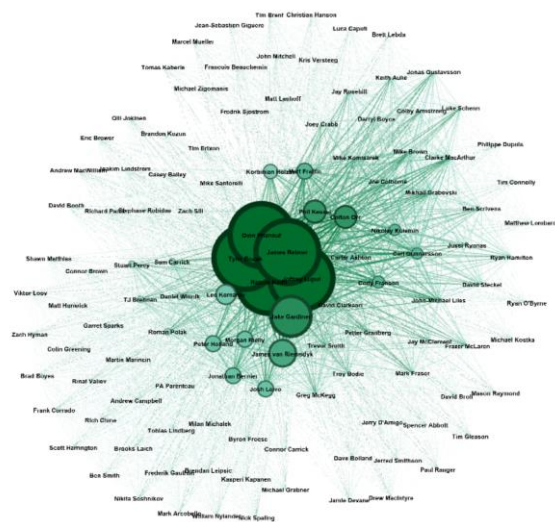


Abbildung 6: Gesamtes Spieler-Netzwerk der Toronto Maple Leafs für 2010 - 2015

Diese Fluktuation wird besonders deutlich, wenn die selbst berechnete Spalte «Avg Weight per Degree» als Filter angewendet wird. Filtert man so auf die Spieler, die nur eine Saison zusammengespielt haben, sind dies fast 58%. Zudem besitzen diese Knoten dann auch fast 20% der Kanten (Abbildung 7).

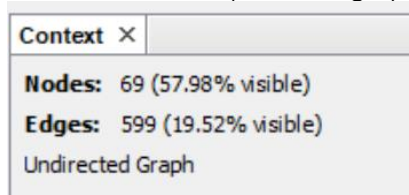


Abbildung 7: Resultat nach Filter mit Wert "1" auf Spalte "Avg Weight per Degree"

2.1.4 Schlussfolgerung

Die Analysen zeigen auf, dass ein eher gleichbleibendes Team grössere Chancen auf Erfolg hat. So haben die beiden «guten» Teams eine weniger hohe Anzahl an verschiedenen Mitspielern. Bei den «guten» Teams wurden jedoch zwei verschiedene Ansätze ersichtlich. So gibt es bei den Chicago Blackhawks einen Kern von Weltklassem Spielern, die die Basis für die Erfolge bildeten. Bei den Los Angeles Kings ist ein so talentierter Kern nicht vorhanden, jedoch ist dort das Netzwerk ungemein dicht (höchste Graph Density) mit einer sehr geringer Fluktuation.

Bei den «schlechten» Teams gibt es zwar auch Spieler, die während dem gesamten Zeitraum zusammengespielt haben. Bei den Edmonton Oilers sind dies jedoch lediglich zwei und bei den Toronto Maple Leafs fehlt diesem Kern einfach die spielerische Klasse. Zudem steht diesem Kern noch eine grosser Anteil an Spielern gegenüber, die nur während einer Saison zusammengespielt haben.

2.2 Analyse von Cliques

Bei dieser Analyse wurde untersucht, ob sich verschiedene Cliques gebildet haben. Für den k-Core wurde die Grenzwerte für k ermittelt. In den gefundenen Cliques wurde die Betweenness-Centrality für eher treue und untreue Spieler im Detail analysiert.

Es wurden alle Knoten und Kanten über den gesamten Zeitraum genommen.

Als minimales k damit alle Spieler eine Clique zugewiesen werden können, wurde 27 ermittelt. Dies kommt zustande, da alle Teams mindestens 28 Spieler pro Saison hatten.

Bei der Suche nach dem maximalen k wurden mehrere Cliques mit einem k-Core von 81 gefunden.

Die Auswertung mit einem k-Core von 81 ergab 1074 Knoten, was 36.68% aller Spieler entspricht und 68397 Kanten, was 42.12% der Verbindungen entspricht. Da dieser Graph immer noch sehr viele Knoten beinhaltet, konnte die einzelnen Cliques mit keinem Layout gut dargestellt werden.

Bei der Analyse in der 81-Core Clique nach treuen und untreuen Spieler, fanden wir heraus, dass Lee Stempniak den höchsten Betweenness Centrality Wert hatte. Dies kommt zustande, da Lee in dieser Zeit bei 11 verschiedenen Mannschaften angestellt war. Patrick Kaleta ein Spieler, der in dieser Zeit nur einmal das Team innerhalb der NHL gewechselt hat, hat in dieser Auswertung einen tieferen Wert.

Als drittes wurde noch ein k-Core analysiert, welcher mindestens 50% aller Spieler beinhaltet. Erzielt wurde dies mit einem k von 67. Der Graph beinhaltet dabei 1488 von 2928 Knoten. Auch hier erzielte Lee Stempniak den höchsten Betweenness Centrality Wert. Bei Spielern mit einem tiefen Betweenness Centrality Wert konnten wir nichts über treu oder untreu. Zum Beispiel fanden wir treue Spieler wie Mike Duco, oder untreue Spieler wie Eric Selleck.

Entgegen unserer Erwartung konnten wir keine treuen Spieler finden, welche eine gewichtete Rolle in einer Clique einnehmen. Stattdessen stellte sich heraus, dass grosse k-Core Cliques vor allem durch Spieler gebildet werden, welche häufig die Mannschaft wechselten.

Chicago Blackhawks

Um die Auswertungen der k-Core zu verifizieren, wurde eine Auswertung über weniger Daten gemacht. Es wurden dafür die Daten von Chicago Blackhawks von 2010 – 2015 genommen. Als Gephistartpunkt, wurden die gleichen Daten wie im Kapitel **2.1.2.1 Chicago Blackhawks** verwendet. Von den 102 Spieler in diesem Zeitraum konnten alle einem k-Core mit $k=31$ zugewiesen werden. Dies kann dadurch bestätigt werden, dass in jeder Saison das Team mindestens 32 Spieler hatte. Als maximales k konnte 39 bestimmt werden. Diese Clique enthält dabei genau 40 Spieler.

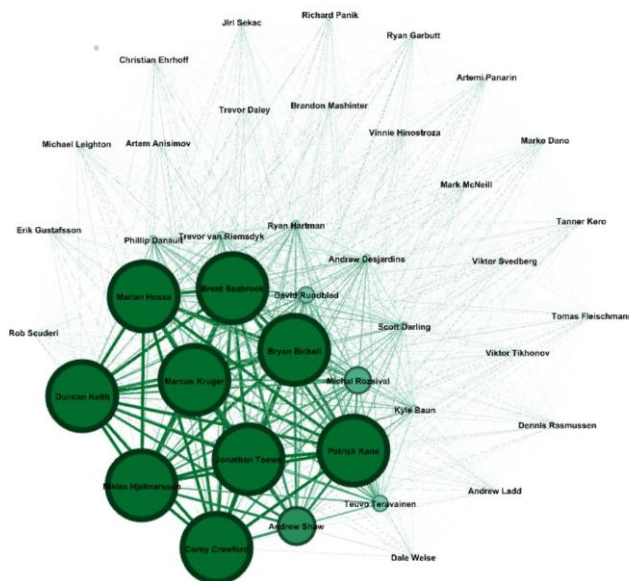


Abbildung 8 Chicago k-Core

2.3 Vertraute Mitspieler

Bei dieser Untersuchung sollen die Spieler identifiziert werden, welche auf eine grosse bzw. kleine Anzahl vertrauter Mitspieler setzen können.

Um eine Aussage diesbezüglich machen zu können, werden zwei folgende Überlegungen über die Spieler-Knoten gemacht:

- Spieler mit hoher "Degree Centrality" und verhältnismässig niedriger "Weighted-Degree Centrality", wechseln oft das Team und sollten daher eher weniger vertraute Spieler aufweisen.
- Spieler mit niedriger "Degree Centrality" und verhältnismässig hoher "Weighted-Degree Centrality", bleiben ihrem Team eher treu und sollten daher auch eher eine hohe Anzahl vertrauter Spieler aufweisen.

Um eine erste Analyse über das gesamte Netzwerk machen zu können, ist also besonders die Differenz zwischen "Degree Centrality" und "Weighted Degree Centrality" als aussagekräftiger Wert zu betrachten. Aus diesem Grund fügen wir dem Node-Table eine zusätzliche Column zu, welche jeweils diese Differenz je Spieler berechnet. Als Ergebnis erhalten wir den Graphen aus (Abbildung 10).

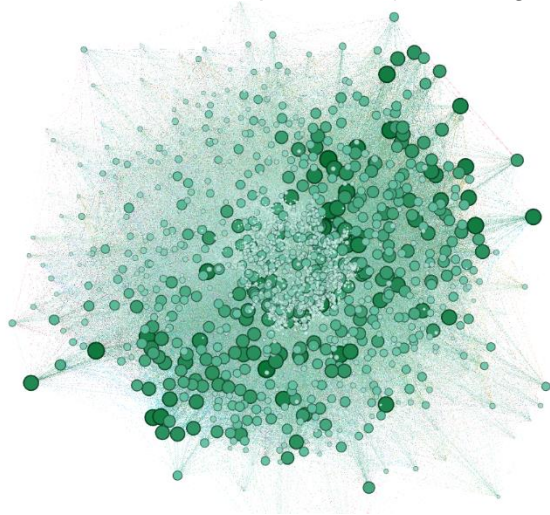


Abbildung 9 Spieler mit hoher Anzahl vertrauter Mitspieler

Je grösser und dunkler ein Node dargestellt wird, je grösser ist die Differenz zwischen Degree und Weighted-Degree. Die Grafik hebt also bereits heraus, welche Spieler auf eine besonders grosse Anzahl treue Mitspieler deuten und welche eher nicht.

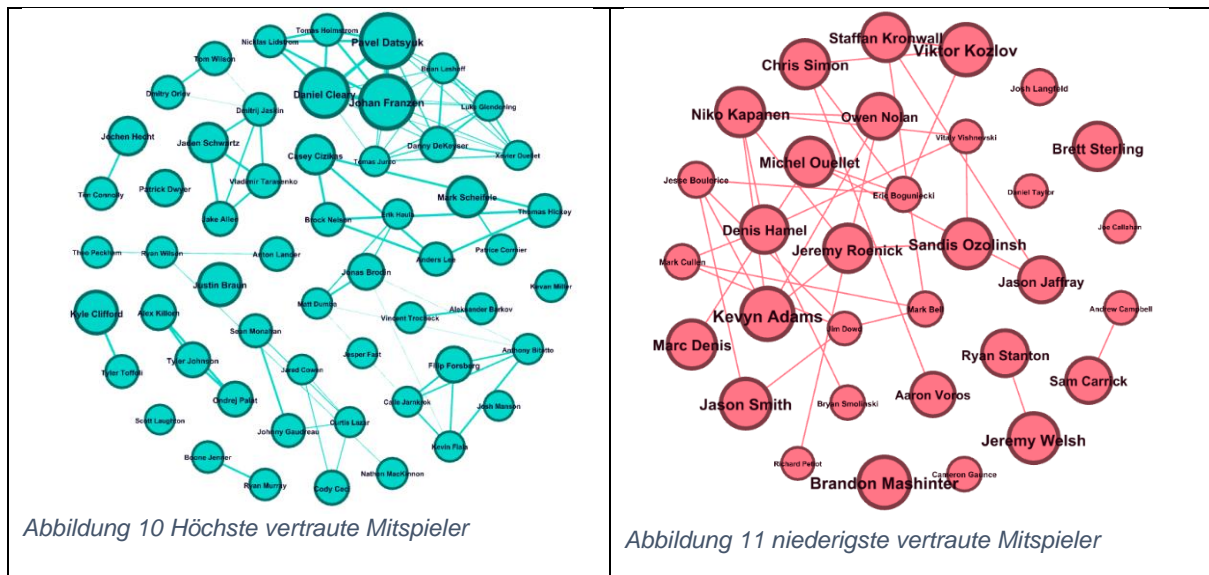
Um den vielen Daten gegen zu wirken und die Spieler, mit den am vertrautesten bzw. wenig vertrautesten Spieler besser visualisieren zu können, werden in einem weiteren Schritt folgend beschriebene Filter angewendet. Filter «Einschränkung hohe Anzahl vertraute Mitspieler»: Degree ≤ 111 (Die durchschnittliche Degree über das ganze Netzwerk ist 111) und Weighted Degree ≥ 195 . Als Ergebnis erhalten wir ein kleines Netzwerk, das gerade noch 55 Spieler enthält.

Mit gegenteilig Filtern machen wir dasselbe erneut in einem zweiten, separaten Netzwerk. Filter «Einschränkung geringe Anzahl vertraute Mitspieler»: Degree ≥ 111 und Weighted Degree ≤ 140 . Hier resultiert ein Netzwerk von gerade noch 32 Spielern.

Um in den beiden Netzwerken noch deutlicher die Unterschiede hervorzuheben, passen wir zudem die Node-Grösse so an, dass Spieler mit höherer Degree-/Weighted-Differenz noch grösser dargestellt werden.

Als Ergebnis erhalten wir die beiden folgend abgebildeten Graphen.

Spieler mit den am höchsten vertrauten Mitspielern	Spieler mit den am wenigsten vertrauten Mitspielern
--	---



Wie in der grünen Grafik (Abbildung 11) zu sehen, stechen Spieler wie Pavel Datsyuk oder Johan Franzén heraus. Vergleichen wir dessen Ergebnisse mit den Spielerdaten auf der offiziellen NHL-Webseite (Abbildung 13) erkennt man sofort, dass diese Spieler nicht einmal (zwischen 2015 und 2018) ihr Team gewechselt haben. Die Begründung liegt daher nah, dass sie demzufolge auch eine hohe Anzahl vertrauter Spieler aufweisen.

Johan Franzen Karrierestatistik

Saison	Team	Kyle Clifford Karrierestatistik			
Saison	Team	GP	G	A	Pts
2015-2016	DET				
2014-2015	DET				
2013-2014	DET				
2012-2013	DET				
2011-2012	DET				
2010-2011	DET				
2009-2010	DET				
2008-2009	DET				
2007-2008	DET				
2006-2007	DET				
2005-2006	DET				

Saison	Team	GP	G	A	Pts
2019-2020	LAK	45	1	1	2
2018-2019	LAK	72	1	1	2
2017-2018	LAK	50	1	1	2
2016-2017	LAK	73	1	1	2
2015-2016	LAK	56	1	1	2
2014-2015	LAK	80	1	1	2
2013-2014	LAK	71	1	1	2
2012-2013	LAK	48	1	1	2
2011-2012	LAK	81	1	1	2
2010-2011	LAK	76	1	1	2

Abbildung 12 Ausschnitt der offiziellen NHL-Webseite über die Spieler Johan Franzen und Kyle Clifford

In der roten Grafik (Abbildung 12) suchen wir nach den eher kleineren Knoten. Als Beispiel sei hier Mark Cullen oder Daniel Taylor genannt und vergleichen sie wieder mit den Daten der NHL-Seite (Abbildung 14).

Daniel Taylor Karrierestatistik

Saison	Team
2007-2008	LAK
2012-2013	CGY
2017-2018	OTT

Mark Cullen Karrierestatistik

Saison	Team	GP	G
2005-2006	CHI	29	0
2006-2007	PHI	3	0
2011-2012	FLA	6	0

Abbildung 13 Ausschnitt der offiziellen NHL-Webseite über die Spieler Daniel Taylor und Mark Cullen

Ab diesem Punkt wird klar, dass die Daten einen signifikanten Qualitätsmangel aufweisen, da ein wichtiges Spieler-Attribut fehlt. Nämlich die Information über die Anzahl gespielte Saisons über die Jahre 2005 bis 2018. Schliesslich ist es nicht ganz unwesentlich, wie viele Saisons gespielt wurden, wenn die Anzahl gewechselte Teams als Faktor für die Bestimmung an Anzahl vertrauten Mitspielern einfließt. Nichtsdestotrotz erkennt man anhand der Vergleichsdaten der NHL-Webseite, dass diese Spieler nach jeder Saison das Team gewechselt haben. Entgegen der roten Grafik dürfte also auch hier die Aussage zutreffen, dass diese Spieler, durch ihren häufigen Teamwechsel, über wenig vertraute Mitspieler verfügen.

3 Ausblick

Die Datenquelle würden wir als passend einstufen. Mit dem direkten Bezug von der NHL gehen wir von einer sauberen Quelle aus. Uns haben von Seiten der API keine Daten gefehlt, wir hätten darauf noch viele weitere interessante API-Requests machen können.

Nützlich wäre als zusätzlicher Parameter die Anzahl gespielte Seasons je Spieler gewesen. Damit hätte man besonders in der Analyse "Vertraute Mitspieler" eine aussagekräftigere Statistik machen können.

Aufgrund der vielfältigen NHL-API hätte man zusätzlich zu den gespielten Seasons auch die gespielten Spiele abfragen können und dann als Kantenbeziehung «Spieler hat ein Spiel zusammen gespielt mit» erstellen können. Dies hätte noch genauere Aussagen geliefert, wie "eng" die Spieler untereinander verknüpft sind, da z.B. Ausfälle wegen Verletzungen oder Wechsel während der laufenden Saison berücksichtigt würden.

Bei der NHL werden auch sehr viele Statistiken zu den einzelnen Spielern erfasst und auch via API zur Verfügung gestellt. Diese Statistiken hätte man einbeziehen können, um Spieler oder auch Teams automatisch als eher «gut» oder «schlecht» einzustufen, was vor allem in der Analyse der Gegenüberstellung solcher Teams nützlich gewesen wäre. Eine Unterteilung in «gut» und «schlecht» lässt sich jedoch nicht nur auf einzelnen Statistikwerten, wie z.B. der Anzahl Tore und Assists, aufbauen. Hier hätte man noch einen Schritt weiter gehen und die immer populärer werdenden «[Advanced Stats](#)» einbeziehen können. Für solche hätte aber wohl eine weitere Schnittstelle hinzugezogen werden müssen.

4 Anhang

4.1 Excel-Dokument für Kategorisierung «gute»/«schlechte» Teams

Rangierungen und Berechnungen für Unterteilung in «gut» und «schlecht»:

[NHL Teams Rangierungen.xlsx](#)

4.2 Werte in Tabellenform für Gegenüberstellung Teams

Chicago Blackhawks:

	Min	Max	Avg
Weighted Degree	31	203	69.784
Degree	31	101	48.176
Closeness Centrality	0.591	1	
Betweenness Centrality	0	192.702	

Los Angeles Kings:

	Min	Max	Avg
Weighted Degree	26	171	73.449
Degree	26	68	41.594
Closeness Centrality	0.618	1	
Betweenness Centrality	0	57.545	

Edmonton Oilers:

	Min	Max	Avg
Weighted Degree	30	221	73.017
Degree	30	115	52.569
Closeness Centrality	0.575	1	
Betweenness Centrality	0	503.222	

Toronto Maple Leafs:

	Min	Max	Avg
Weighted Degree	29	215	67.866
Degree	29	118	51.58
Closeness Centrality	0.570	1	
Betweenness Centrality	0	376.037	

4.3 KNIME Workflow

Der angepasste KNIME-Workflow ist hier zu finden:

[SNA_Projekt.knwf](#)