Projekttitel / Projekt Kurzbeschrieb:

NHL Spielerverbindungen

Teammitglieder

In meinem Team befinden sich die folgenden Team-Mitglieder (min. 2, max. 3):

* Riedi Manuel
* Hürzeler Janick
* Winter Joël

Tools

Gephi

Python

KNIME

# Datenquelle

Die Daten erhalten wir direkt von der NHL über ihre Stats-API (<https://statsapi.web.nhl.com/api/v1>, Beispiel-Abfrage: <https://statsapi.web.nhl.com/api/v1/teams/1>).

Datenzugriff mittels API.

# Datenmodellierung

Knoten

Die Knoten im Netzwerk bilden die einzelnen Spieler. Die Kanten haben die Bedeutung «hat zusammen gespielt mit» und können mit verschiedenen Werten, wie z.B. Anzahl Spiele, Anzahl Saisons oder Anzahl verschiedene gemeinsame Teams gewichtet werden. Es handelt sich somit um ein One-Mode-Netzwerk.

Netzwerk-Grösse

Anzahl betrachtete Saisons: 15  
Anzahl NHL-Teams: ~30 (Aktuell 31)  
Anzahl Spieler im aktiven NHL-Kader: 23 (hier zählen wir noch einen geschätzten Wert von 7 hinzu, da es im Kader auch während der Saison Änderungen gibt wie z.B. Trades oder AHL/NHL-Reassignemnts, das Maximum pro Saison wäre 50 Spieler)  30

15 \* 30 \* 30 = 13'500, das würde aber bedeuten, dass nach jeder Saison der komplette Kader ausgetauscht wird.

Nehmen wir nun an, ein Spieler spielt durchschnittlich 5 Saisons in der NHL, so teilen wir diese Zahl von 15 Saison noch durch 3:

13'500 / 3 = 4’500

Pro Saison gehen wir von (30 – 1) Mitspieler aus. Wenn ein Spieler durchschnittlich 2 Jahre (von den 5) im selben Verein spielt und man davon ausgeht, dass pro Saison ¼ des Kaders ausgetauscht wird, so kann man folgende Formel aufstellen:

Anz. Mitspieler für 1 Spieler: (30 – 1) + (30 – 1)\*0.25 + (30 – 1) + (30 – 1)\*0.25 + (30 - 1) = 101.5

Wenn wir dies nun auf alle Spieler (= unsere Anzahl Knoten) hochrechen, kommen wir auf 101.5 \* 4'500 = 456'750. Diesen Wert kann man sicher durch 2 Teilen, da wir ungerichtete Kanten haben  456'750 / 2 = 228'375. Nicht berücksichtigt sind nun die Mitspieler, mit denen mehrmals zusammen gespielt wird, was dem entspricht, was wir hier herausfinden wollen. Deshalb runden wir einfach mal ab auf 220'000 Kanten.

Anzahl Knoten: 4’500 Anzahl Kanten: 220’000

Attribute

* Knoten
  + Name
    - Evtl. Namensänderungen, aber die API liefert eine ID zurück
  + Position
    - Schwierigkeit: Spieler mit Positionswechsel  aktuellste Position nehmen
  + Alter
    - Aktuelles Alter nehmen oder sich auf die Aktivzeit beziehen?
  + Nationalität
    - Wie sind Doppelbürger erfasst und wie wollen wir diese zählen?
  + Captain / Assistenzcaptain
    - Schwierig, auf einen Spieler aufzusummieren, wenn dieser nicht immer Captain war
* Kanten
  + Anzahl gemeinsame Spiele
    - Zählen wir Einsätze im Spiel oder nur Aufführungen auf dem Matchblatt?
  + Anzahl gemeinsame Saisons
    - Zählt man eine Saison schon bei einem gemeinsamen Spiel? Was wenn ein Spieler z.B. verletzt war und gar kein Spiel gespielt hat?
  + Anzahl verschiedene Vereine
    - Gleiche Frage wie oberhalb

Grundsätzlich erwarten wir eine hohe Datenqualität, da die Daten direkt von der NHL stammen und keine Umwege über Drittanbieter machen. Ausserdem wird diese API z.T. auch auf der Webseite der NHL selber verwendet (z.B. für den Banner mit den aktuellsten Spielen).

# Analysen

|  |  |
| --- | --- |
| These / Frage: | Welche Spieler können auf eine grosse Anzahl an vertrauter Mitspieler setzen und welche haben in ihrer Karriere eher viel verschiedene Mitspieler gehabt. |
| Filterung: | Die Operationen werden auf dem kompletten Netzwerk ausgeführt |
| Analyse: | Die Degree Centrality wird eine wichtige Rolle spielen. Ein tiefer Wert an Anzahl Verbindungen (mit tiefer Gewichtung) sagt aus, dass ein Spieler weniger oft den Klub gewechselt hat. Eine eher weniger hohe Anzahl (dafür mit hoher Gewichtung), deutet auf einen «treuen» Spieler hin. |
| Erwartung: | Spieler, welche lange beim selben Klub sind, werden tiefere (ungewichtete) Degree Centrality Werte haben, jedoch sind die einzelnen Kantengewichte deutlich höher. Als Beispiel wäre hier Joe Thornton zu nennen, der seit der Saison 05/06 für denselben Klub spielt und zuvor nur bei einem weiteren war. |

|  |  |
| --- | --- |
| These / Frage: | Gibt es Cliquen um besonders treue Spieler, die evtl. eine Ära in einem Verein geprägt haben? |
| Filterung: | Die Operationen werden auf dem kompletten Netzwerk ausgeführt |
| Analyse: | Mit K-Core soll ermittelt werden, ob sich bestimmte Cliquen gebildet haben. Für k ist ein sinnvoller Wert währende der Analyse noch zu definieren. Innerhalb der Cliquen kann ein hoher Betweenness-Centrality Wert auf eine wichtige Rolle hinweisen. |
| Erwartung: | Spieler, die in der ersten Analyse als «treue Spieler» analysiert wurden, werden eine gewichtige Rolle in allfälligen Cliquen spielen. |

|  |  |
| --- | --- |
| These / Frage: | Haben «treue Spieler» oder eher gleichbleibende Kader einen Einfluss auf den Erfolg eines Teams? |
| Filterung: | Es werden zwei Teilnetzwerke erstellt, welche jeweils auf ein Team gefiltert werden. Eines ist eher ein erfolgreiches Team, eines ein eher weniger erfolgreiches. Es können auch mehrere solche Gegenüberstellungen gemacht werden. |
| Analyse: | Graph-Density, Netzwerkzentralisierungswerte (Degree-, Betweeness-, Closeness-Zentralisierung) |
| Erwartung: | Tendenziell haben gleichbleibende Kader höhere Erfolgsschancen, jedoch gibt es sicher auch Ausnahmen, z.B. wenn ein gleichbleibender Kader einfach nicht gut genug ist, oder wenn sich eine neu zusammengestellter Kader als Glücksfall erweist und sofortigen Erfolg hat. |