Merging Shot Data and Measuring Player Importance to Team's Offensive Flow

Louis Boulet[0009−0000−3432−2604]

Hockey Data Analyst, LB-Hockey, Ottawa ON, Canada
lbhockey.social@gmail.com

1 Bridging the Data Gap

For the last few years, the quality of available shot data features in the public hockey analytics sphere has really fallen behind what is being used within teams and companies. Although the NHL provides various parameters for each shot attempt, passes leading to them are not registered anywhere. This prevents the use of situational expected goals to measure skills such as expected assists for one's playmaking strength, or expected points from entry shots for the ability to create off the rush.

MoneyPuck (MP) does an excellent job of cleaning and formatting what the NHL offers with precise coordinates, previous events, and their own expected goal model attached to every shot attempt. But this still leaves room for more. Now it does not mean that this type of tracking is entirely unavailable. Corey Sznajder manually tracks hundreds of games every season as part of his incredibly insightful AllThreeZones (A3Z) project. He records more situation context like offence type (cycle, forecheck, rush), screen presence, and most notably, the three most recent passes leading to a shot.

While the NHL's unique shot ID is not present in Sznajder's shot tracking sheets, the MP and A3Z datasets share several features that allow for a potential merging process. The shot time, type (backhand, slap, wrist, tip, etc.), outcome (miss, on net, goal), and shooter-goalie pair are noted in both sources. This is enough for roughly 90% of the tracked A3Z shot attempts to be matched with their equivalents on the other side of the data pond, taking under 25 seconds for a roughly 400-game sample to be generated. The result is a database that links coordinates and expected goal values from MoneyPuck to the contextual indicators and pass tracking from AllThreeZones.

2 Network Analysis Modelling in Hockey

Network analysis concerns the evaluation of relationships within a structure made of nodes that are connected through edges. Analyzing a passing network within a team felt like a natural way to implement this into my work, with the players and passes between them acting as the nodes and edges respectively. These edges will be directed, from sender to receiver, and weighted, using separation from the play's eventual shot (primary, secondary, tertiary assist) and that shot's threat level (expected goal value).

Extensive research has focused on quantifying importance within these networks, using metrics known as centrality measures. And as is the case throughout data science, there is no end-all-be-all to every solution. We will use four centrality measures to build an offensive puck-movement profile, capturing a diversity of contributions and aiming to outline any given player's inclinations with the puck and role when on the attack.

The first is importance through flow, using Betweenness Centrality. Betweenness is calculated by going over how often a player lies on the shortest path between nodes. Players ranking highly here are crucial facilitators who maximize their team's passing routes toward shots. They tend to act as the main link between any two teammates, resulting in their line's offensive movement flowing through them.

We can estimate a player's influence with PageRank Centrality: an iterative process that considers the entire team dynamics. This benefits skaters who connect highly valued players, usually acting as an intermediary between central portions of the network.

By considering all possible paths within a network, Information Centrality finds skaters who enhance the efficiency of their team's offensive system. This is where offensive support pieces tend to shine. They make themselves valuable by creating pathways that allow for flexible progression even if they are not the most direct options, rendering possessions more robust to interruptions should they arise.

Keeping the simplest for last, Weighted Degree Centrality favours high-volume players as it sums the total edge weights coming in and out of a node. Heavily involved passers stand out here even if their plays do not result in groundbreaking connections.

3 Chemistry Interactions

Using these profiles, we can conduct a quick case study of how the different areas of puck-moving interact with each other to establish chemistry at the line level. We'll be taking the Ottawa Senators' first line as an example here. All the ranks mentioned in this section are league-wide among forwards during the 2023-24 season.

First, Brady Tkachuk is the identity piece in Ottawa. He plays a very high-danger-centric style of game and so much of the Senators' offence is geared towards getting him a chance in-tight whether that is through a screened tip, rebound, bumper play, or general net scramble. As a result, Tkachuk has a lot of sway over how the team's plays unfold, giving him a top 3 Influence (PageRank) score.

Next, we have Tim Stützle, who is the primary carrier and dynamic mover. Many of the Senators' possessions are made possible by virtue of the center connecting plays thanks to his playmaking and skating abilities. With such a high share of Ottawa's offensive pathways going through him, Stützle garners a top 3 Flow (Betweenness) mark.

Lastly on the right wing is Claude Giroux. He tends to play more of a tertiary role here, acting as the reliable veteran piece for his younger linemates. But Giroux goes beyond that, with his anticipation of potential passing routes ensuring Ottawa's puck possessions can progress towards a chance no matter their current state. The adaptability and support he provides rank him in the top 10 on the Efficiency (Information) side.

Each player's distinct specialization in a different facet of play-driving and the meshing of these puck-movement abilities allows the line to consistently produce at a high level. Interactive visualizations of these offensive puck-movement profiles are available to all at LB-Hockey.com along with the full-length article.