Introduction

What actually leads to winning a game of League of Legends? In this project, I analyzed thousands of ranked matches and built a cause and effect map that shows how early actions (like taking the first tower, securing dragons, or building a gold lead) flow into later game control and, ultimately, a win. We learn a separate map for each skill tier so you can see how strategies differ from Platinum to Elite. The core question we address is: what are the causal pathways to winning, and how do these pathways differ between lower skilled and higher skilled players? This can be used for players who want practical advice, coaches who want to communicate priorities, and analysts who want structure behind their insights.

In addition, the project allows users to ask “what if” questions such as “If we defeat the Baron and have a gold lead at 20, how likely are we to win?”

Overview

This project uses causal discovery methods to understand what drives winning in League of Legends matches across different skill tiers. I computed rank-specific Bayesian network structures that reveal how early game advantages (like First Blood and early gold leads) translate into mid game objective control (towers, dragons) and ultimately victory.

We can only very cautiously assume causality here. We can approximately satisfy causal sufficiency since many latent factors such as player skill and draft strength influence winning through the observed game state variables. We also enforce time ordered mechanics rules to enforce an acyclic graph. We also enforce game specific rules and stratify by rank to help reduce spurious independencies. Thus we can weakly assume faithfulness. With these assumptions, we can approximate queries such as $p(Win \mid do(Baron = 1))$ as $\Sigma_z p(Win|Baron = 1, z)p(z)$ where z is the set of observed, non-descendant parents of Baron in the learned graph. This is the backdoor adjustment formula.

Methods

I implemented a complete pipeline using the Greedy Equivalence Search (GES) algorithm to learn causal graph structures from match data. The pipeline preprocesses raw match statistics into 13 discrete game state variables spanning early game (0-14 min), mid game (14-25 min), and late game (25-30+ min), then applies structure learning with temporal constraints to ensure causes precede effects. We enforce domain knowledge through forbidden edges (such as winning cannot cause early-game events) and use bayesian information criterion scoring to penalize spurious correlations. The result is a Completed Partially Directed Acyclic Graph (CPDAG) for each rank tier that represents causal equivalence classes of directed acyclic graphs. We then estimate conditional probability tables (CPTs) using Bayesian parameter estimation and compare structures across ranks to identify strategic differences.

The codebase is organized into different python scripts. preprocessing.py handles data discretization using domain-informed thresholds (for example, 1000 gold = item breakpoint), “ges.py” implements structure learning with temporal constraints, “parameters.py” estimates CPTs using Bayesian Dirichlet equivalent uniform (BDeu) priors, “queries.py” aids probabilistic inference, and “visualize.py” generates graph visualizations. The CLI (through cli.py) makes it easy to run the full pipeline (`python -m src.cli full`) or individual steps. Models are saved as pickled objects (GES results and Bayesian networks) for each rank. The project uses `causal-learn` for structure learning, `pgmpy` for Bayesian network modeling, and standard scientific Python libraries (numpy, scipy, pandas) for data processing.

Results

We show the probability of winning conditioned on every game state for the Diamond rank.

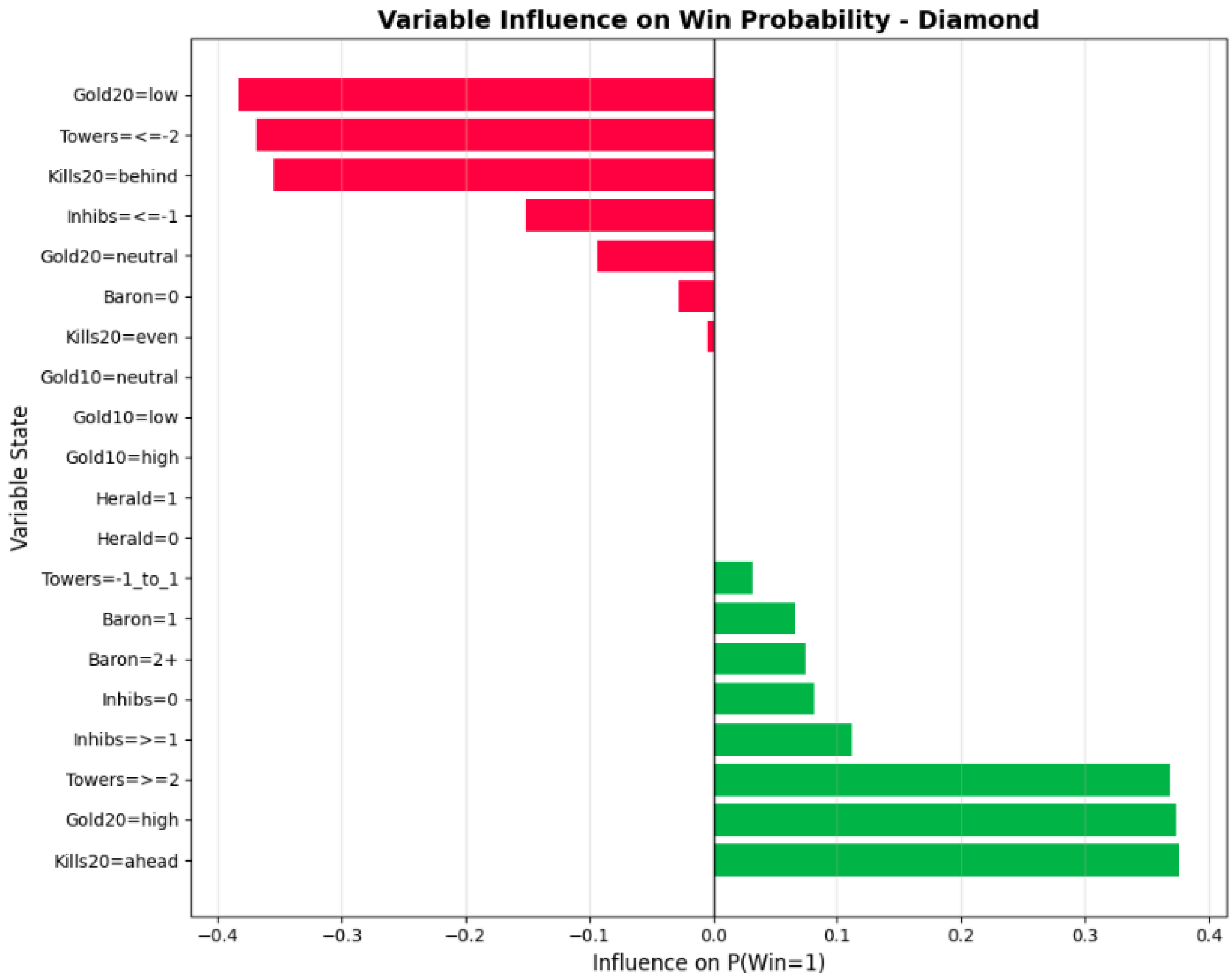

Figure 1: Conditional probability of winning for the Diamond Rank

These results indicate that the greatest indicators of winning in the Diamond rank are not having destroyed towers, having high gold, and having many kills, which aligns with personal experience in the game.

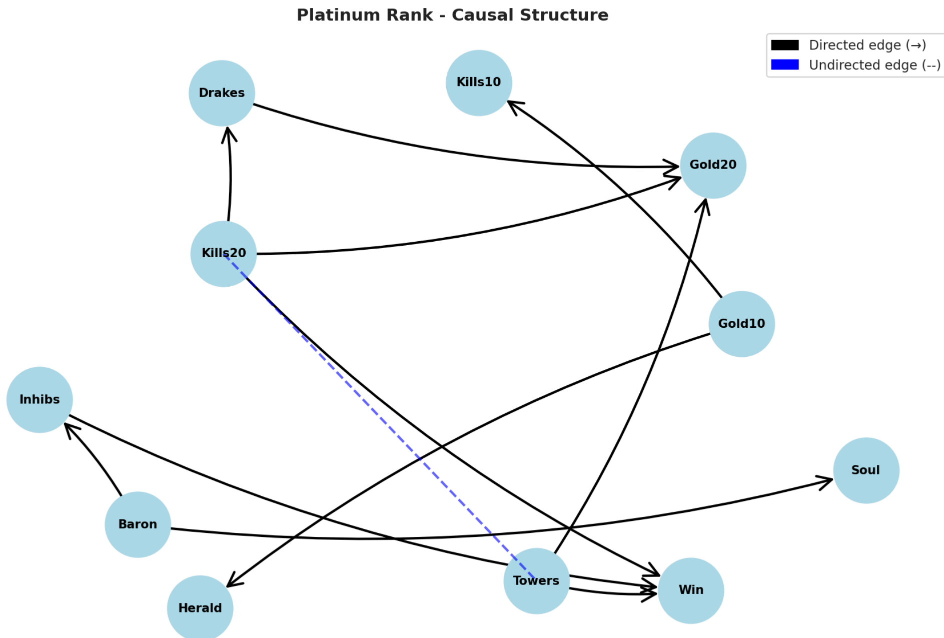

Figure 2: Causal Structure of gameplay for Platinum Rank

We see that platinum games reflect a combat-first style where mid game kill leads tend to unlock the rest of the map. The graph shows kills around 20 minutes feeding into both drake control and gold advantages, while towers are tied to kills but with ambiguous direction. This suggests that teams often win fights first and then safely siege. Winning is closely linked to both towers and mid game kills, and Baron often translates into inhibitor pressure and sometimes Soul. Practically, when ahead in fights, Platinum teams should immediately convert momentum into dragons and towers to secure a durable path to victory rather than lingering in extended skirmishes.

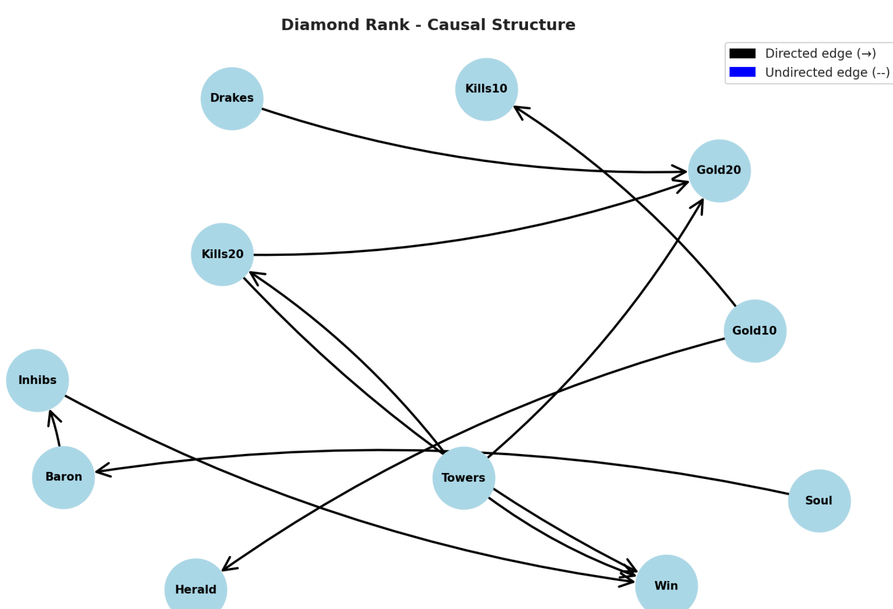

Figure 3: Causal Structure of gameplay for Diamond Rank

Diamond teams look more tower first as towers come before and create better fights. The graph highlights towers pointing to gold and to mid game kills, and directly to winning, with kills still contributing to outcomes but as a second order effect. Taking towers opens the map, increases safe vision, and sets up favorable engagements. Baron continues to map into inhibitor pressure, and there's evidence that acquiring Soul positions teams to secure Baron afterward. The practical takeaway is to prioritize early tower pressure to shape the terrain of fights, then use that structural lead to secure Baron setups and close out games.

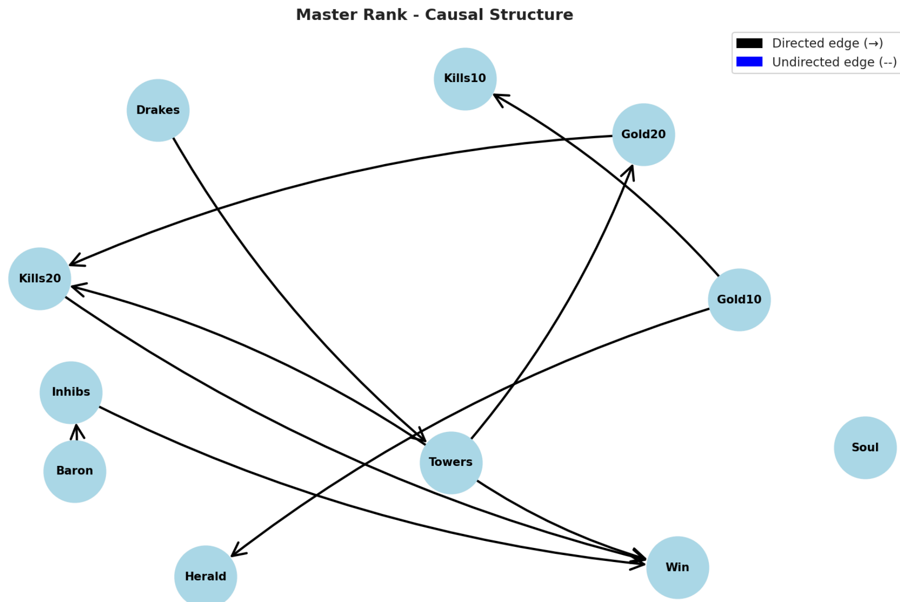

Figure 4: Causal Structure of gameplay for Master Rank

Master play emphasizes objective chaining and timing discipline. An edge from drakes to towers suggests early dragon control helps teams translate pressure into structural gains. Towers then drive both gold and mid game kills, and those kills strongly affect win probability. There's also an explicit path from mid game gold to kills, capturing how itemization timing transforms economic leads into combat power.

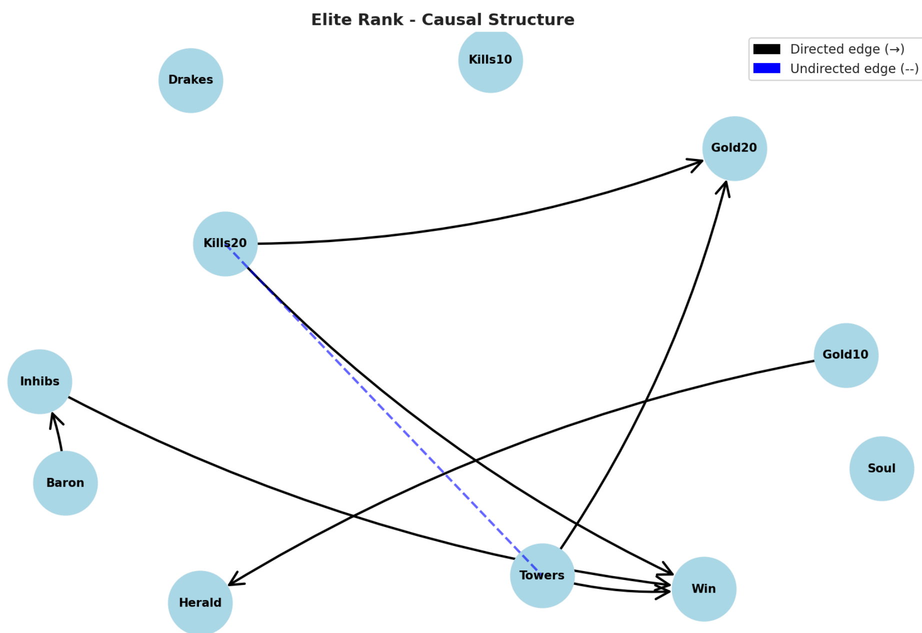

Figure 4: Causal Structure of gameplay for Elite Rank

Elite graphs are sparse, which shows that gameplay has condensed into a few “metas” (or optimal strategies). Towers and inhibitors dominate the path to winning, with Baron feeding directly into inhibitor pressure. This minimal structure suggests that high level teams streamline decisions around essential objectives and let map advantage dictate when and where to fight.

Limitations and Future Steps

The biggest caveat in this project was the causal assumptions (causal sufficiency, acyclicity, and graph faithfulness). An actual controlled study would be needed, not just observation, to assure causality. Also, some edges are undirected (equivalence class), so causal direction is not fully identified without stronger assumptions or interventional data.

Future analysis would involve adding more covariates to improve identifiability, models that are more suited for temporal analysis (dyanmic bayesian networks), or comparing our results with different structure learning models such as NO TEARS.

References

1. Zheng, X., Aragam, B., Ravikumar, P., & Xing, E. P. (2018). DAGs with NO TEARS: Continuous Optimization for Structure Learning. ArXiv.org. https://arxiv.org/abs/1803.01422
2. Chickering, D. (2002). Optimal Structure Identification With Greedy Search. Journal of Machine Learning Research, 3, 507–554. https://www.jmlr.org/papers/volume3/chickering02b/chickering02b.pdf
3. Dataset: https://www.kaggle.com/datasets/californianbill/patch-25-14-lol-league-of-legends-ranked-games