Graph Reinforcement Learning for Resilient Supply Chain Network Design

Nicolas Smits[1,*], Arsh Bhatia[1], Carl D. Laird[1]

[1]Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA

*Corresponding author: nsmits@andrew.cmu.edu

Theme: AI and Intelligent Process Systems

Abstract

Supply chain network design under disruption risk requires deciding which transport links to add to an existing network so that customer demand continues to be met when facilities or links fail. Because the benefit of a candidate design depends on how inventory is allocated and replenished once components have failed, risk is typically estimated by simulating network operations over a large set of failure scenarios, and the combinatorial space of candidate designs renders simulation-based search increasingly expensive as networks grow. This work formulates network design as a finite-horizon Markov Decision Process (MDP) in which the state comprises the observable attributes of the current network and the remaining design budget, and each action either adds one candidate arc or terminates the design. A learned policy adds at most five arcs to a multi-echelon network of plants, warehouses and customers, and each design is evaluated by the fraction of baseline unmet demand it removes across simulated failure scenarios, net of a weighted arc addition cost. The policy is trained by Double Deep Q-Learning (DDQN) and observes the network through a Graph Attention Network (GAT) encoder whose node and arc features are restricted to quantities available in routine operational records (safety stocks, storage and production capacities, throughput, lead times and lane priorities) so the value of an arc must be inferred from network structure alone. Our Q-network evaluates every candidate arc with a single shared scoring function that takes as input the embeddings of the two endpoint nodes, the features of the arc itself, and the remaining design budget. Terminating the design is assigned a fixed value of zero, so that an arc is added only when its estimated contribution to the design value, net of its cost, is positive. As the number of parameters is independent of network size and arc costs and rewards are normalized by the number of customers, a policy trained on small networks can be applied without modification to considerably larger ones. The policy was trained on networks ranging from 2×3×4 to 3×6×10 (plants × warehouses × customers) and evaluated on held-out networks excluded from training that extend to 10×20×40, a fourfold increase over the largest training size. All methods were evaluated on a common set of failure scenarios and greedy search served as the baselines. In our experiments, the learned policy outperformed the unmodified network and two randomised design baselines at every size considered. It attained between 64% and 80% of the design value of greedy search at the training sizes and designed a 10×20×40 network in approximately 1 CPU-second, compared with 6 CPU-hours for greedy search.

Keywords: reinforcement learning, graph neural networks, supply chain network design, resilience, size generalization