Novel research has been conducted to transform the structure of networks into tensors suitable for Deep Learning (DL) – motivating whether DL can be used to discover network weaknesses and exploit them. Should DL agents gain the ability to exploit and attack networks, the security and robustness of infrastructure and information networks comes in question. Are there any limits on the capacity for DL to discover network weaknesses and selectively attack them? If not, can this capacity be mitigated through efficient strategies which conceal some of the information about the network, allowing for a defensive response?

We will create a framework which maps this targeted network attack to a two-player game, allowing us to leverage the field of Deep Reinforcement Learning. Through this novel framework, the first DL player will be presented a network and will be rewarded for choosing nodes which quickly break apart the network – simulating network attack. The second DL agent will receive the network before the first agent and will hide a fraction of the network links before it is shown to the first agent. The second agent will be rewarded if the partially concealed network requires more moves to be broken by the first DL agent – simulating network defense. To this end, we will determine the limits to malicious DL attackers and develop defensive strategies to reduce their effectiveness.