Influence Maximization in Unknown Social Networks: Learning Policies for Effective Graph Sampling

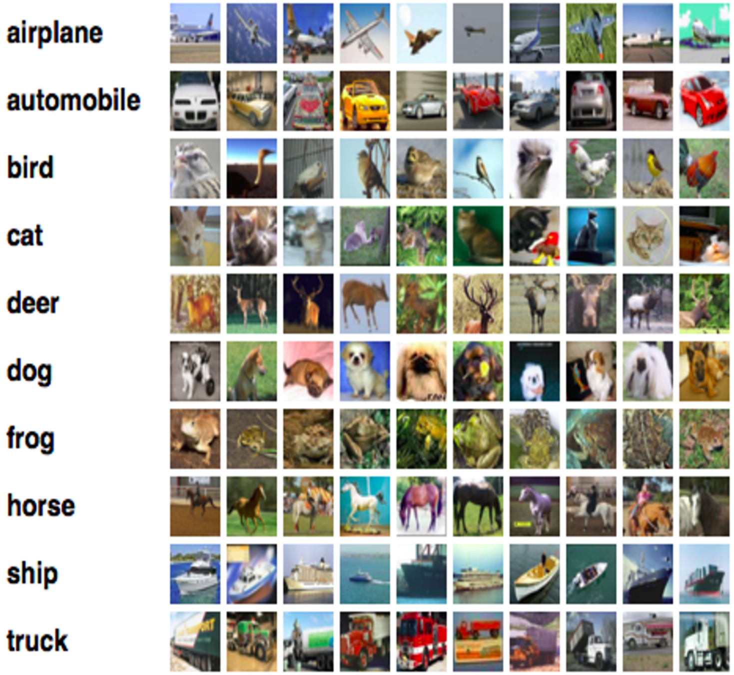
B. Ravindran

Reconfigurable and Intelligent Systems Engineering (RISE) Group
Department of Computer Science and Engineering

Robert Bosch Centre for Data Science and Artificial Intelligence (RBC-DSAI)
Mindtree Faculty Fellow
Indian Institute of Technology Madras

Machine Learning

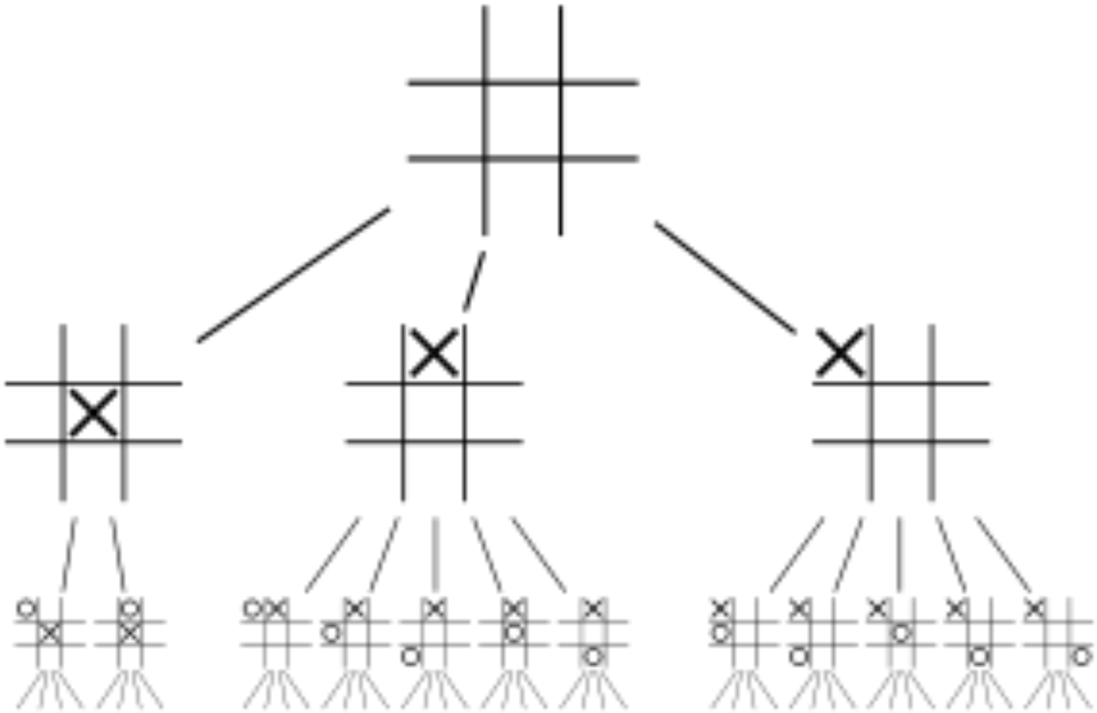
Learn functions from input to output from data



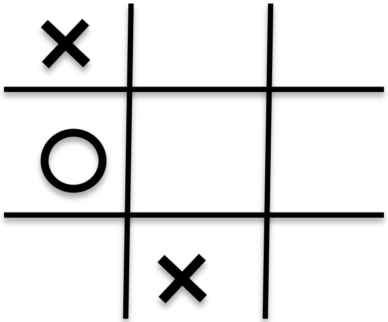
Learning to Control

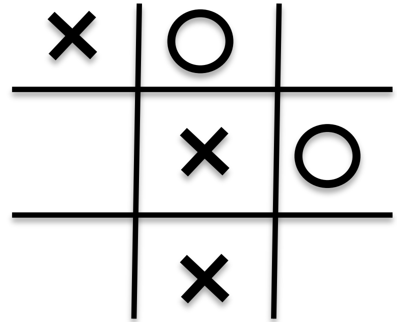
- Familiar models of machine learning
  - Learning from data.
- How did you learn to cycle?
  - Not from Data!
  - Trial and error!
  - Falling down hurts!

Tic-Tac-Toe

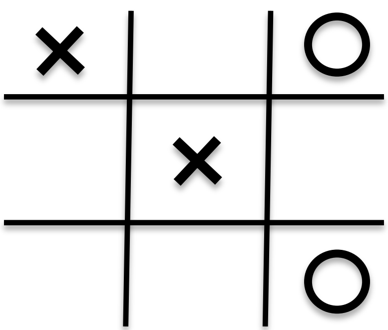


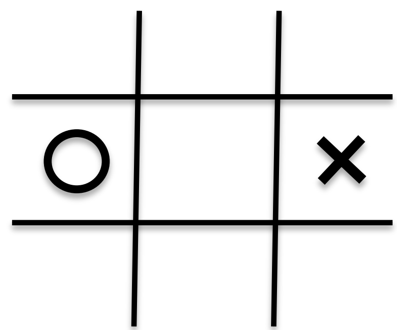
Supervised Learning



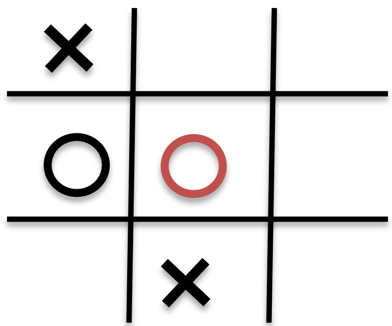


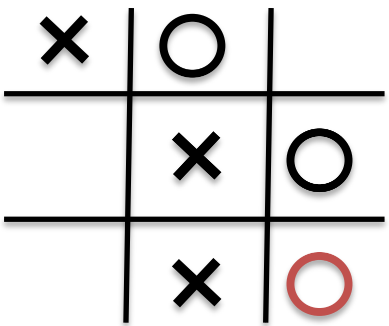
Current Positions



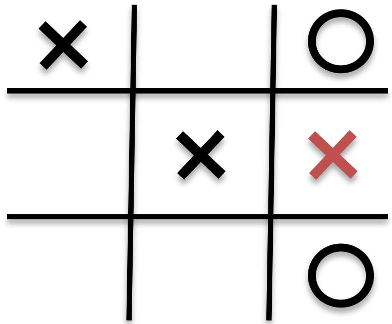


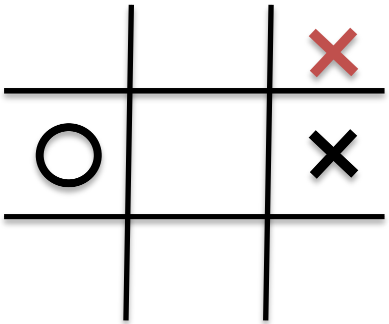
Supervised Learning





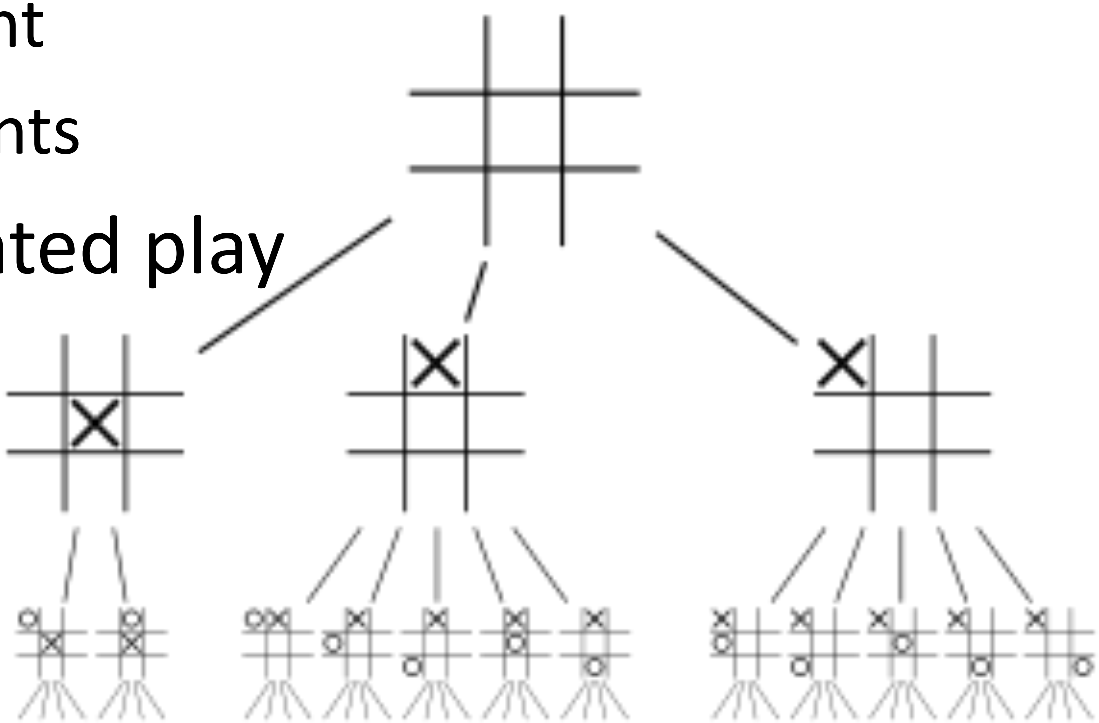
Expert Moves



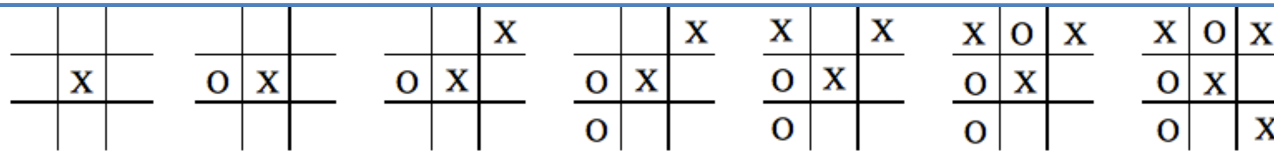


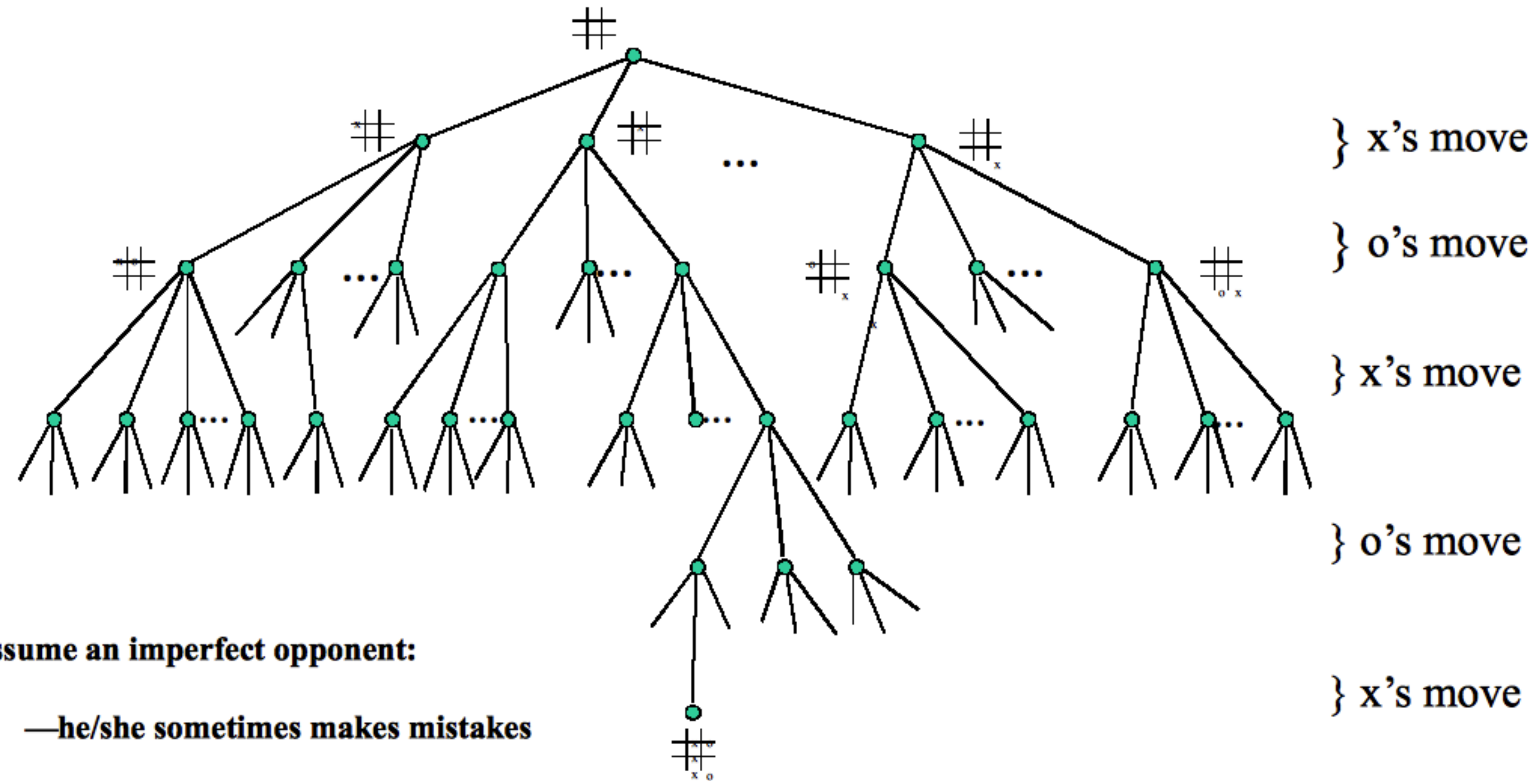
Reinforcement Learning

- Learn from evaluation
  - Win gives 1 point
  - Loss gives -1 point
  - Draw gives 0 points
- Learn from repeated play



More Tic-Tac-Toe

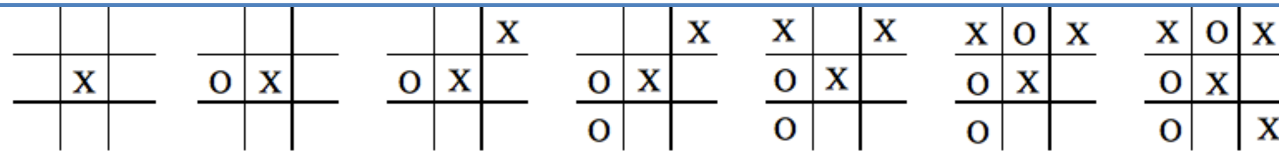


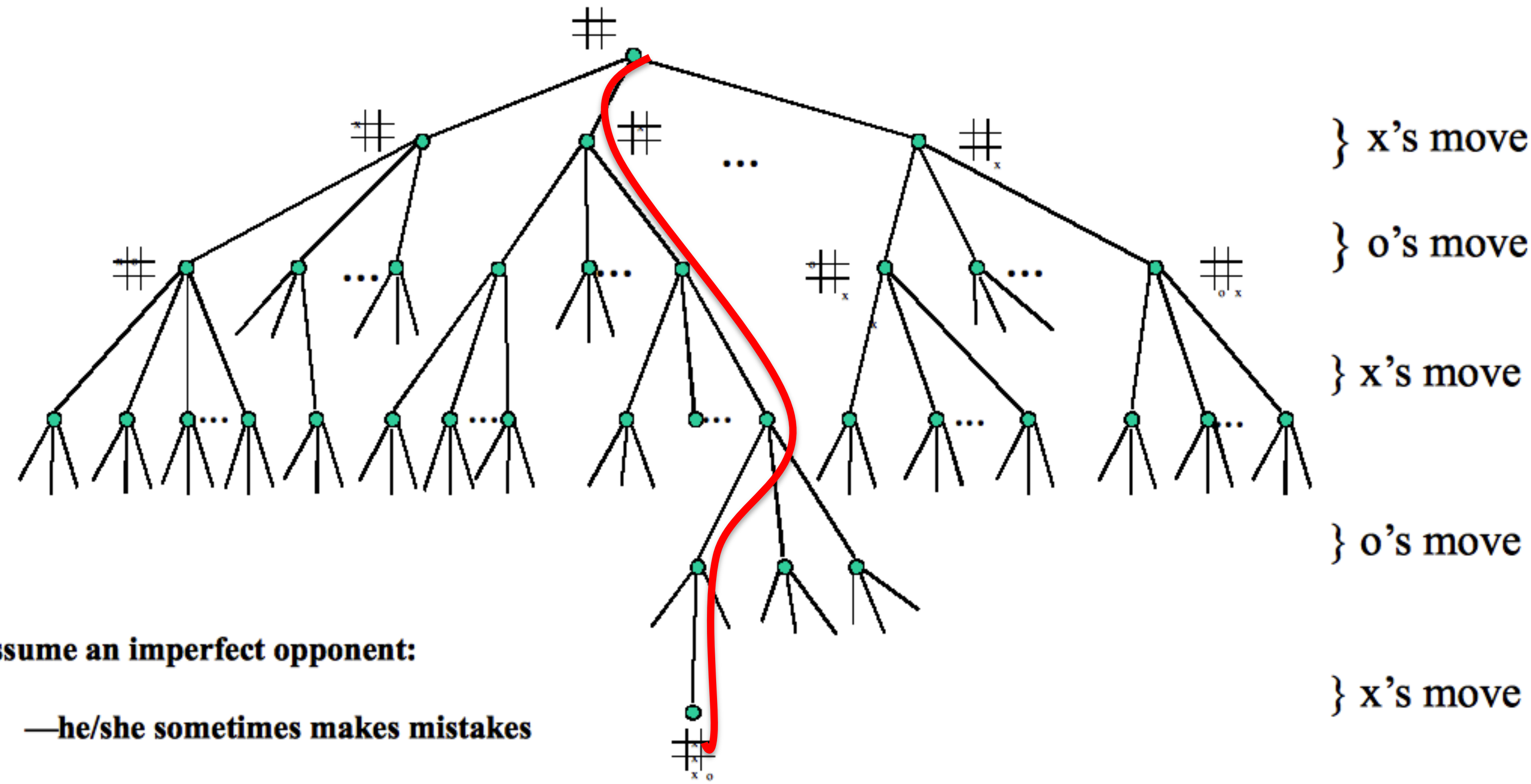


Assume an imperfect opponent:

—he/she sometimes makes mistakes

More Tic-Tac-Toe

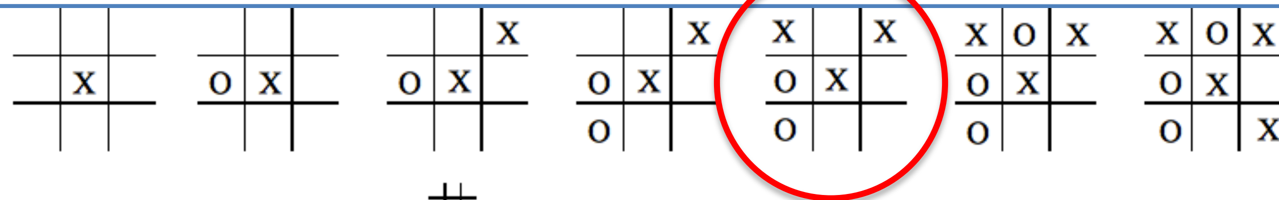


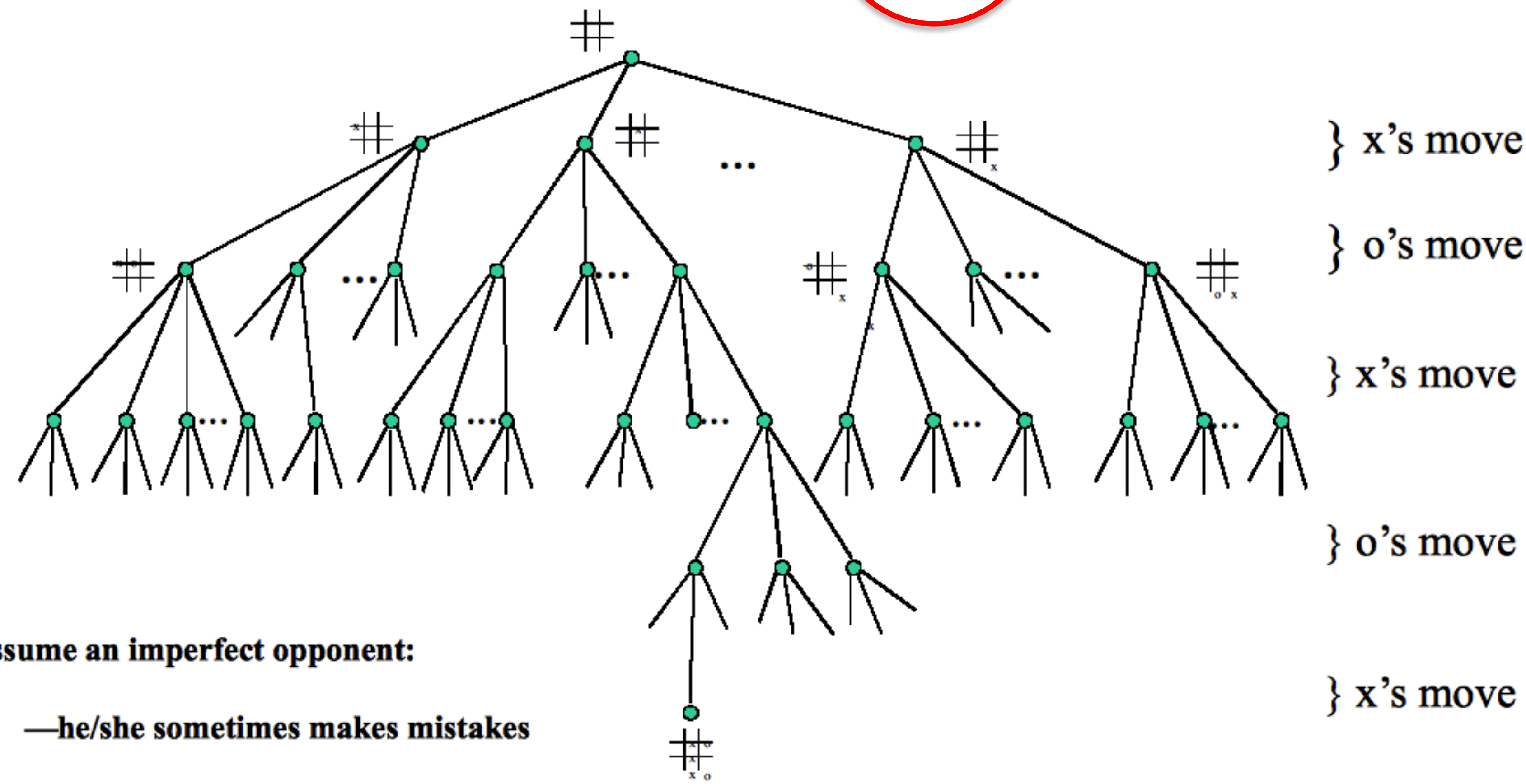


Assume an imperfect opponent:

—he/she sometimes makes mistakes

More Tic-Tac-Toe





Assume an imperfect opponent:

—he/she sometimes makes mistakes

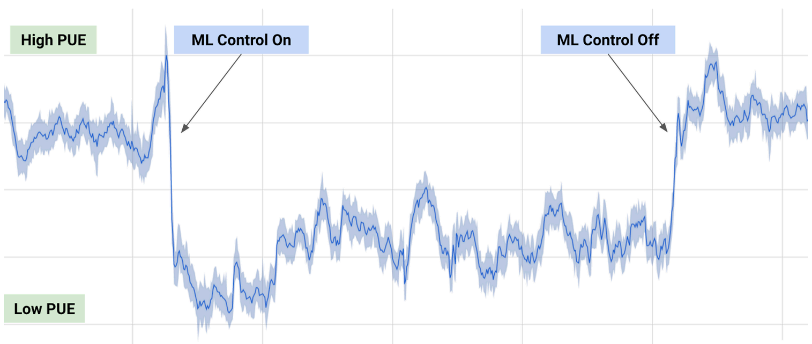
Temporal Difference

Barto, Sutton, Anderson ‘83

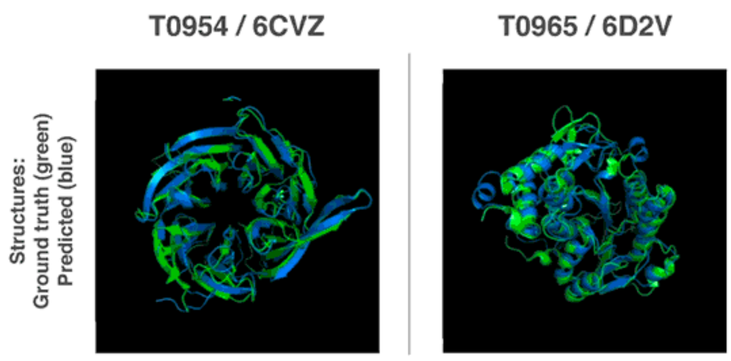
- Simple rule to explain complex behaviors
- Intuition: Prediction of outcome at time $t+1$ is better than the prediction at time t. Hence use the later prediction to adjust the earlier prediction.
- Has also had profound impact in behavioral psychology and neuroscience!



RL Works



High PUE
ML Control On
ML Control Off
Low PUE



T0954 / 6CVZ
T0965 / 6D2V
Structures:
Ground truth (green)
Predicted (blue)

To watch such a strong programme like Stockfish, against whom most top players would be happy to win even one game out of a hundred, being completely taken apart is certainly definitive.
Viswanathan Anand
It's like chess from another dimension.
Demis Hassabis

nature
ALL SYSTEMS GO

Shop for Florists in Chennai on Google
THE ULTIMATE GO CHALLENGE
15 MARCH 2016
AlphaGo
Lee Sedol
280

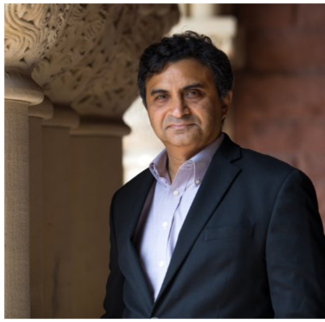
Designing an RL solution

- States
  - Enough information to take decisions
  - Raw inputs often not sufficient
- Actions
  - The control variables
  - Discrete – items to recommend, moves in a game
  - Continuous – torque to a motor, rate of mixing,
- Rewards
  - Define the *goal* of the problem

Influence Maximization in Unknown Social Networks: Learning Policies for Effective Graph Sampling







Harshavardhan Kamarthi (IITM->Georgia Tech), Priyesh Vijayan (IITM-> McGill/Mila), Bryan Wilder (Harvard), B. Ravindran (IITM), Milind Tambe (Harvard/Google)

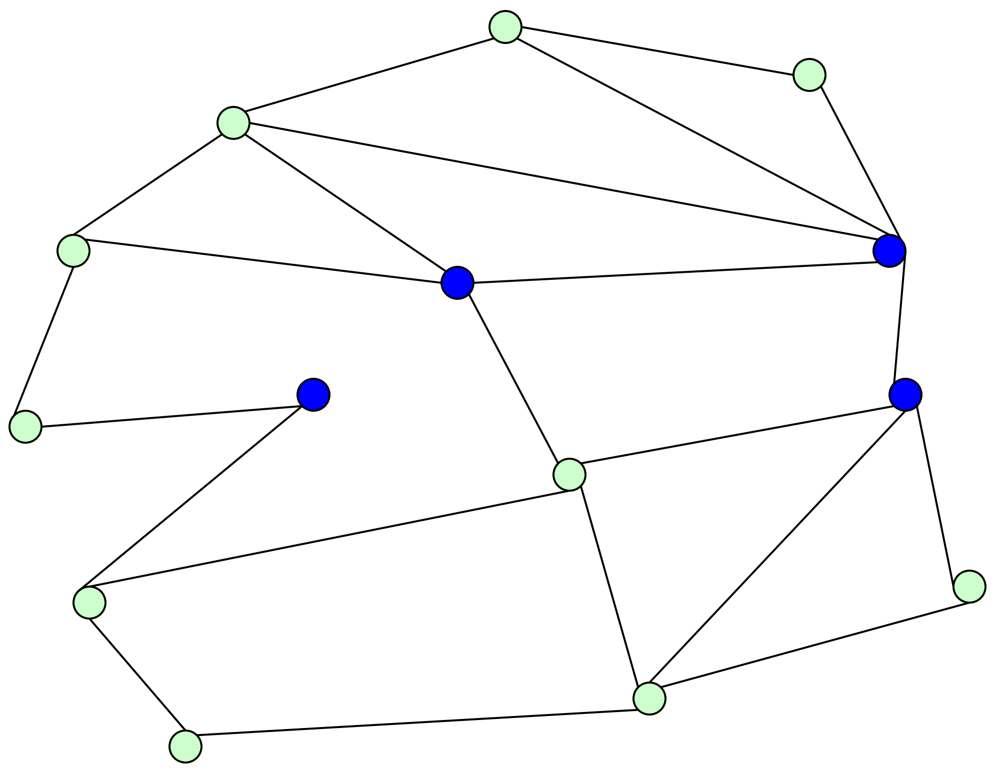
Best Paper Runner-up AAMAS 2020.

http://arxiv.org/abs/1907.11625

Influence Maximisation Problem

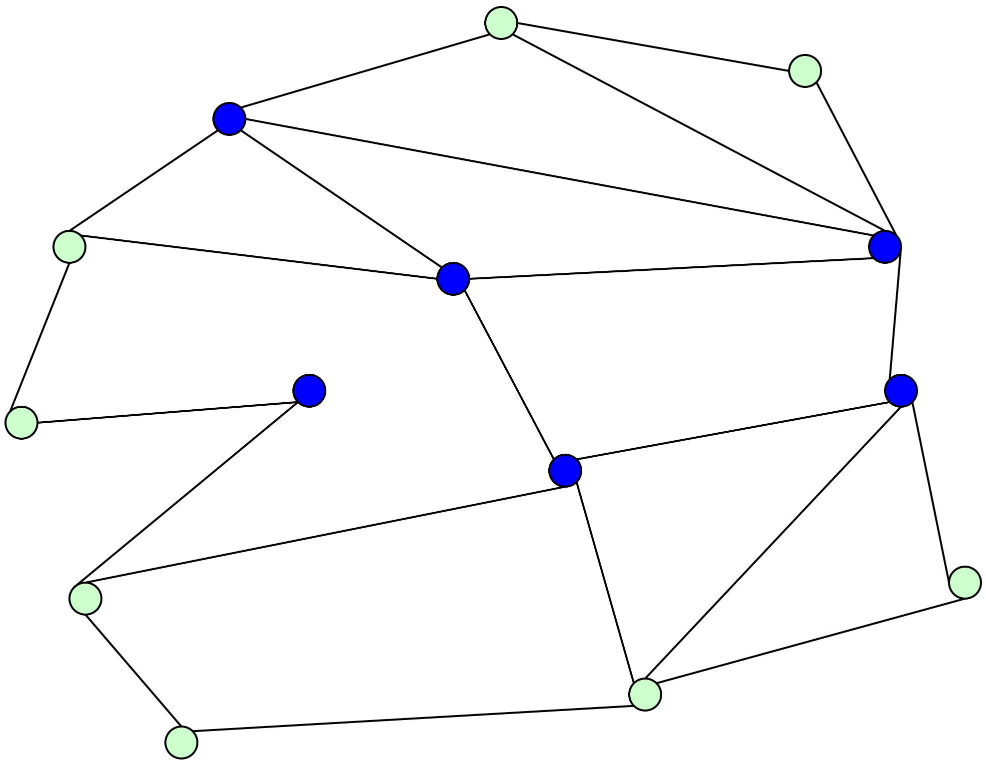
- Objective: Pick influential nodes from a social network as peer leader to help disseminate information to maximum number of nodes in the network

Spread of Influence



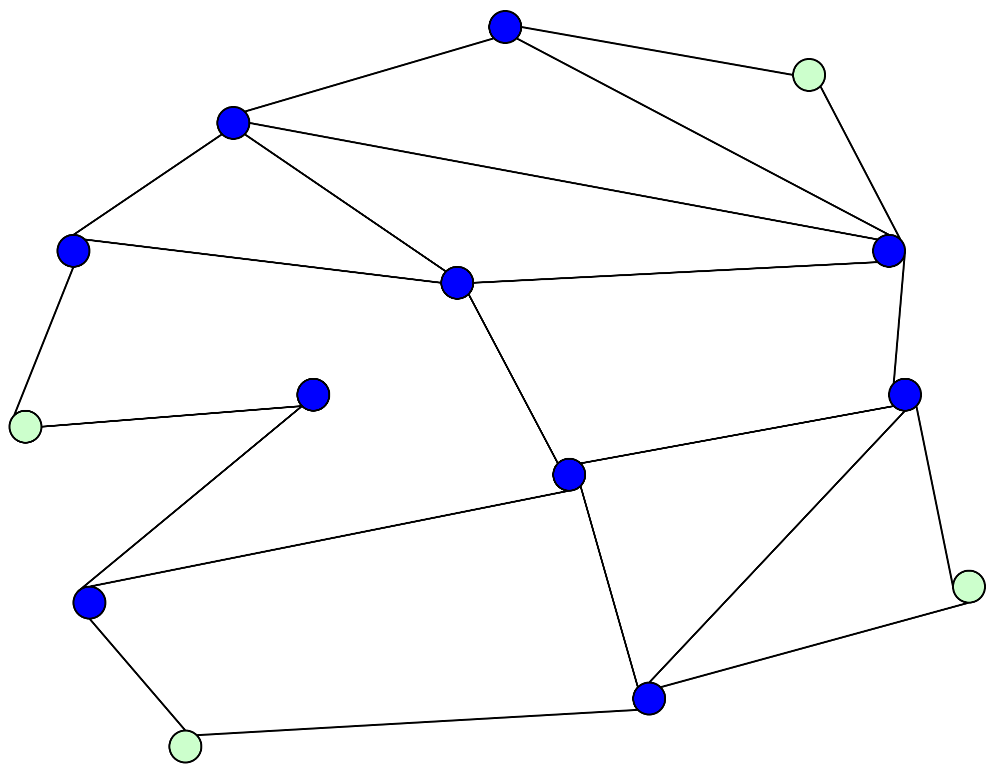
A node is influenced is at least two of its neighbours are influenced.

Spread of Influence



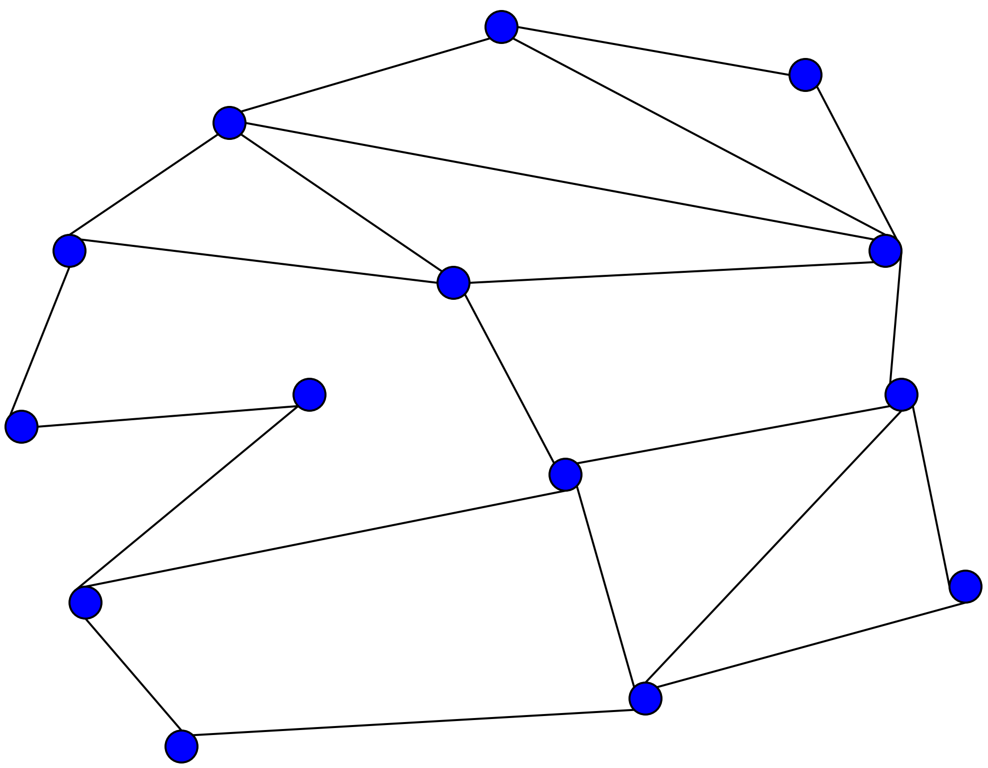
A node is influenced is at least two of its neighbours are influenced.

Spread of Influence



A node is influenced is at least two of its neighbours are influenced.

Spread of Influence



A node is influenced is at least two of its neighbours are influenced.

Influence Maximisation Problem

- Objective: Pick influential nodes from a social network as peer leader to help disseminate information to maximum number of nodes in the network

- More specifically, given a graph $G = (V, E)$, select K nodes to activate such that the information flows across the edges from V results in maximum number of nodes receiving information.

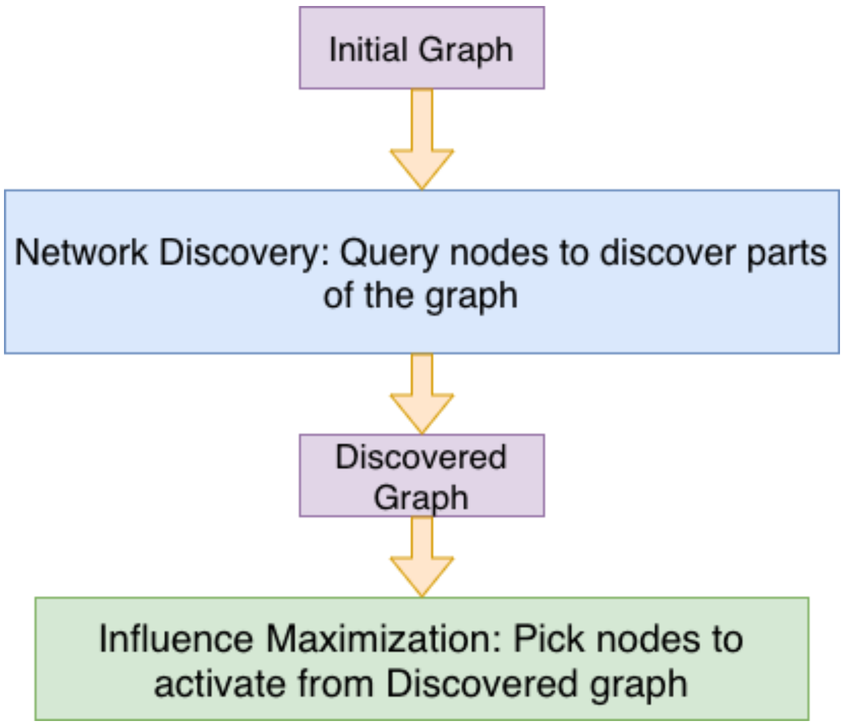
- Have found applications in substance abuse [VP07] interventions, micro-finance adoption [Ban+13], HIV prevention [Yad+18; Wil+18a] , etc.

- Previous works use a greedy algorithm [KKT05] to minimise computational cost of simulating the influence spread over entire networks.

Influencing real world social networks

- We have little data on the structure of social network
- Cost of collecting data in terms of time and effort is high
  - Typically operate with a budget of queries

Network Discovery for influence maximisation

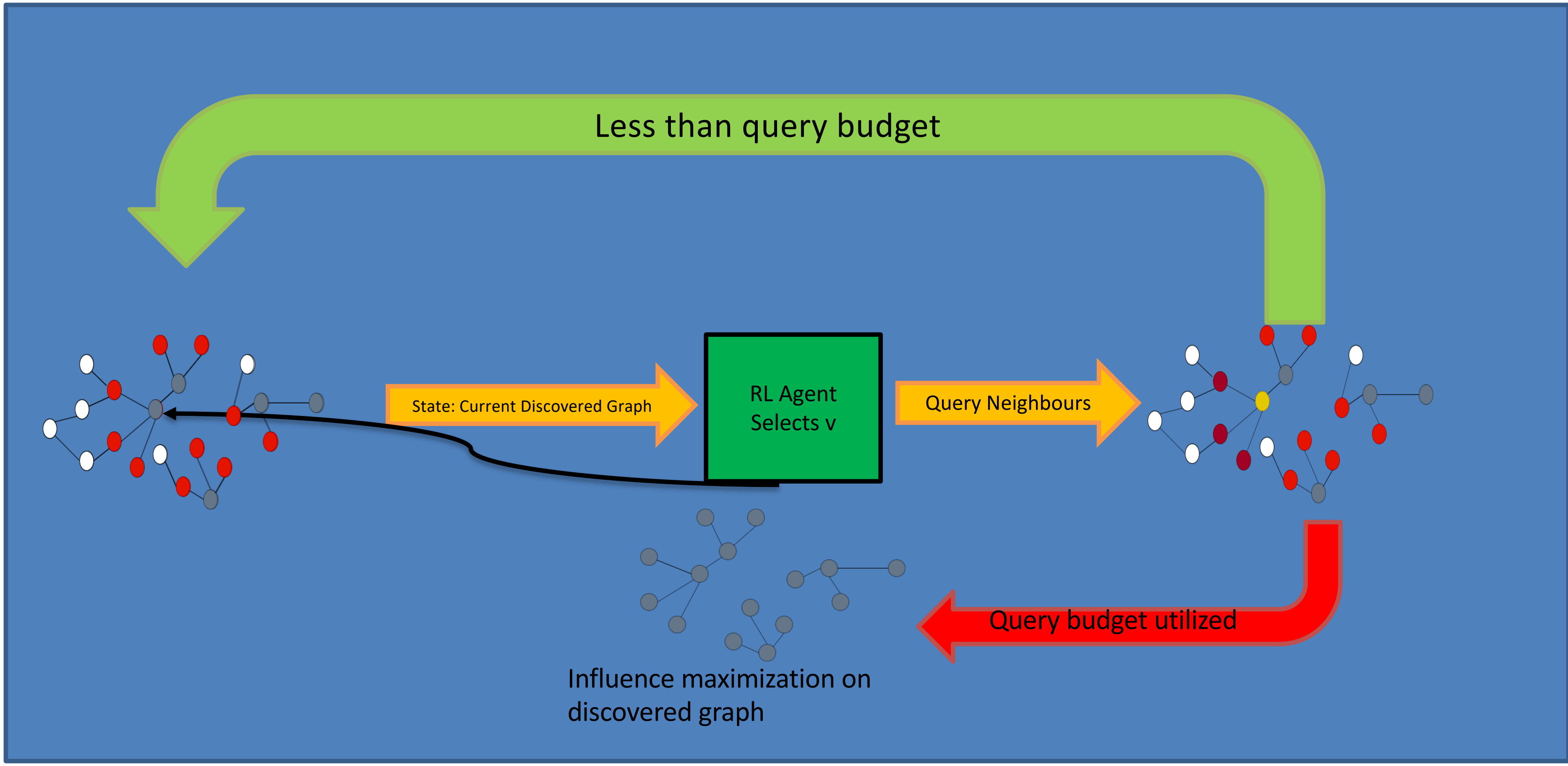


- Sample subset of nodes in the network to query.
- The queried nodes reveal their respective neighbours.
- Then we use any influence maximisation algorithm on discovered graph to pick K nodes to activate.

RL Formulation

- State: the current state is the discovered graph G_t.
- Action: All nodes in G_t that have not been queried.
- Reward at the end is based on no. of nodes influenced in G^* after discovering graph G_T (denoted by $I_{G^*}(G_T)$).

RL Solution

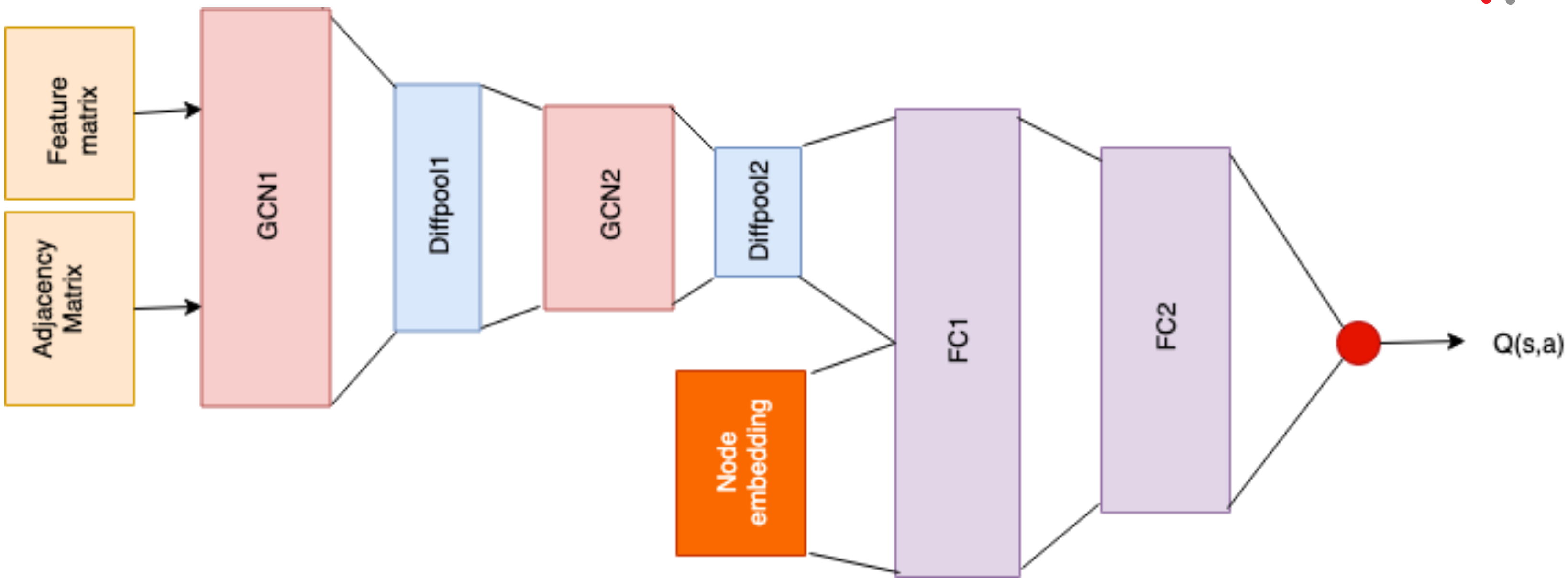

Less than query budget
State: Current Discovered Graph
RL Agent
Selects v
Query Neighbours
Query budget utilized
Influence maximization on
discovered graph

RL Formulation

- State: the current state is the discovered graph G_t.
- Action: All nodes in G_t that have not been queried.
- Reward at the end is based on no. of nodes influenced in G^* after discovering graph G_T (denoted by $I_{G^*}(G_T)$).

Challenges

- Do not have access to the actual network to run multiple iterations of learning
- Have to work ***zero-shot***
- Need a way of generalising across graphs
    - across an *appropriate class* of graphs
- Typical heuristics make apriori assumptions about graph structure



- Differential pooling based Graph convolutional (GCN) architecture [HYL17] to obtain graph representation.
- Deepwalk representation of nodes ϕ [PAS14] for actions as well as node features for (GCN) input.
- We input both state s and action a representation to the DQN and train it to predict the state-action value $Q(s, a)$.

Datasets

	Description	Nodes	Edge relations
Animal	Wildlife contact networks collected by [Dav+15]	Voles	2 voles caught is same trap
Retweet	Each network models retweets about hashtags of specific topics. (Source: networkrepository.com)	Twitter user	one of the users retweets the tweets of the other
Rural	Networks gathered by [Ban+13] to the study diffusion of micro-finance in Indian rural households.	household	health, finance, family, friendship
Homeless	Homeless youth networks from HIV intervention campaigns [Wil+18a; Wil+18b]	Individual	acquaintance

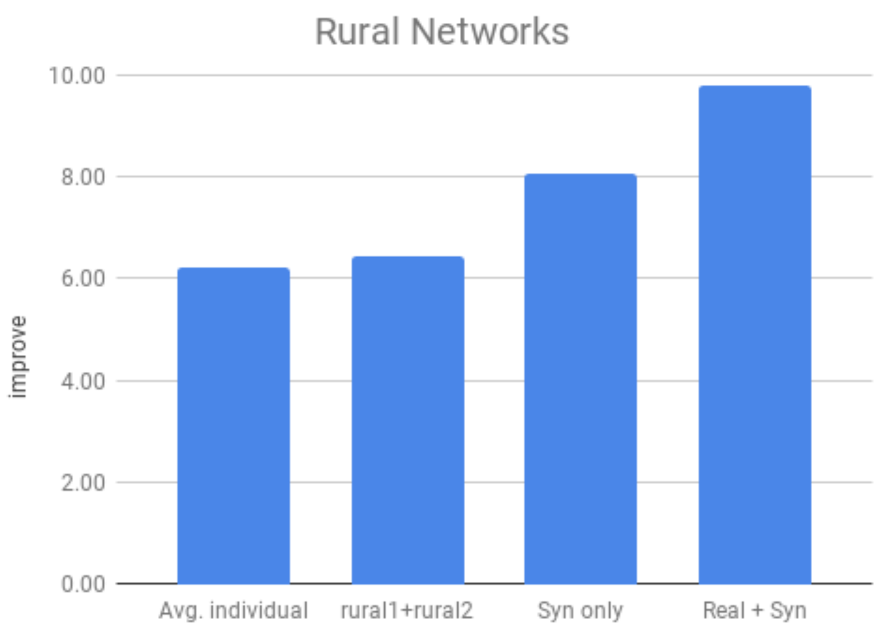
Synthetic Graph generation

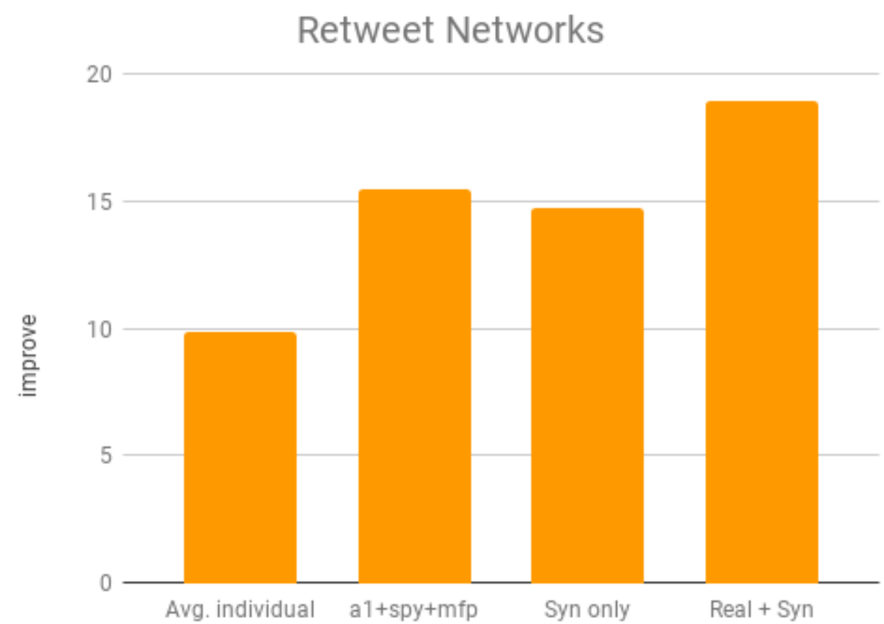
- Why?
  - Don’t have too many actual networks for training
  - Use known structural properties of the family of networks
  - Data augmentation for richer experience from data.
- Used Stochastic block models to generate graphs similar to training graphs.
- For retweet networks enforced that each community is a star graph

Results: Overview

Network Family	increase %	improve %
Rural	10.54	23.76
Animal	36.03	26.6
Retweet	33.87	19.7
Homeless	21.03	7.91

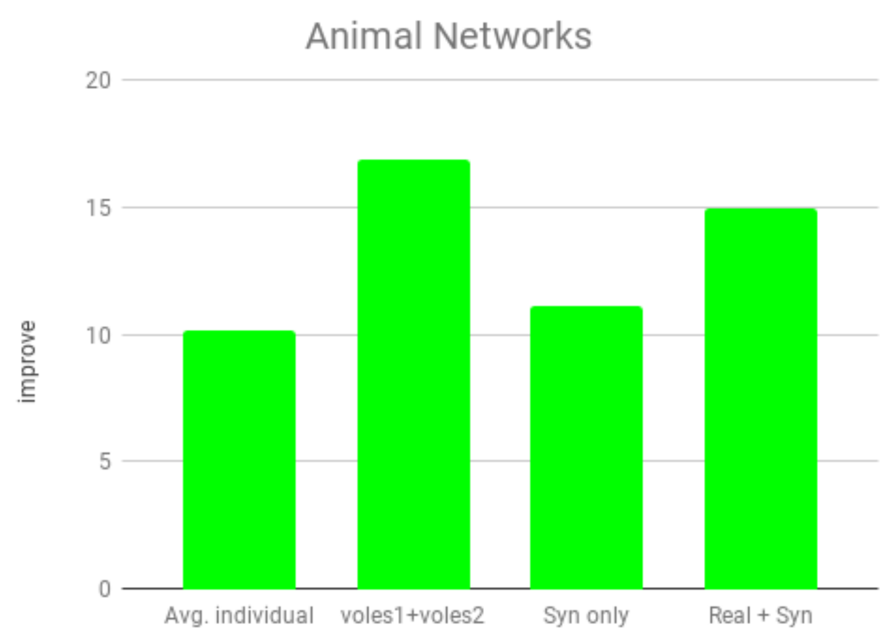
Scores are averaged over test networks for each class


Rural Networks
improve
10.00
8.00
6.00
4.00
2.00
0.00
Avg. individual
rural1+rural2
Syn only
Real + Syn


Retweet Networks
improve
20
15
10
5
0
Avg. individual
a1+spy+mfp
Syn only
Real + Syn

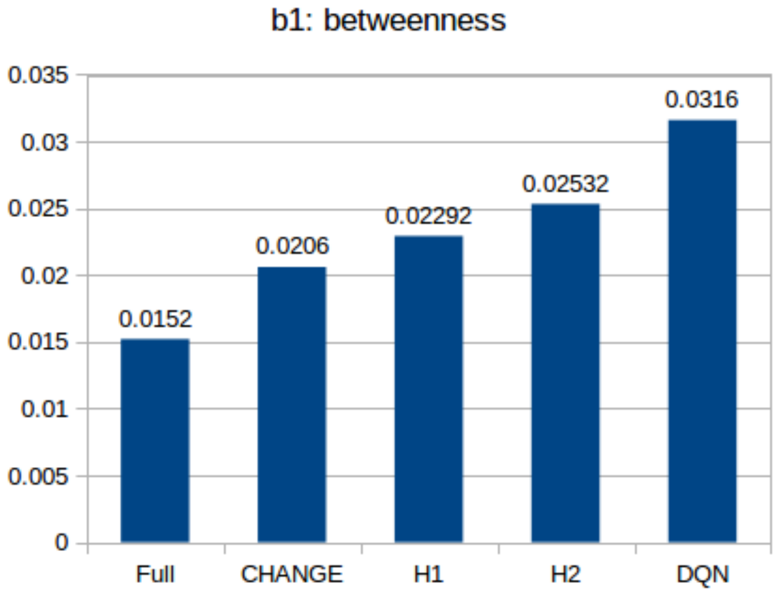

Homeless Networks
improve
30
20
10
0
Avg. individual
a1+spy+mfp
Syn only
Real + Syn

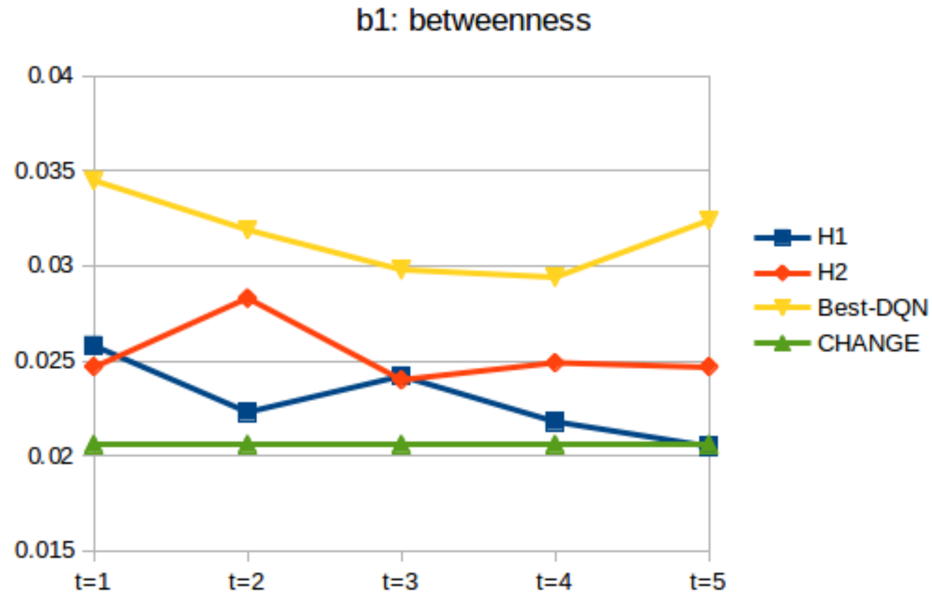
X

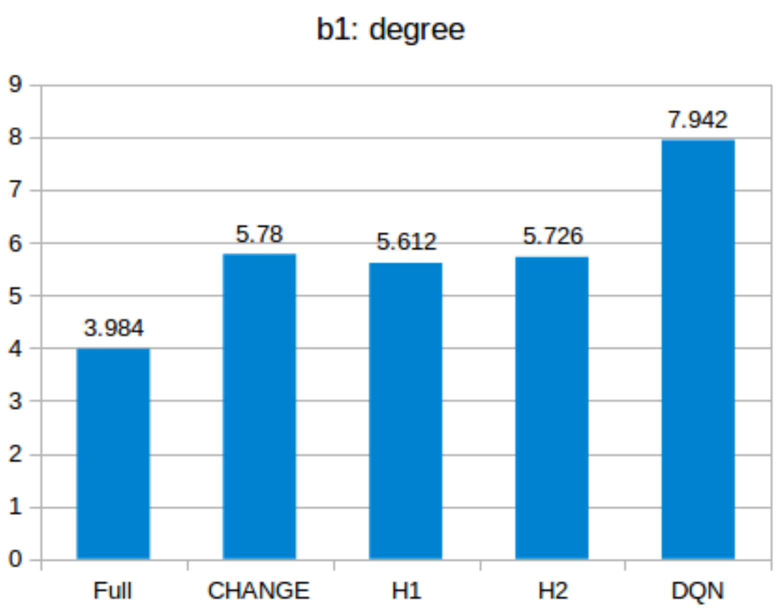

Animal Networks
improve
20
15
10
5
0
Avg. individual
voles1+voles2
Syn only
Real + Syn

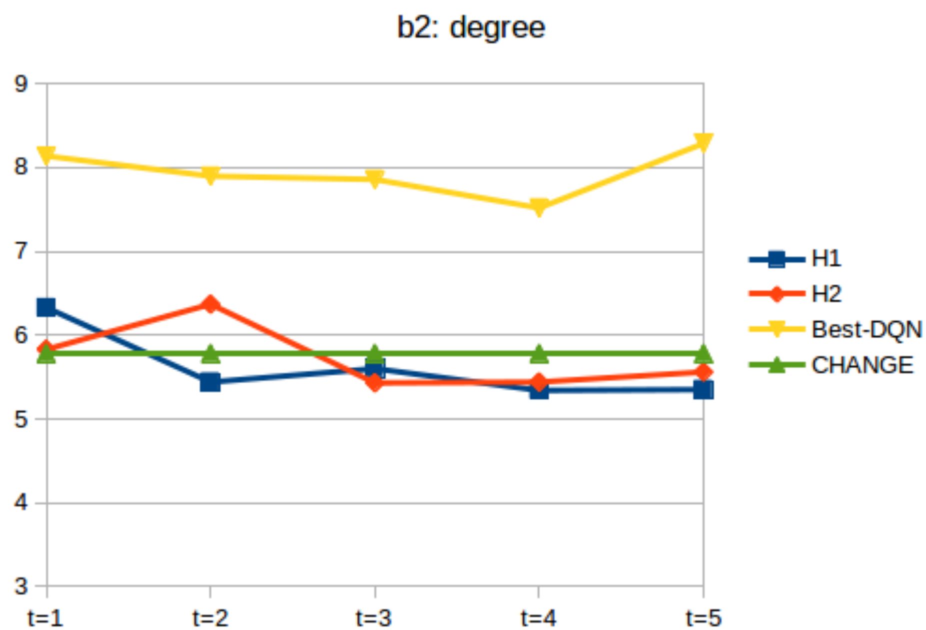
Properties of nodes queried

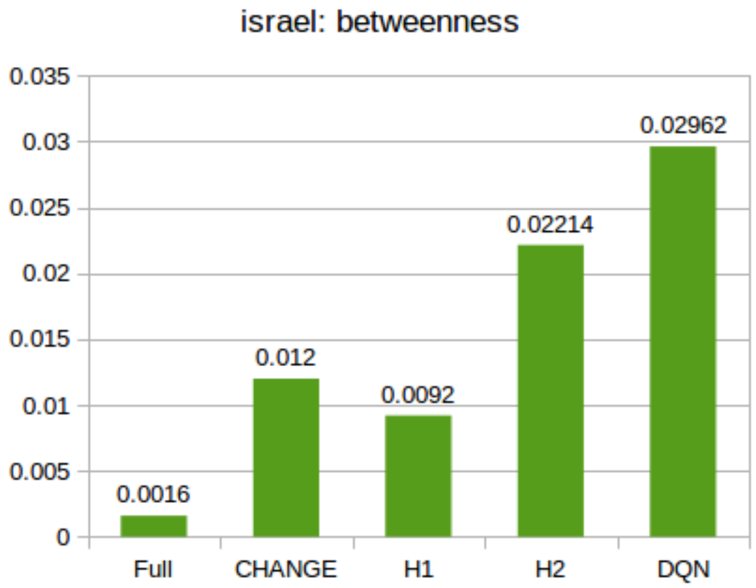
- We study two properties of nodes queried by DQN, H1, H2 and CHANGE on graphs *israel(retweet)* and *b1(homeless)*
- Degree centrality: can cover larger number of nodes during discovery.
- Betweennes centrality: act as bridge between different strongly connected communities of nodes.

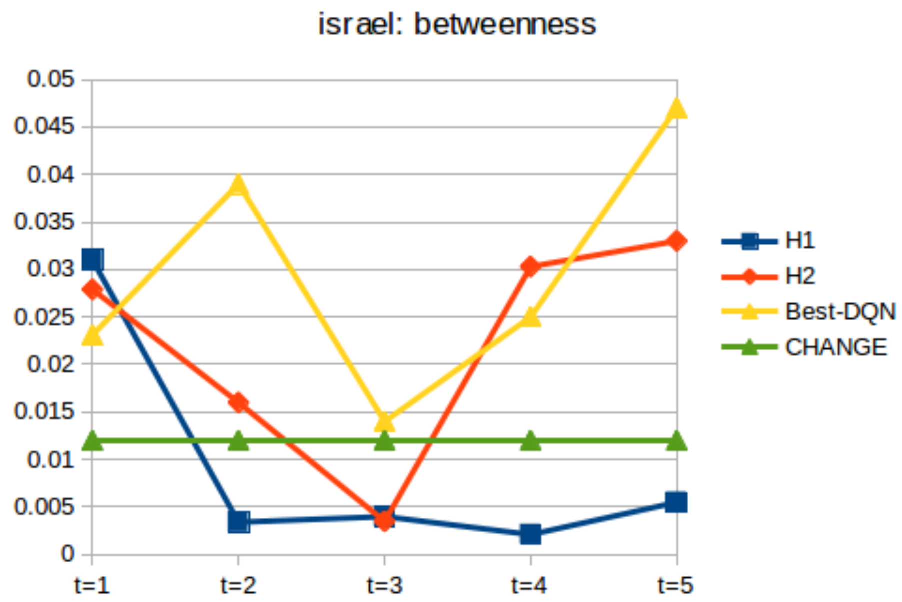

b1: betweenness
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0.0152
0.0206
0.02292
0.02532
0.0316
Full
CHANGE
H1
H2
DQN

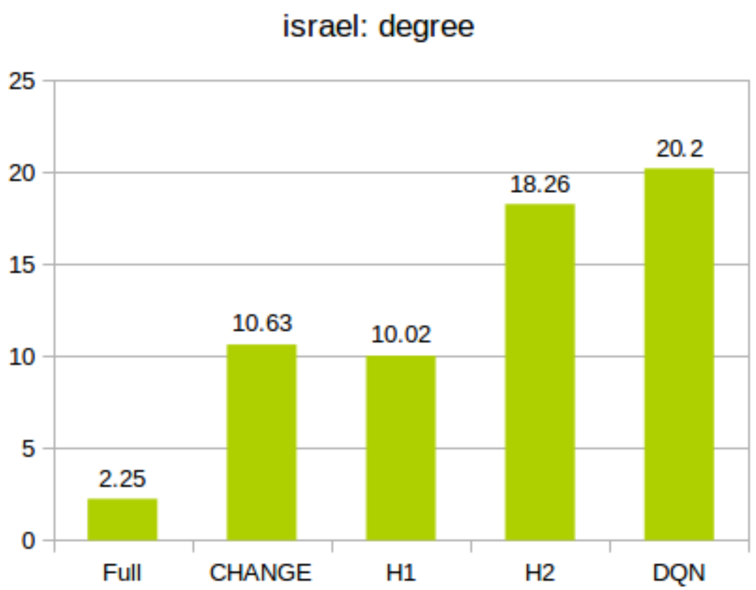

b1: betweenness
0.04
0.035
0.03
0.025
0.02
0.015
t=1
t=2
t=3
t=4
t=5
H1
H2
Best-DQN
CHANGE

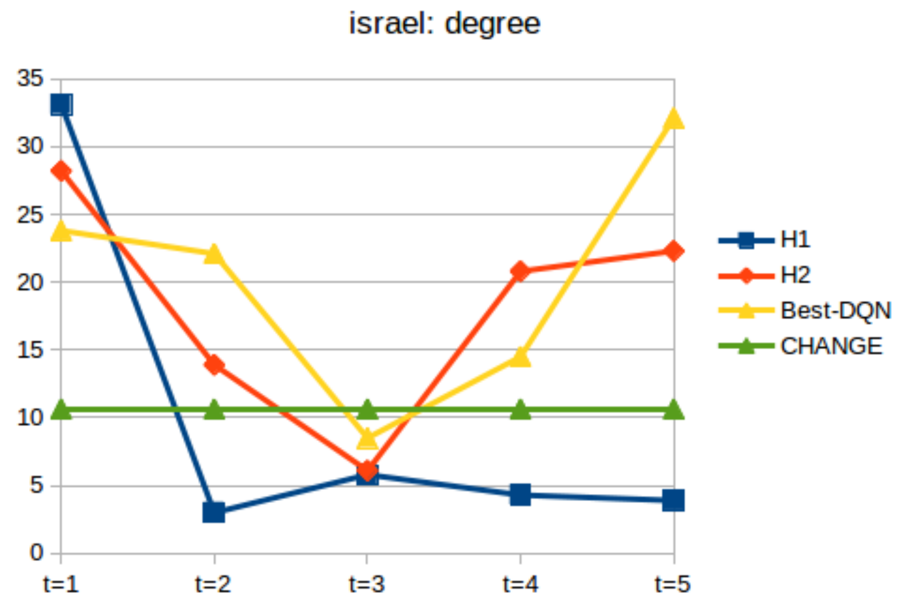

b1: degree
9
8
7
6
5
4
3
2
1
0
3.984
5.78
5.612
5.726
7.942
Full
CHANGE
H1
H2
DQN


b2: degree
9
8
7
6
5
4
3
t=1
t=2
t=3
t=4
t=5
H1
H2
Best-DQN
CHANGE


israel: betweenness
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0.0016
0.012
0.0092
0.02214
0.02962
Full
CHANGE
H1
H2
DQN


israel: betweenness
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
t=1
t=2
t=3
t=4
t=5
H1
H2
Best-DQN
CHANGE


israel: degree
25
20
15
10
5
0
2.25
10.63
10.02
18.26
20.2
Full
CHANGE
H1
H2
DQN


israel: degree
35
30
25
20
15
10
5
0
t=1
t=2
t=3
t=4
t=5
H1
H2
Best-DQN
CHANGE

Conclusions

- Proposed a deep learning based method to leverage structural properties learn effective policies for the network discovery problem for influence maximisation on undiscovered social networks.

- Observed 10-36% improvement over CHANGE.

- Graph embeddings learned by our models was used to pick nodes with high betweenness centrality with respect to the entire network, which was key to discovering important portions of the social network.

RL, really?

- RL needs much larger volumes of data than DL

- RL works only in games where a detailed simulation is available

RL, really?

- RL needs much larger volumes of data than DL
  - Naïvely, yes.
  - Domain models, Historic data, Data augmentation, Transfer learning, etc.
- RL works only in games where a detailed simulation is available
  - Realistic simulations are not needed
  - Rollouts are often enough

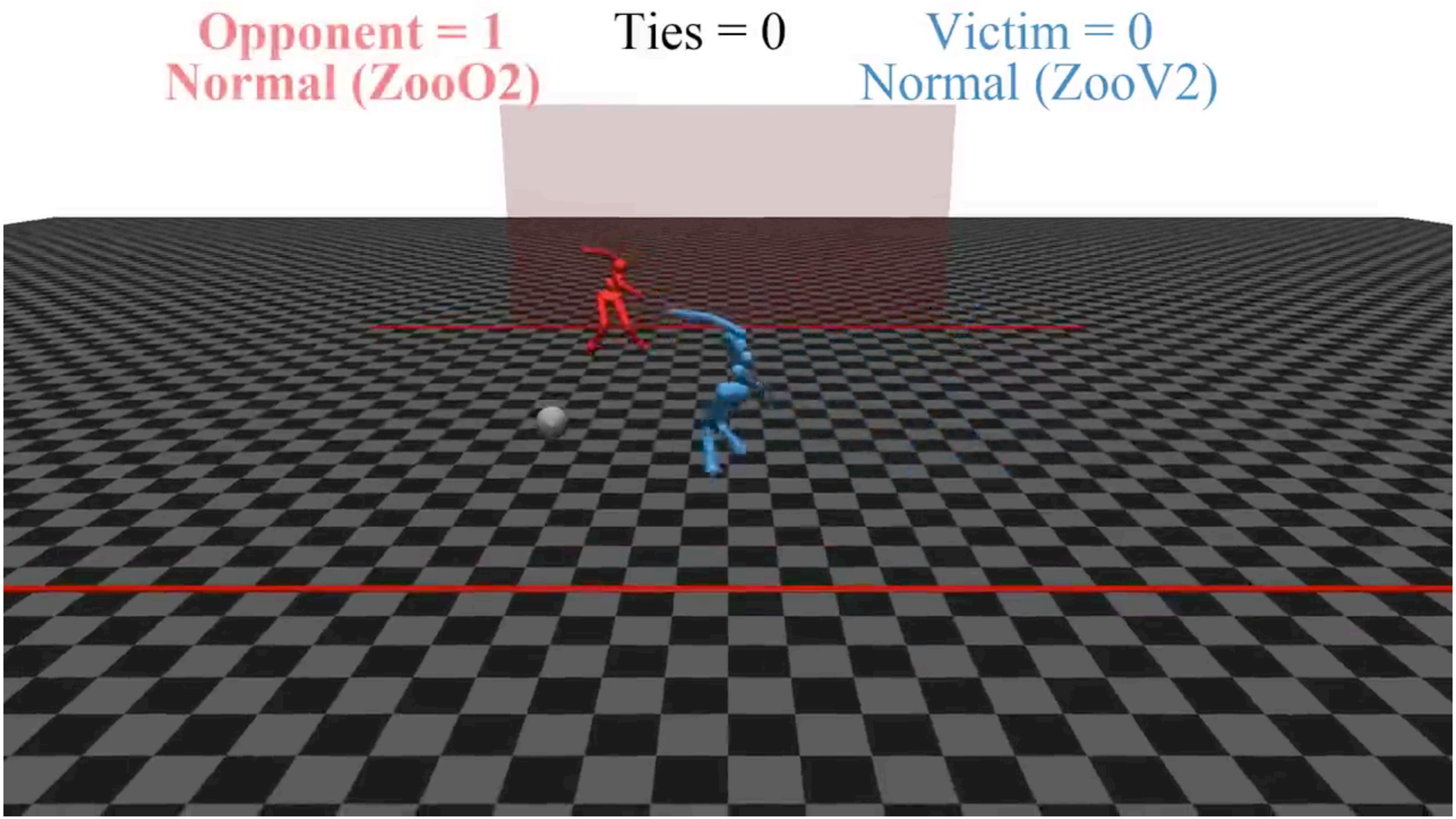
Final Thoughts

- Finally we have building blocks to use RL for real problems
- We need a concerted effort to widen the scope of RL
- Evolve a set of best practices for building data science applications, enterprise applications built on RL
- Put RL to work! Go beyond fun and games! :-)

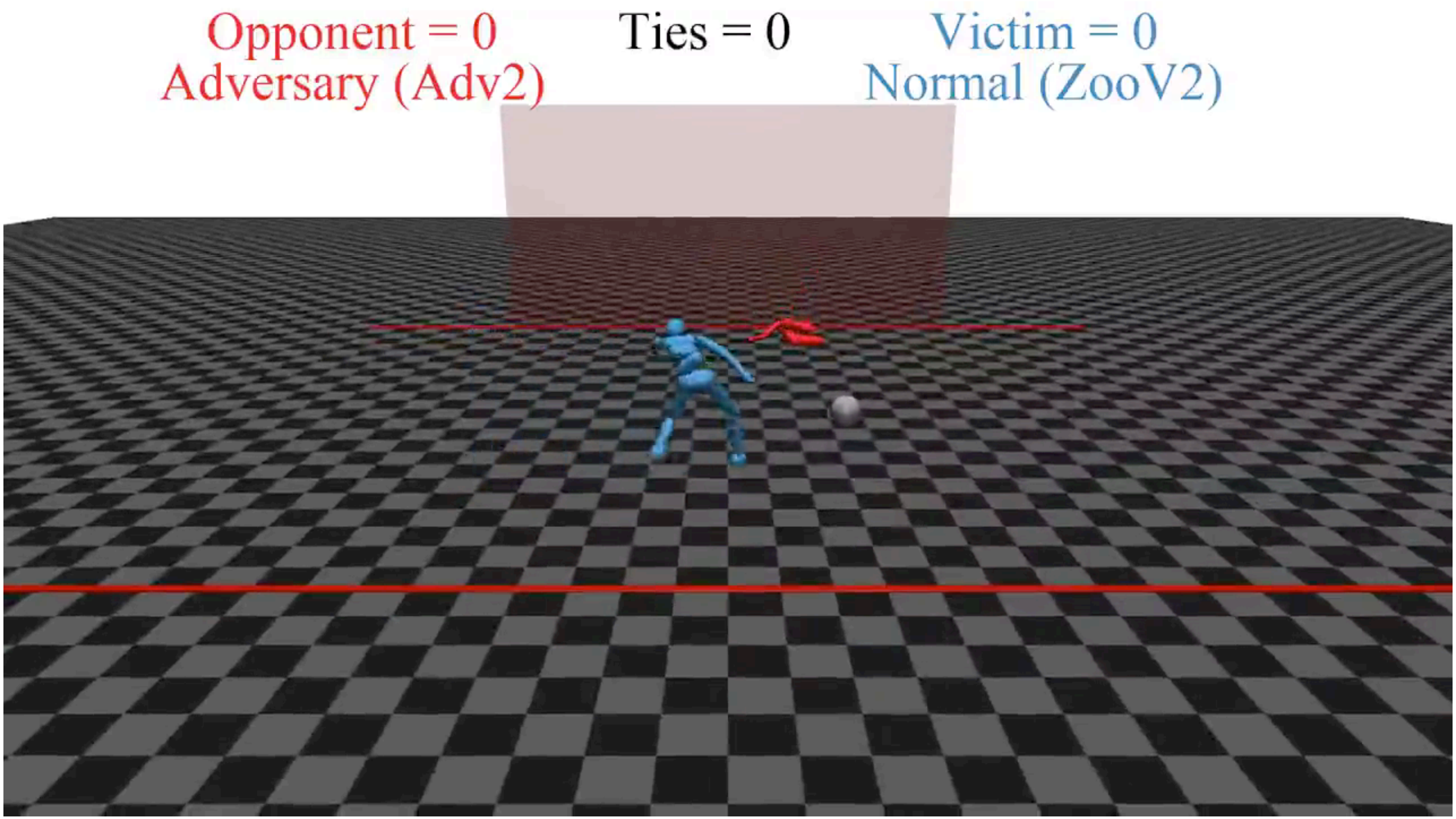
Fundamental Problem

- Learn associations from data
- Current ML based systems' understanding of the world is about that of a 2 year old
  - But no knowledge of the structure and processes of the world

RL and football!



Adversarial Goalie!



Fundamental Problem

- Learn associations from data
- Current ML based systems' understanding of the world is about that of a 2 year old
  - But no knowledge of the structure and processes of the world
- Long way to human-like performance
- But handle it with care in the meanwhile...

AI/ML is more like this kid

Source: Disney

More Learning

Recent short tutorial at Purdue:

http://nanohub.org/resources/30809

Full 12-week course:

http://www.cse.iitm.ac.in/~ravi/courses

RBC DSAI

ROBERT BOSCH CENTRE FOR DATA SCIENCE AND ARTIFICIAL INTELLIGENCE IIT MADRAS

Vision & Mission

- To achieve global prominence in artificial intelligence and data science research with significant societal impact, by carrying out foundational research, and creating top-quality engineers.

- Fundamental Research in the areas of deep learning, network analytics, reinforcement learning, learning with limited and partial data, causal modelling, ethics, fairness and explainability in AI.

- Applied Research in four verticals viz., manufacturing analytics, financial analytics, smart cities, systems biology & health care.

- Quality Education and Societal impact

RBCDSAI Collaborations

RBCDSAI IITM

APPLIED MATERIALS make possible
DVARA Trust
SOCIETE GENERALE
DASSAULT AVIATION
TATA CONSULTANCY SERVICES
intel
INDIAN RAILWAYS
Microsoft Research
GAVS
ERICSSON
Honeywell
NATIONAL INSTITUTE FOR RESEARCH IN TUBERCULOSIS
virinchi
KLA Tencor
SAINT-GOBAIN
The Jackson Laboratory
BUDDIHEALTH
GENOME 10K
thsti Translational Health Science and Technology Institute
Northwestern University
IBM Research
Government of India Ministry of Statistics and Programme Implementation
Ashok Leyland
itihaasa Research and Digital
DEFENCE RESEARCH & DEVELOPMENT ORGANISATION
SERB SPECIALTY PHARMACEUTICALS
Google
SANKARA NETHRALAYA
freshworks
WADHWANI AI AI FOR SOCIAL GOOD
NVIDIA
Mindtree A Larsen & Toubro Group Company
websciencе TRUST

Socially Relevant Projects

RBCDSAI IITM

Government of India
Ministry of Statistics and Programme Implementation

Capacity building

R & D

Committee participation

108 Ambulances

Data Analytics based performance Evaluation, Location, Dispatch and Routing

Google AI

armman
Delivering India's Future

ARMMAN

GenomeINDIA
'Cataloguing the Genetic Variation in Indians'

Data analysts for whole genome sequencing project for 10K Indians

DEPARTMENT OF BIOTECHNOLOGY
Ministry of Science & Technology

thsti
Translational Health Science and Technology Institute

Preterm birth prediction in GARBH-Ini cohort

Grand Challenges India

BILL & MELINDA GATES foundation

Research & Education

Industry Conclave

Quarterly workshops

Summer/winter schools and training programs

Five-year dual degree program in data science at IIT Madras

Certificate in Technology and Management jointly with IIMBx

Outreach & Scholarships

- Postdoctoral fellowship
- Post-Baccalaureate fellowship
- Research Travel Fellowships
- Associate Researcher Program
- Research Experience Program
- Researcher Program

Thank You

RBC DSAI

ROBERT BOSCH CENTRE FOR DATA SCIENCE AND ARTIFICIAL INTELLIGENCE IIT MADRAS

5th Floor, Block II,
Bhupat and Jyoti Mehta School of Biosciences,
Indian Institute of Technology Madras, Chennai, India

https://rbc-dsai.iitm.ac.in/

RBCDSAI IITM