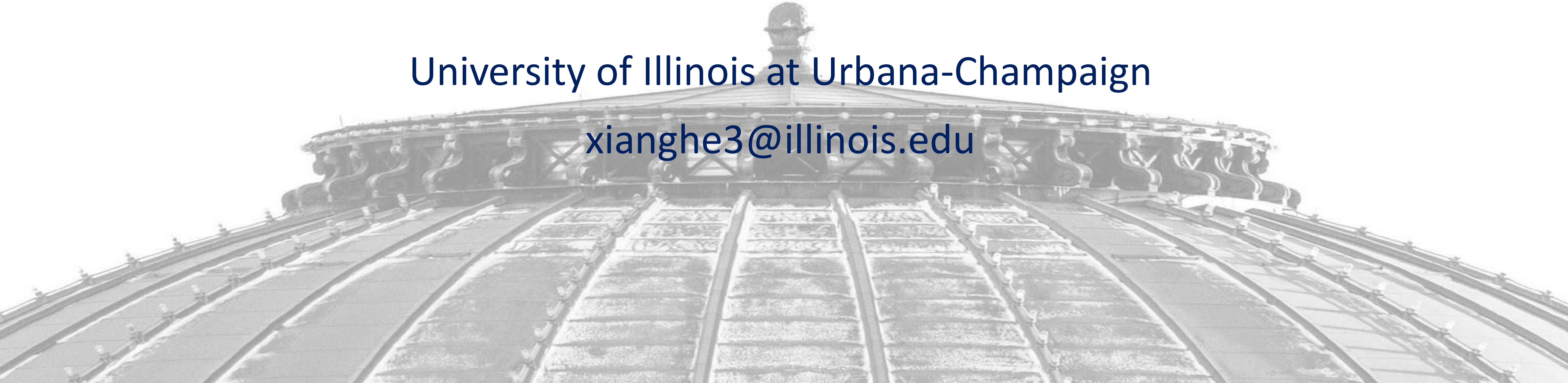
Fairness Contrastive learning on Graphs

Xianghe Xu
Graduate Mentor: Jian Kang
Faculty Mentor: Hanghang Tong

University of Illinois at Urbana-Champaign

xianghe3@illinois.edu



Content

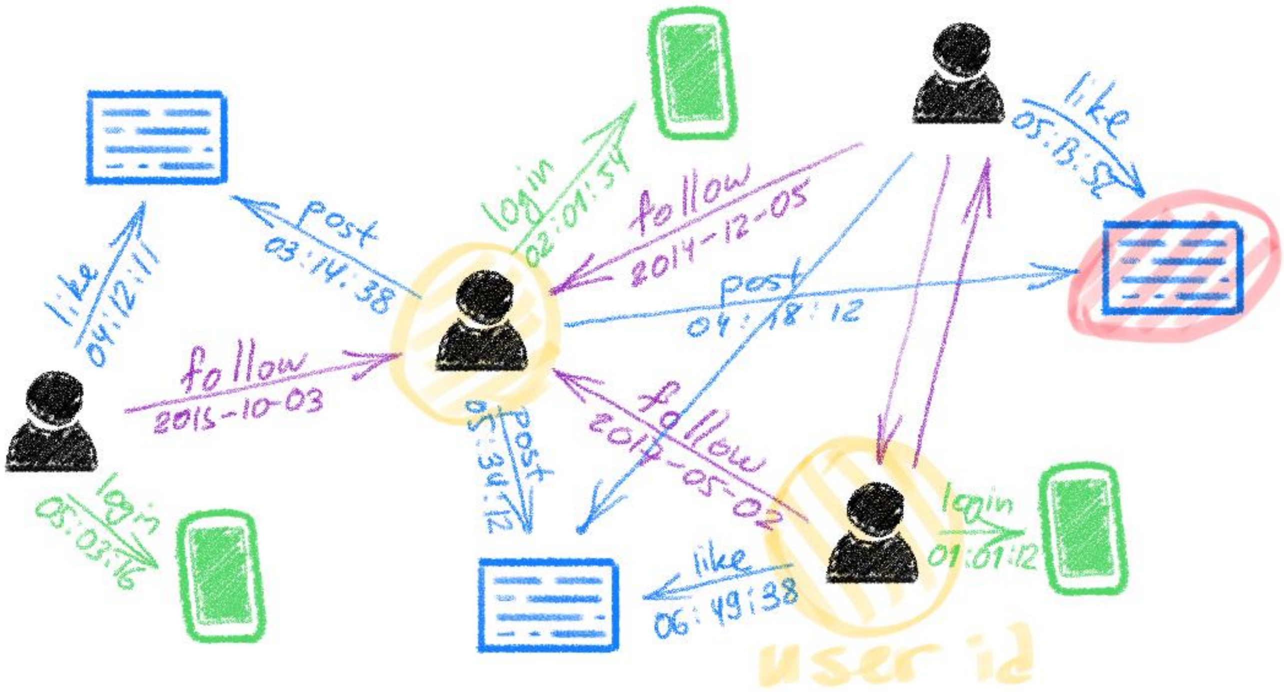
- Part1 Background:
  - Data mining on graphs
  - Fairness in machine learning
  - Contrastive learning
- Part2 Problem Setting
- Part3 Experiment
- Part4 Conclusion and Future work

Part1 Background

Data mining on graphs

- Graph is a kind of complicated data with complex relationships and interdependency between objects and can reflect real-life situation

- e.g. social networks

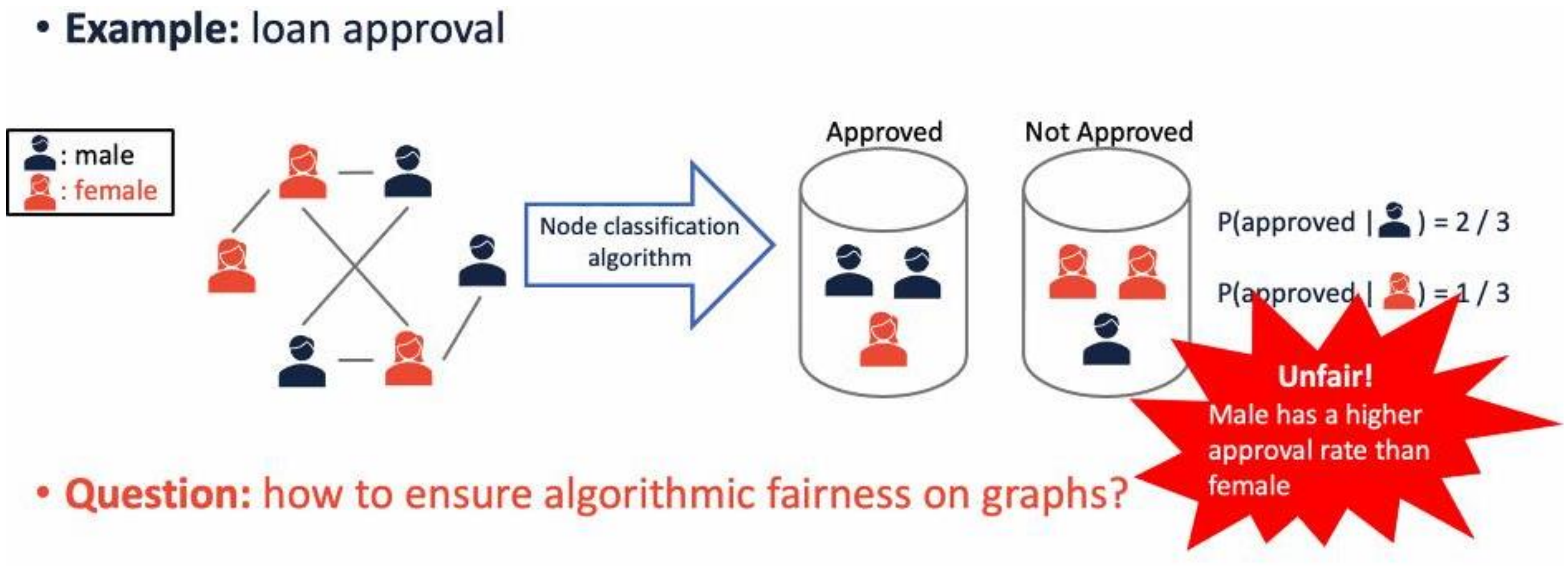


- Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, Philip S. Yu: A Comprehensive Survey on Graph Neural Networks
- Michael Bronstein: Deep learning on graphs: successes, challenges, and next steps

Part1 Background

Fairness on graph learning

- Definition: Lack of favoritism from one side or another



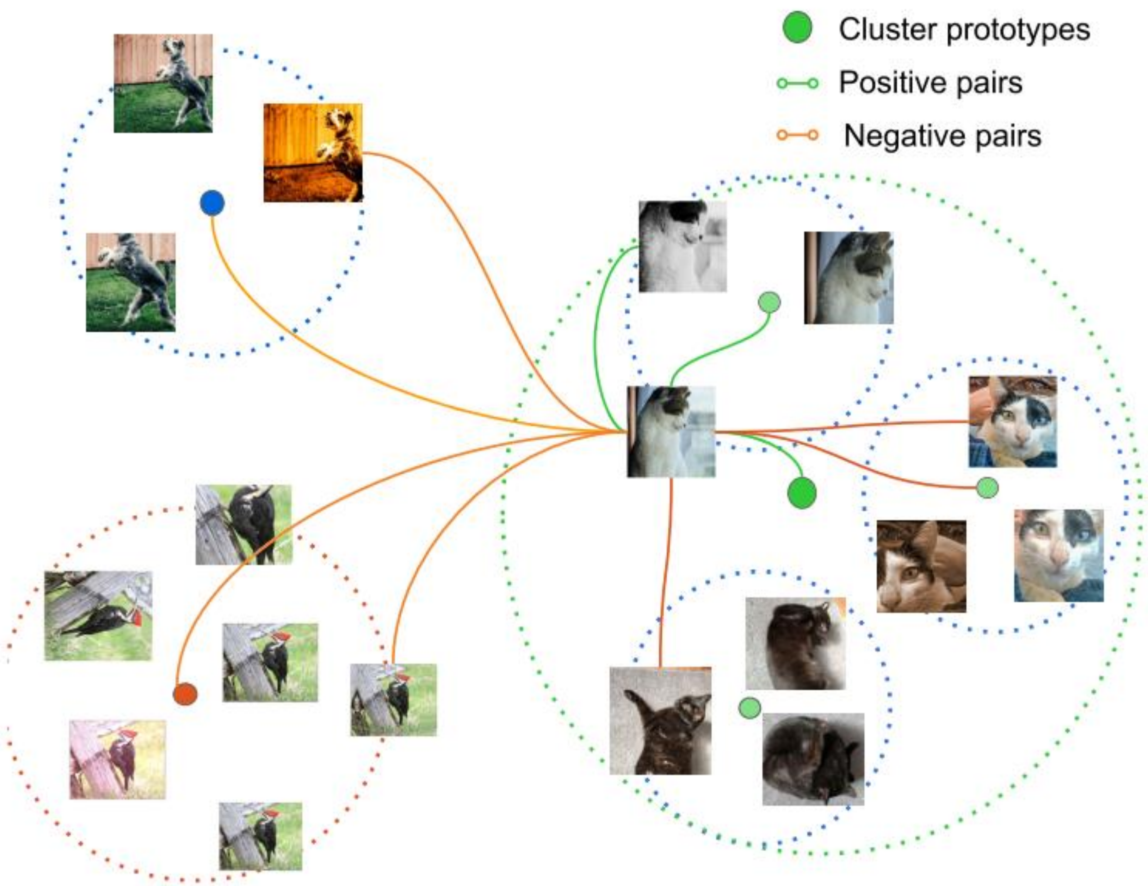
One practical problem: Balance utility and fairness

- Jian, K, Hanghang Tong: Fair Graph Mining
- Fairness in graph mining: a survey

Part1 Background

Contrastive learning

- Goal: learning an embedding space to make similar sample pairs close to each other and dissimilar pairs are far apart

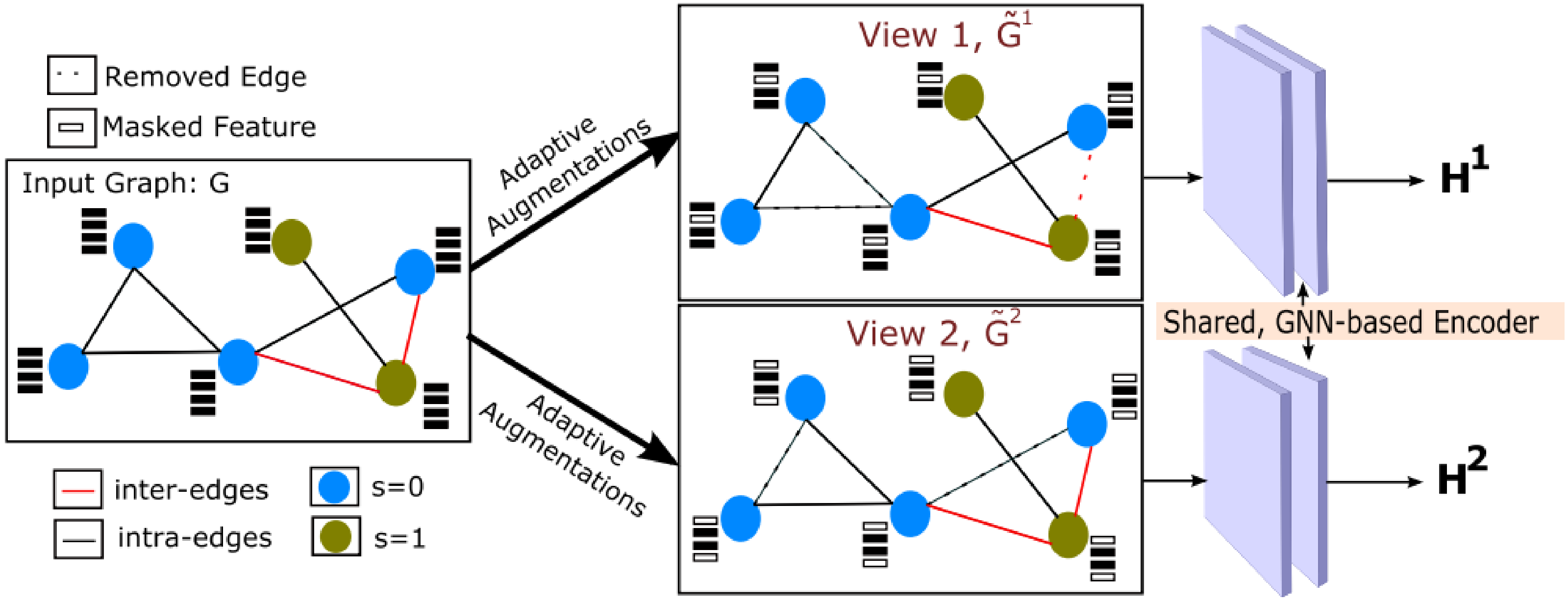


- Contrastive Representation Learning https://lilianweng.github.io/posts/2021-05-31-contrastive/
- PHUC H. LE-KHAC, GRAHAM HEALY, ALAN F. SMEATON: Contrastive Representation Learning: A Framework and Review

Part2 Problem setting

- **The general problem:**

Whether there exists fairness issue in contrastive learning algorithms on graphs and how should we alleviate the issue if it exists?



Part3 Experiment

- Method:
- Fair sampling:

Balance the number of sensitive samples and insensitive samples when using the objective function of CL algorithm

Number of sensitive samples: n1

Number of insensitive samples : n2

p = n1/n2

pi ~ Bern(1 - p)

pi is probability of whether dropping ith node in insensitive groups when calculating objective functions in contrastive learning

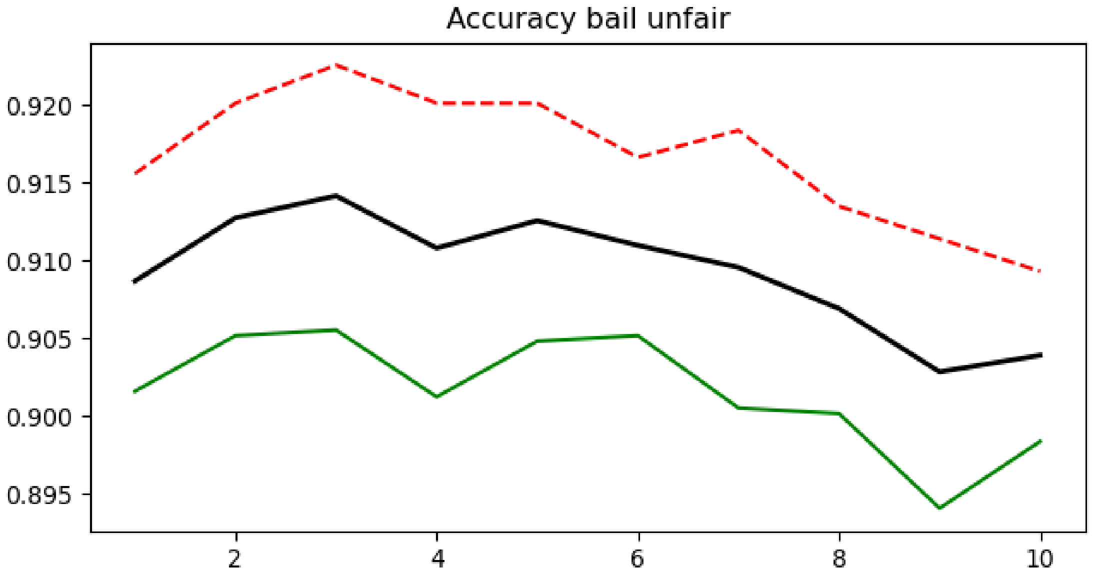
Part3 Method

Experiment

- Dataset:

Cora, Pokec-c, pokec-z, credit, bail

- Result:



	pokec fair	pokec unfair	pokec2 fair	pokec2 unfair	bail fair	bail
total ACC	0.63561	0.64562	0.62695	0.63019	0.90233	0.90215
sensitive group ACC	0.63230	0.64124	0.62789	0.62871	0.89553	0.89159
insensitive group ACC	0.64133	0.65321	0.62519	0.63297	0.90896	0.91245
gap	-0.00903	-0.01197	0.00269	-0.00426	-0.01344	-0.02086
	0.00294		0.00695		0.00742	

Part4 Conclusion and Future work

Conclusion

- Problem: unbalanced sample numbers of different groups generates fairness problem
- Method: Balance the number of sensitive samples and insensitive samples when calculating the objective function
- Result: There does exist accuracy gap between sensitive groups and insensitive groups but the

Future work

- Analysis the reason of performance difference of GCA on different datasets
- Processing the features that have high correlation to the sensitive feature

Thank you