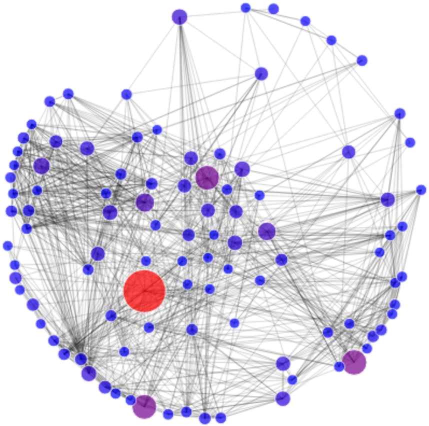
Los Alamos NATIONAL LABORATORY

The Automated Design of Network Graph Algorithms with Applications in Cybersecurity

Aaron Scott Pope, Ph.D.
A-4: Advanced Research in Cyber Systems
apope@lanl.gov

LA-UR-24-23429

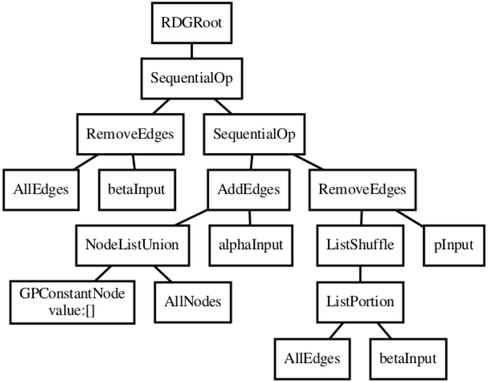
Automated Design of Network Algorithms



Use automated heuristic search techniques to improve off-the-shelf algorithm performance for specific applications.

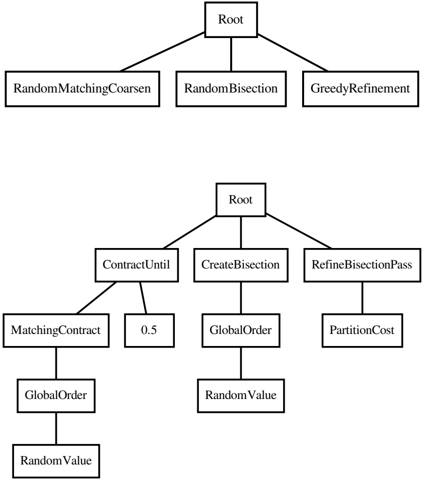
- Complex network applications typically rely on approximation heuristics for efficiency
- These heuristics can be tailored to leverage problem characteristics for an application to improve accuracy, speed, etc.
- Doing this manually can be expensive and time-consuming
- The optimization can be automated using bio-inspired search techniques

Automated Heuristic Optimization



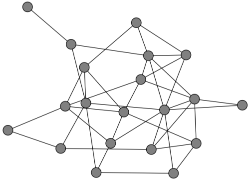
- Extract functionality from related algorithms to build a set of "algorithmic primitives"
- Construct entire algorithms from primitives (e.g., parse tree)
- Measure algorithm quality based on the application
- Use heuristic search algorithm (e.g., genetic programming) to optimize algorithm structure

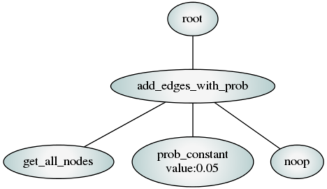
Heuristic Search Scalability for Real-world Applications



- Granularity level of primitive operations has a huge impact on scalability
- Automated primitive granularity control can help address scaling issues for heuristic searches on complex real-world problems

Application: Data-Driven Network Model Generation

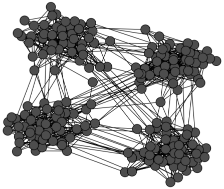




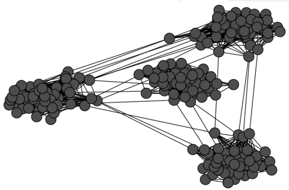
- Automate the design of algorithms for generating random networks with characteristics of interest
  - Investigate network properties
  - Make predictions
  - Generate synthetic data
- Can be trained on a single or multi-objective definition of graph quality:
  - Similarity to sample networks
  - Graph or application specific metrics

Static Modeling: Reproducing Random Community Graphs

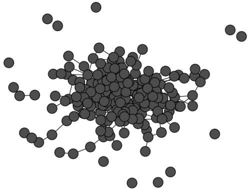
Actual Graph



Granular Model Generation

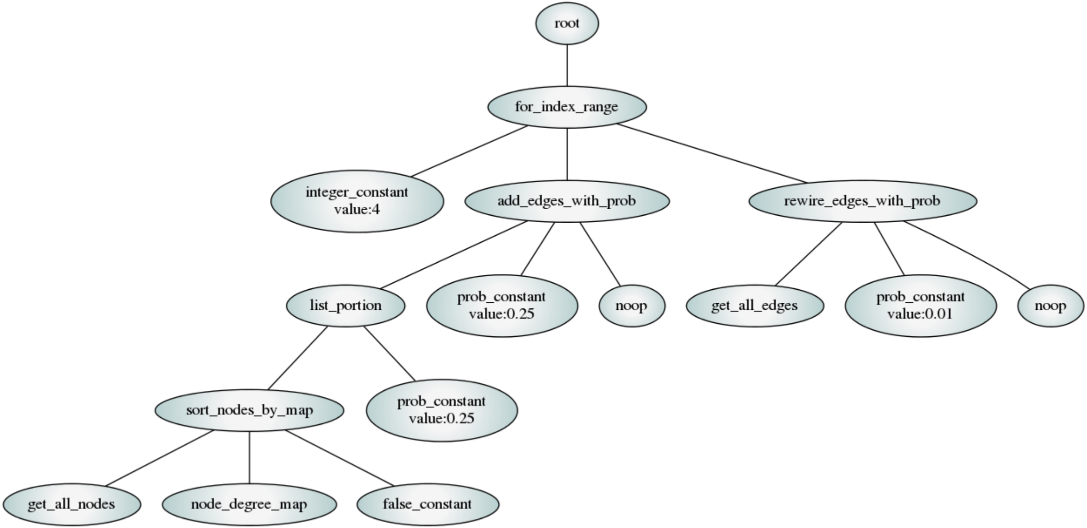


Naive Model Fitting



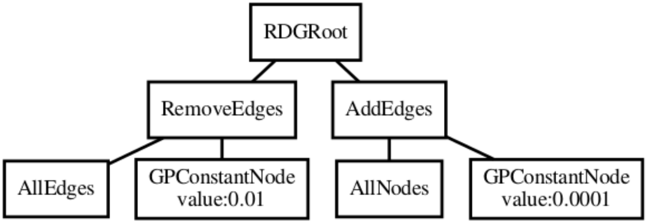
Similarity	Granular			Naive	
Metric	Mean	σ	Comparison	Mean	σ
Degree	0.436	0.075	$<$	0.458	0.055
Betweenness	0.209	0.105	$<$	0.320	0.126
PageRank	0.127	0.029	$<$	0.150	0.036

Static Modeling: Random Community Network Generator

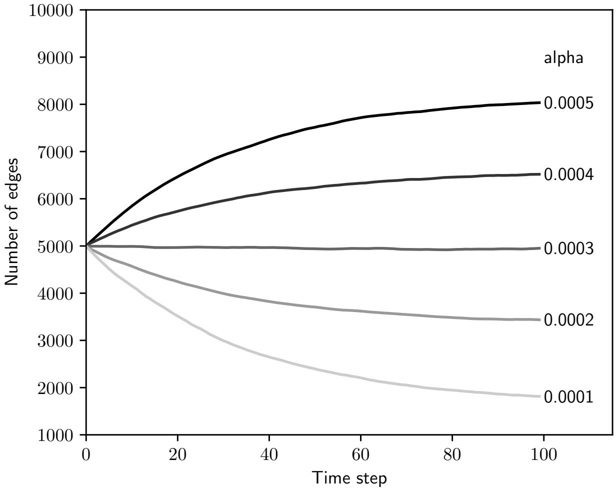


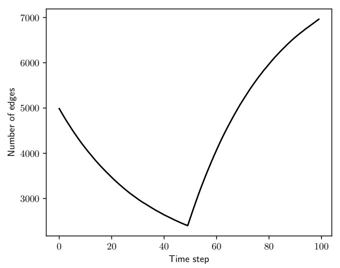
Application: Data-Driven Dynamic Network Modeling

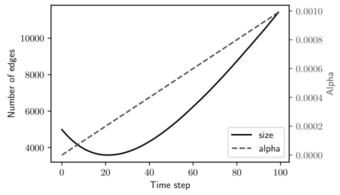
- Extends model generation to dynamic networks
- Generate algorithm that "updates" the network at each time step
- Learn to mimic target network behavior



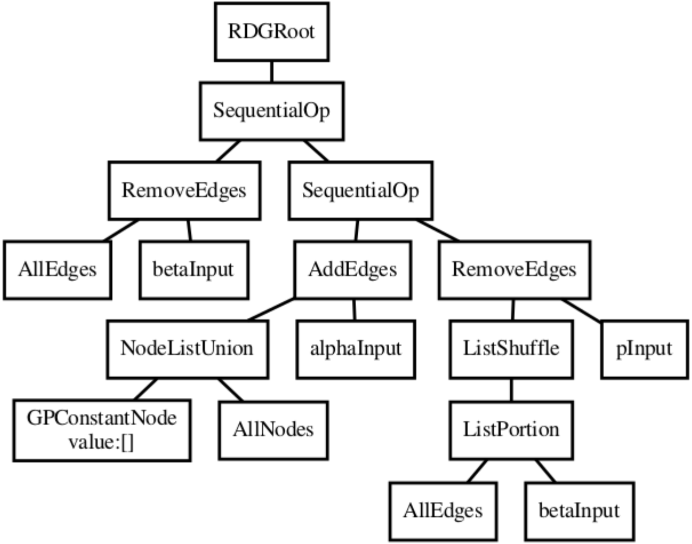
Dynamic Modeling: Dynamic Erdös-Rényi Model







Dynamic Modeling: Example Generated Algorithm

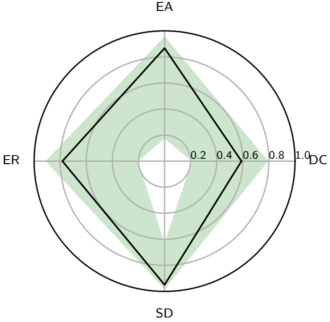


Dynamic Modeling: Real-World Enterprise Network Behavior

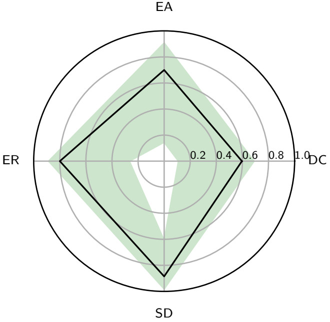
Model activity of Los Alamos National Laboratory (LANL) enterprise computer network

- User-computer authentication events
- NetFlow communication sessions between pairs of computers

Authentication

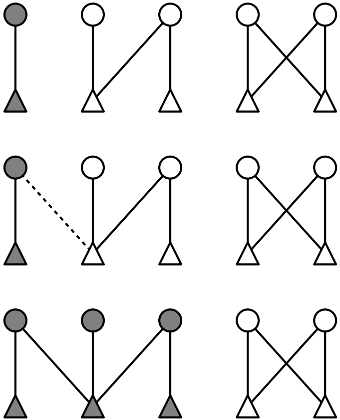


NetFlow



Application: Automated Network Security Metric Design

Attack Simulation



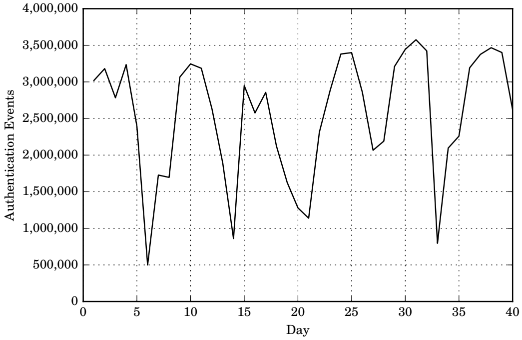
- Automated the design of network security metrics for large networks
- Trained on real or simulated event data
- Simulated attacks using real LANL network data

LANL Authentication Dataset Details

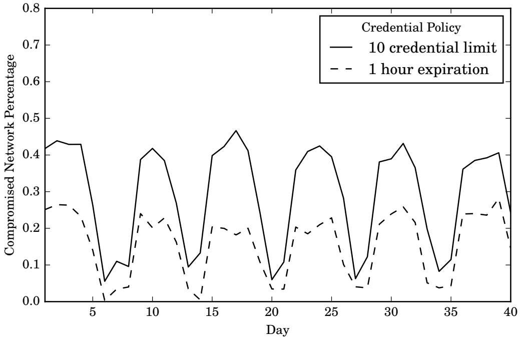
Unique Users	10,044
Unique Computers	15,779
Unique (User, Computer) Pairs	124,020
Total Authentication Events	101,918,344
Average Daily Authentication Events	2,547,959

Application: Automated Network Security Metric Design

Daily Authentication Events

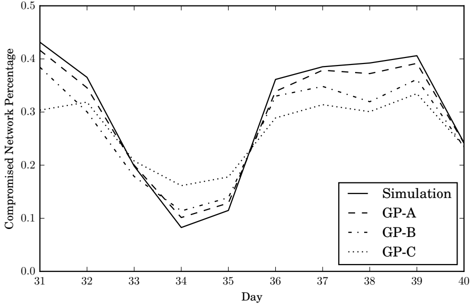


Simulation Results

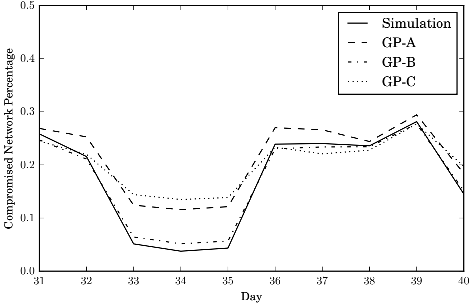


Application: Automated Network Security Metric Design

10 credential limit policy



1 hour expiration policy



Application: Tailored Anomaly Detection Heuristics

- Automated the design of novel link prediction heuristics for anomaly detection
- Link prediction: predict the existence of a relationship or rank relationships by likelihood
- Relies on historical or contextual information
- Predictive performance can be optimized by tailoring for an application

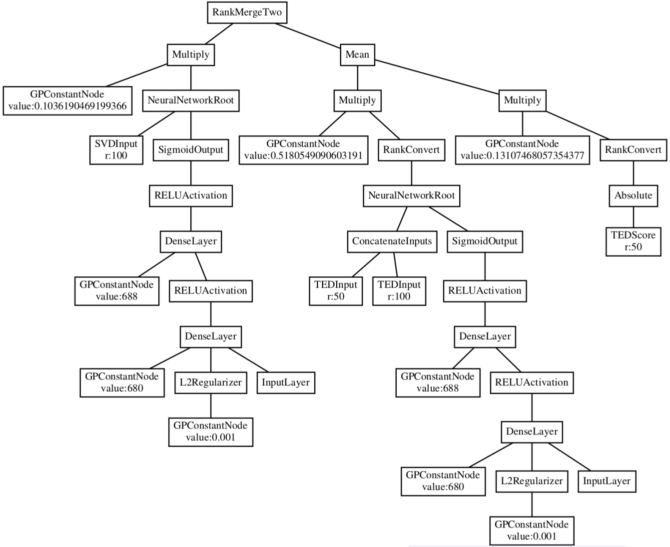
Tailored Link Prediction Heuristics: Experiment

- Data from the network at Los Alamos National Laboratory
  - User-Process (UP), Computer-Process (CP), NetFlow (NF)
- Differentiate legitimate activity from anomalies
  - Positive "new" links
  - Randomly generated negative links
- Use heuristic to calculate scores for a set of input links
- Fitness: area under ROC curve (AUC)
- AUC $\in [0, 1]$, maximized when positive and negative samples are clearly differentiated by scores

Results

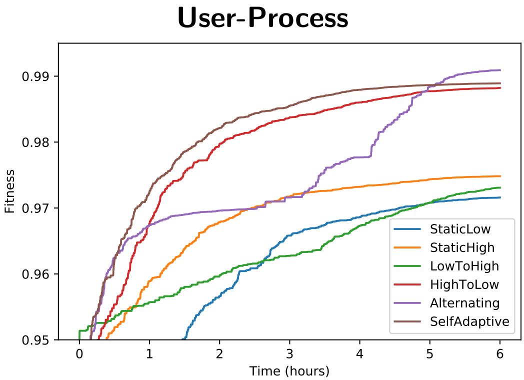
Method	Application		
	UP	CP	NF
NP	0.76963	0.74226	0.52967
TSVD	0.94186	0.90334	0.92936
TED	0.97478	0.97697	0.92390
NN	0.98725	0.98661	**0.98836**
GP-UP	**0.99066**	**0.98718**	0.98051
GP-CP	**0.98897**	**0.98996**	**0.99090**
GP-NF	**0.98867**	**0.98874**	**0.99241**

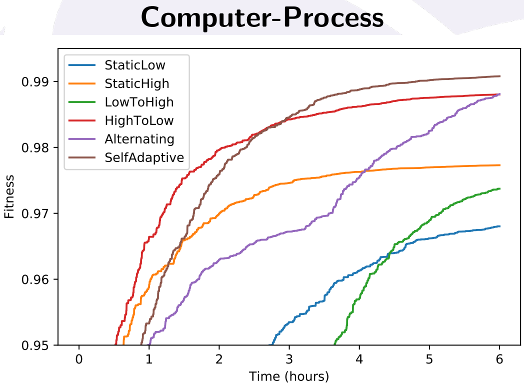
Tailored Link Prediction: Generated Heuristic



Los Alamos NATIONAL LABORATORY

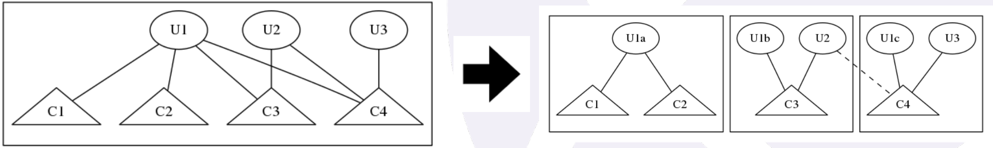
Tailored Link Prediction: Dynamic Granularity Control





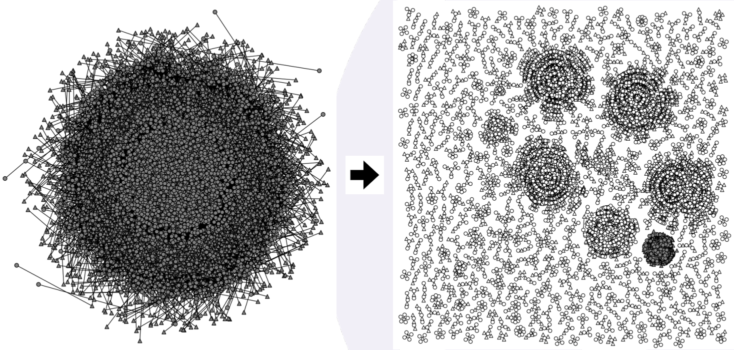
Application: Network Segmentation Algorithms

- Automate the segmentation of a network to limit adversarial traversal using stolen credentials
- Reduce the size of connected components within the network by:
  - Revoking a user's access to a computer to remove a path
  - Split a user into multiple accounts (different credentials)
- Minimize changes to reduce impact on user productivity



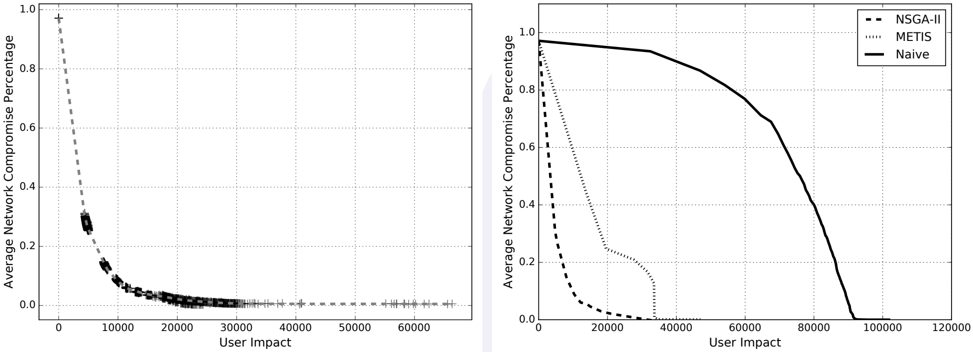
Application: Network Segmentation Algorithms

Segmenting LANL network bipartite authentication graph (BAG)



Application: Network Segmentation Algorithms

BAG Partitioning Results

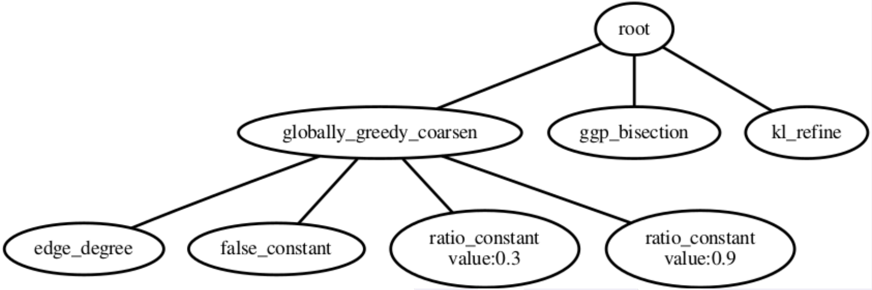


- 1-2 orders of magnitude lower user impact compared to traditional graph partitioning
- Significant reduction in network vulnerability to intrusion

Los Alamos
NATIONAL LABORATORY

Application: Design of Network Segmentation Algorithms

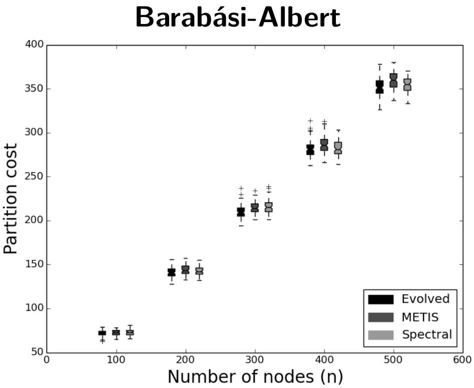
Leverage heuristic search to automate the design and optimization of multi-level graph partitioning algorithms that are tailored to specific applications

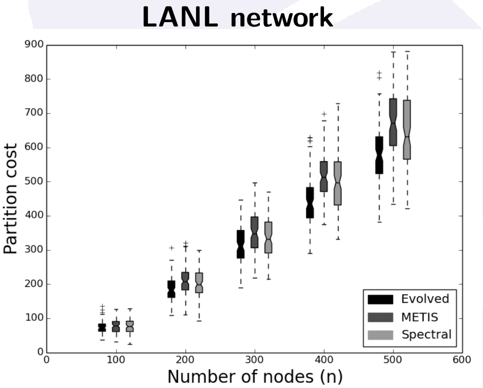


Application: Design of Network Segmentation Algorithms

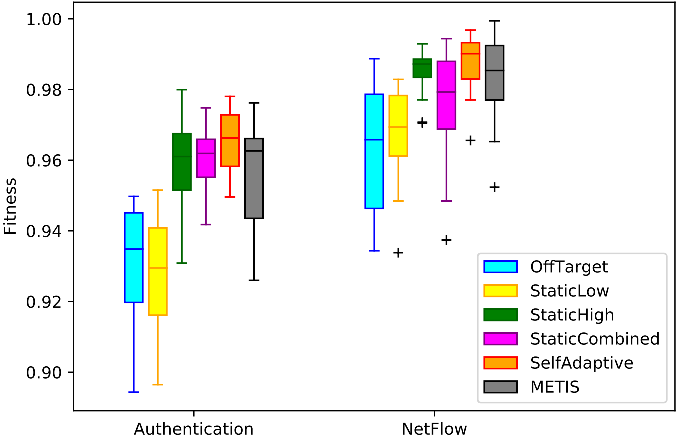
Target graph classes:

- Random graph models (Erdös-Rényi and Barabási-Albert)
- Los Alamos National Laboratory (LANL) authentication graphs



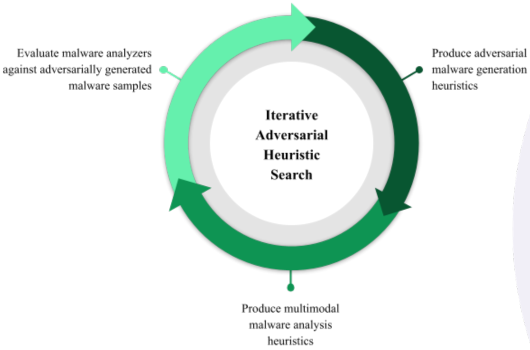


Network Segmentation: Dynamic Granularity Control



Authentication	
OffTarget	0.93089
StaticLow	0.92903
StaticHigh	0.95844
StaticCombined	0.96001
SelfAdaptive	**0.96510**
METIS	0.95539
NetFlow	
OffTarget	0.96292
StaticLow	0.96839
StaticHigh	0.98556
StaticCombined	0.97700
SelfAdaptive	**0.98787**
METIS	0.98374

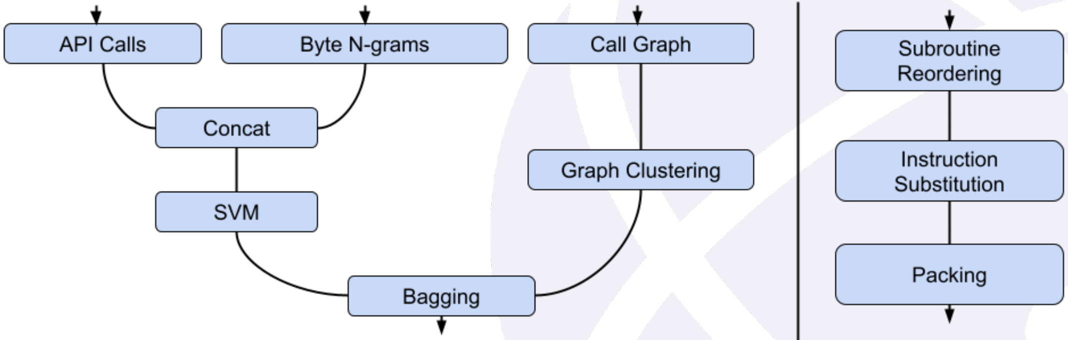
Proposed Work: Automated Algorithm Design for Adversarial Malware Analysis



Design and optimize novel algorithms for detecting and classifying malicious software

- Machine-learning based malware analyzers can be easy to defeat with simple obfuscation methods
- Automate the design of both malware analyzers and adversarial malware generators
- Use competitive co-evolution to train robust malware classifiers

Proposed Work: Automated Algorithm Design for Adversarial Malware Analysis



Summary

Bio-inspired heuristic search techniques can be used to automate the design and optimization of application-tailored algorithms. Demonstrated on:

- Complex network modeling, both static and dynamic
- Network segmentation
- Anomaly detection using link prediction
- Novel network security metrics
- Co-evolving attacker and defender strategies
- Proposed: Adversarial malware analysis

Questions?

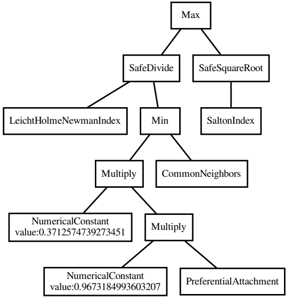
Dynamic Primitive Granularity Control: Motivation

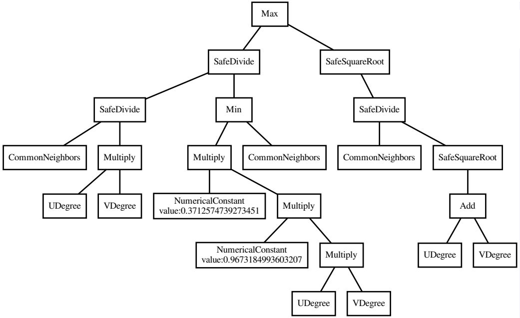
- Conventionally, primitive operation set is decided a priori
- Proper construction of set is crucial to heuristic search
- Functionality can be implemented at different levels of abstraction or granularity
- Complex, high-level operations:
  - Leverage more domain knowledge
  - Improve early results
  - Limit search flexibility to fine-tune
- Basic, low-level operations:
  - Allow greater algorithmic expressiveness
  - Dramatically increase search space
  - Requires “reinventing the wheel”

Dynamic Primitive Granularity Control: Approach

- Implement operations at multiple granularity levels
- Construct high-level “macro” primitives from basic operations
- Granularity level can be set dynamically throughout search
- Controls operations available to variation mechanics
- Macro primitives can be decomposed into basic components

Dynamic Primitive Granularity Control: Example



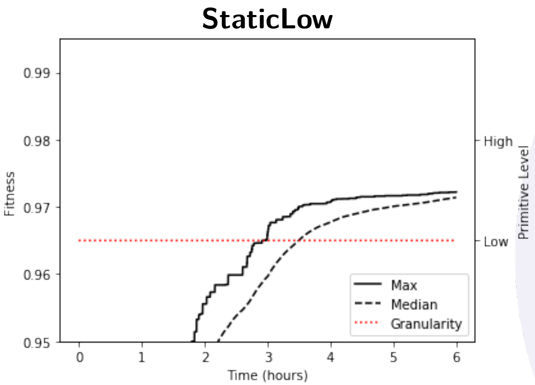


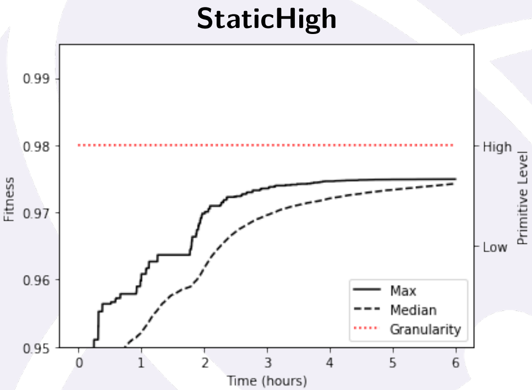
Dynamic Primitive Granularity Control

Dynamic Granularity Control Schemes:

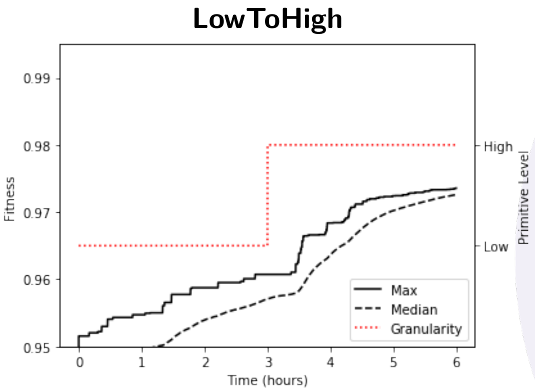
StaticLow:	low throughout evolution
StaticHigh:	high throughout evolution
LowToHigh:	low initially, change to high at midpoint
HighToLow:	high initially, change to low at midpoint
Alternating:	random initially, alternate on convergence
SelfAdaptive:	self-adaptive granularity level

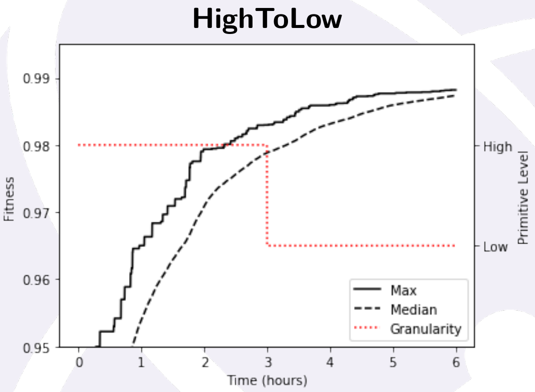
Tailored Link Prediction: Dynamic Granularity Control



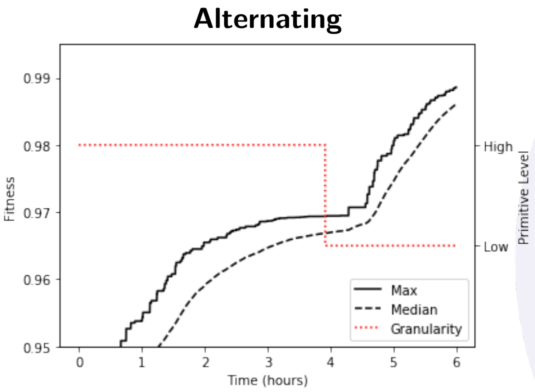


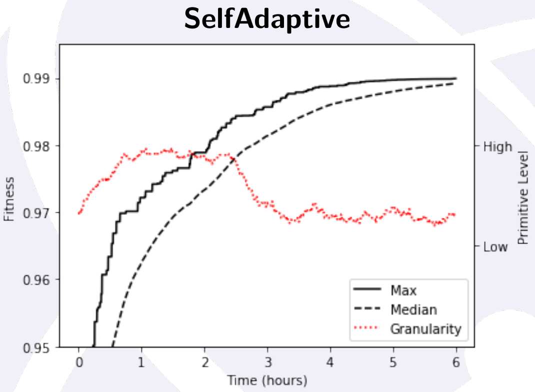
Tailored Link Prediction: Dynamic Granularity Control





Tailored Link Prediction: Dynamic Granularity Control



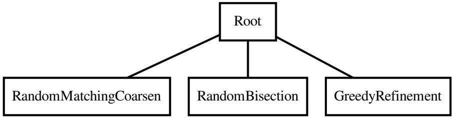


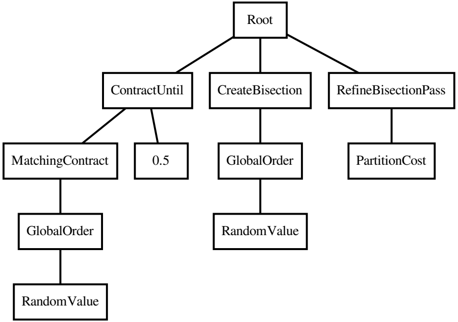
Tailored Link Prediction: Dynamic Granularity Control

	Method	Application		
		UP	CP	NF
Link Prediction Accuracy	**Ensemble**	0.98757	0.98734	0.9884
	Best-UP	——	0.97995	0.98133
	Best-CP	0.98277	——	0.97816
	Best-NF	0.98518	0.98098	——
	StaticLow	0.97269	0.97005	0.9296
	StaticHigh	0.975	0.97748	0.94082
	LowToHigh	0.97428	0.97625	0.95065
	HighToLow	0.98863	0.98835	0.9895
	Alternating	**0.9911**	**0.99019**	0.98343
	SelfAdaptive	0.98906	**0.99106**	**0.99285**

Self-Adaptive Granularity Control for Network Segmentation

- Evolution of MLP heuristics can be improved using dynamic primitive granularity control
- Leverage self-adaptive control scheme
- Target real-world networks for improving security through segmentation





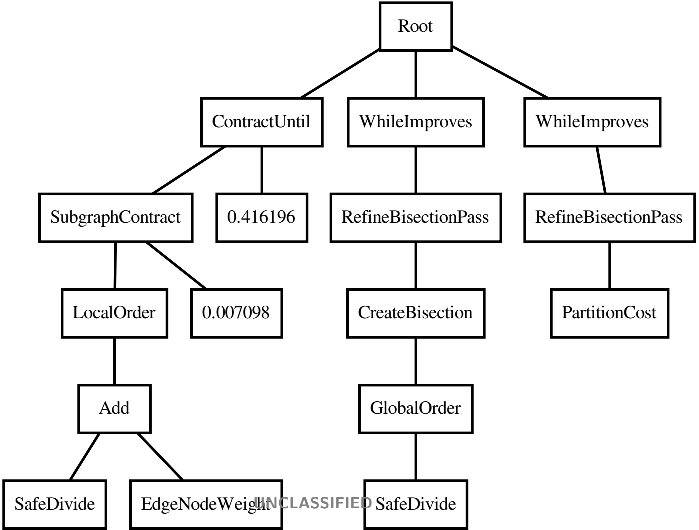
Self-Adaptive Granularity Control for Network Segmentation

Authentication	
Unique users	9,924
Unique computers	14,822
Unique user-computer pairs	106,693
NetFlow	
Unique devices	60,185
Unique communication pairs	1,136,854

- Segmenting **Authentication** graphs revokes user-computer access to limit traversal of insider or intruder with stolen credentials
- Segmenting **NetFlow** graphs identifies low-cost plans for separating network domains or placing intrusion detection monitors

Self-Adaptive Granularity Control for Network Segmentation

Example Heuristic Evolved for NetFlow Application



Los Alamos NATIONAL LABORATORY