CERIAS

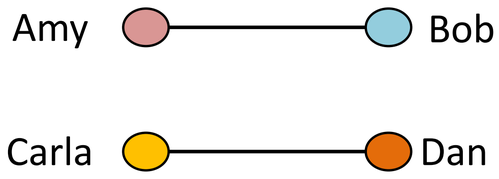
The Center for Education and Research in Information Assurance and Security

Generalized Network Privacy

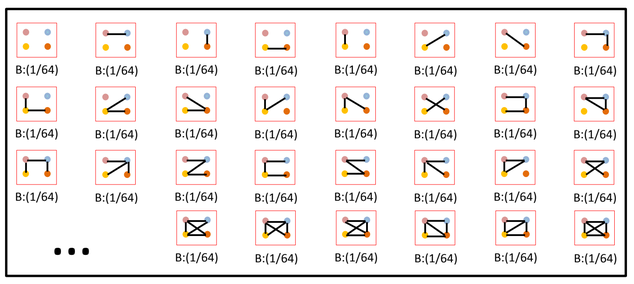
Objective: Define a universal platform for quantitatively understanding and comparing the behavior of a wide variety of graph privatization techniques.
Motivating Example: How does anonymization compare to aggregation for privacy protection?

An example: Information is collected from four individuals.

Edges are drawn in the graph to connect friends.

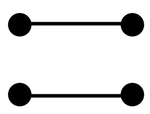


An attacker who knows nothing believes all 64 possible graphs are equally likely.



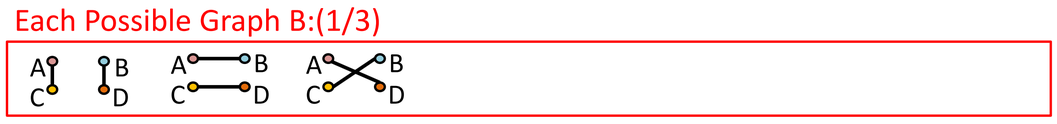
Anonymization

The easiest approach to privatization is always anonymization. A researcher might publish his data with the names removed:



The anonymized graph $\mathbf{G_{Anon}}$ is consistent with $\mathbf{4!/|Aut(G_{anon})| = 24/8 = 3}$ possible graphs over the individuals, where **Aut(G)** is the set of all automorphisms of **G**. An attacker with no prior knowledge will believe these graphs are equally likely.

Each Possible Graph B:(1/3)

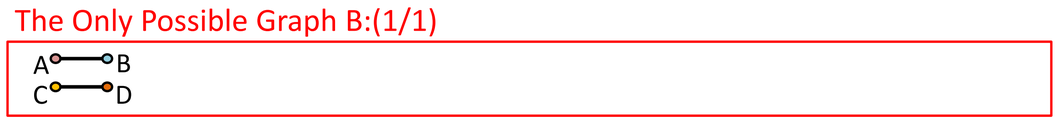


Edge Ambiguity: Each possible friendship appears in 1/3 of the possible graph set, so the attacker cannot make a strong guess about the existence of any particular friendship.

We examine the resilience of a privatization technique by considering how it performs when the attacker has some background knowledge.

If the attacker knows that Amy is friends with Bob, then there is only a single graph consistent with the attacker's total knowledge:

The Only Possible Graph B:(1/1)



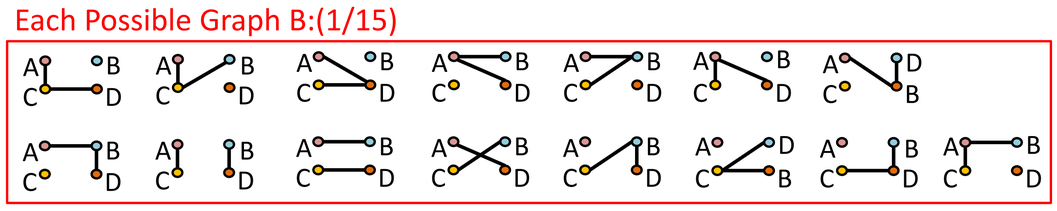
Edge Ambiguity: Anonymity is not as resilient to attacker knowledge as Aggregation. There is no ambiguity, all edges are known, the attacker knows the true graph.

Aggregation

Rather than publicly release the graph itself, a researcher might publish meaningful statistics about the graph, such as its edge count, triangle count or degree distribution.

The ***aggregate statistic*** that: **Graph G has 2 friendships** is consistent with **15** possible graphs over the individuals. An attacker with no prior knowledge will believe these graphs are equally likely:

Each Possible Graph B:(1/15)

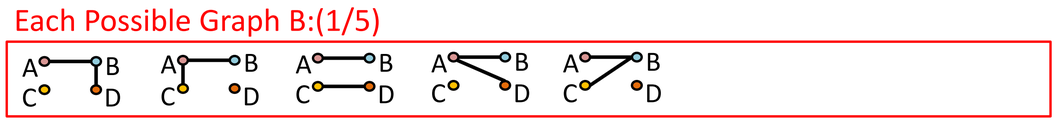


Edge Ambiguity: Each possible friendship appears in 5/15 = 1/3 of the possible graph set; The attacker cannot make a strong guess about the existence of any particular friendship.

We examine the resilience of a privatization technique by considering how it performs when the attacker has some background knowledge.

If the attacker knows that Amy is friends with Bob, then there are 5 graphs consistent with the attacker's total knowledge. With no other information these graphs are considered to be equally likely.

Each Possible Graph B:(1/5)



Edge Ambiguity: Each (unknown) possible friendship appears in 1/5 possible graphs. The attacker can not make a strong guess about the existence of any particular (unknown) friendship.

Future Work:

- Randomized Structural Noise
- Differential Privacy
- Combined Techniques
- Alternate Aggregation Statistics
- Approximation Algorithms

CERIAS

PURDUE UNIVERSITY