Publishing Attributed Social Graphs with Formal Privacy Guarantees

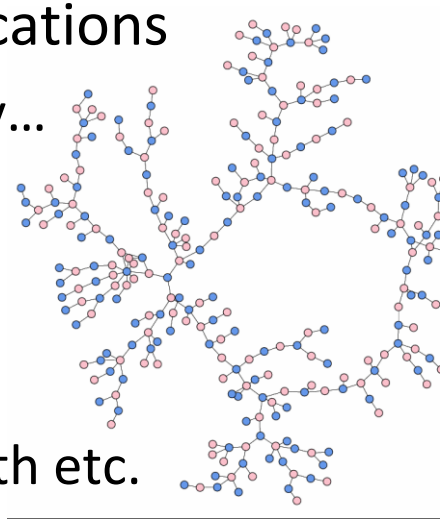
Zach Jorgensen

Graham Cormode

g.cormode@warwick.ac.uk

Ting Yu

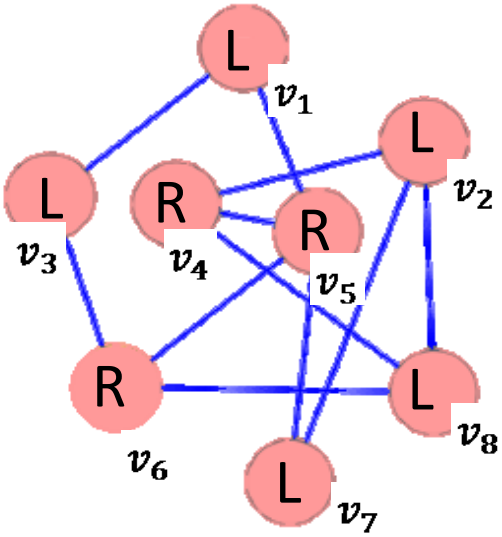
Releasing Attributed Graph Data



- Social Network Analysis has a wide range of applications
  - Marketing, disease transmission analysis, sociology…
- Real graphs (e.g. social networks) have attributes
  - Different types of node, different types of edge
- Information in social graphs is very sensitive
  - Religious, political, sexual, financial, personal, health etc.
  - We want realistic social graph data with privacy guarantees
- Prior work releases core statistics under (differential) privacy
  - Counts of small subgraphs like stars, triangles, cliques etc.
  - These counts are parameters for graph models
  - Sensitivity of these counts is large: one edge can change a lot
- We aim to release (private, synthetic) attributed graphs

Attributed Social Graphs

- Graph represented by nodes N, edges E, and attributes X
  - For every $v_i \in N$, there is a w-dimensional attribute vector $x_i \in X$
- For simplicity, assume undirected edges, binary attributes



Example:

$w = 1$ attribute, *political views*

L = Left-wing (0) R = Right-wing (1)

$N = \{v_1, \ldots, v_9\}$

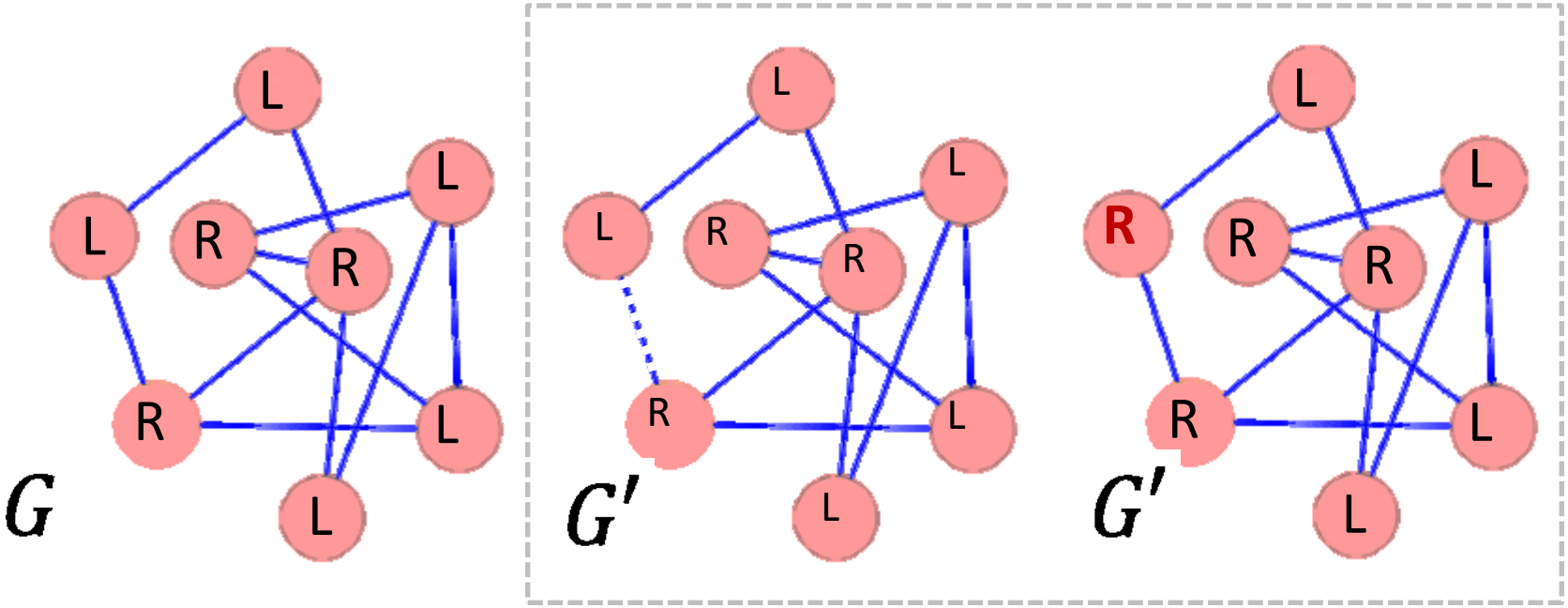
$E = \{e_{13}, e_{15}, e_{24}, e_{27}, e_{29}, \ldots\}$

$X = \{\langle 0\rangle, \langle 0\rangle, \langle 0\rangle, \langle 1\rangle, \ldots, \langle 0\rangle\}$

Privacy Model

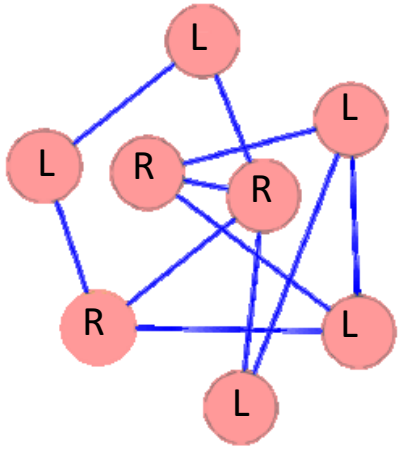
- Differential Privacy for Attributed Graphs
  - Neighboring graphs differ in the **presence of a single edge** or the **attributes associated with a single node**. [Blo13]

Two (of many) possible neighbors of *G*

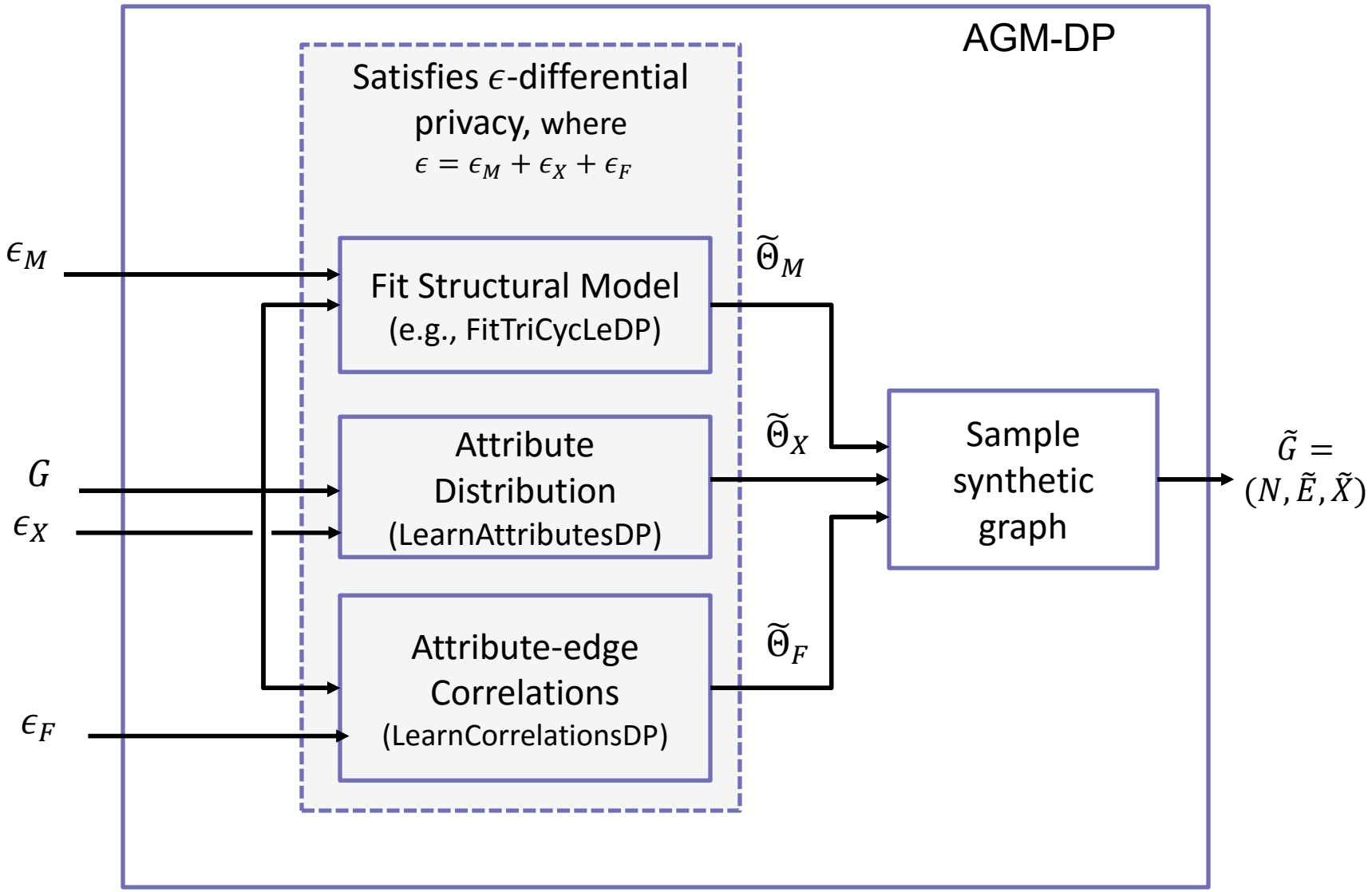


Building blocks for the private model

- **Node-attribute distribution, Θ_X**: prior distribution of attributes
  - Compute 2^w counts, add Laplace noise (histogram q
- **Attribute-Edge correlations, Θ_F**:
  probability of an edge given the two node values
  - Query has high "sensitivity" if node degrees are larg
  - Use edge truncation to bound the degree of nodes
- **Structural model for the graph edges , Θ_M**:
  - We propose a new privacy-friendly model called TriCycle
  - The parameters are the degree sequence and number of triangles
    - These can be found accurately under DP

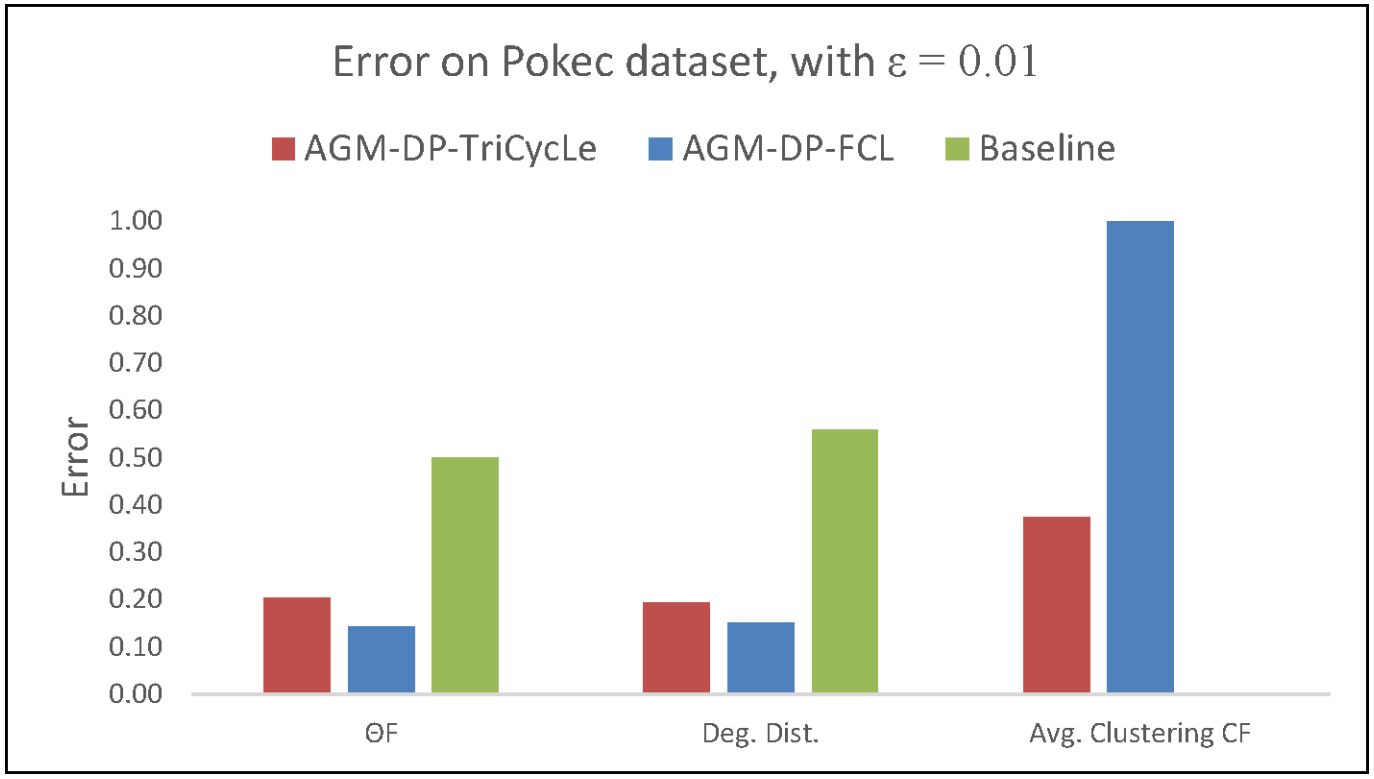


System overview

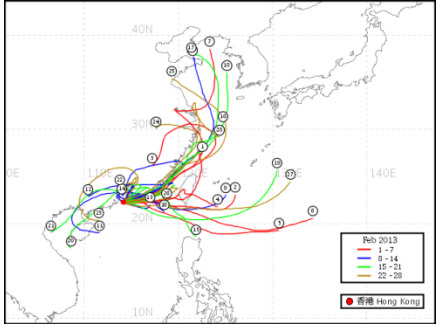


Experimental Snapshot

- Results on a large social network with strong privacy (ε=0.01)
  - Measure mean absolute error for different parameters



Summary



- Important to release social graphs with privacy
  - Full paper proposes a framework for these releases
  - Can accommodate different graph and correlation models
- Experiments show good fidelity of synthetic graphs
  - Larger inputs allow better (private) estimation of parameters
- Many natural extensions to richer graph models are possible
  - E.g. include directed edges, more attribute types
- Yet stronger privacy models (e.g. node differential privacy) remain a particular challenge

Work supported by Royal Society, European Commission