Graph Perturbation as Noise Graph Addition:

A New Perspective for Graph Anonymization

Vicenç Torra, Julián Salas

Data Privacy Management
Luxembourg, 26 September 2019

Cryptography and Information Security for Open Networks

R&I
IN3
Internet Interdisciplinary Institute

CYBER[||||]CAT

Outline

1. Introduction
   - Motivations and objectives
   - Random graph models

2. Formalizing noise addition for graphs

Motivations

- *Several masking methods for graphs:*

There is a large number of adhoc methods based on removing/adding edges/nodes.

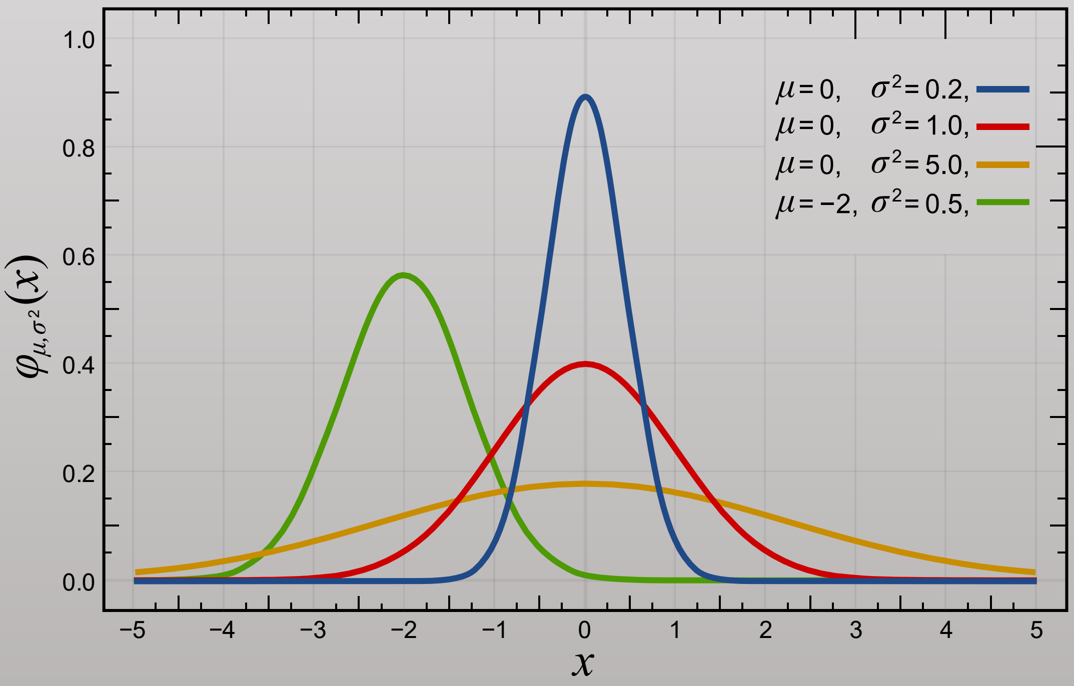
Most of them are evaluated empirically.

- *Noise addition for standard databases:*

Is a well-structured approach with a solid mathematical/statistical basis.

Noise addition

For standard databases



- Given a value x for variable V with mean μ and variance σ^2

Replace x by $x+\varepsilon$ with $\varepsilon \sim N(0, \sigma^2)$.

Privacy models

- *K-anonymity:* Modify the data so that intruders cannot find a record in the database. Protect record among k indistinguishable records.

- *Differential privacy:* Given a query, avoid disclosure from the outcome of the query. Add noise into the outcome.

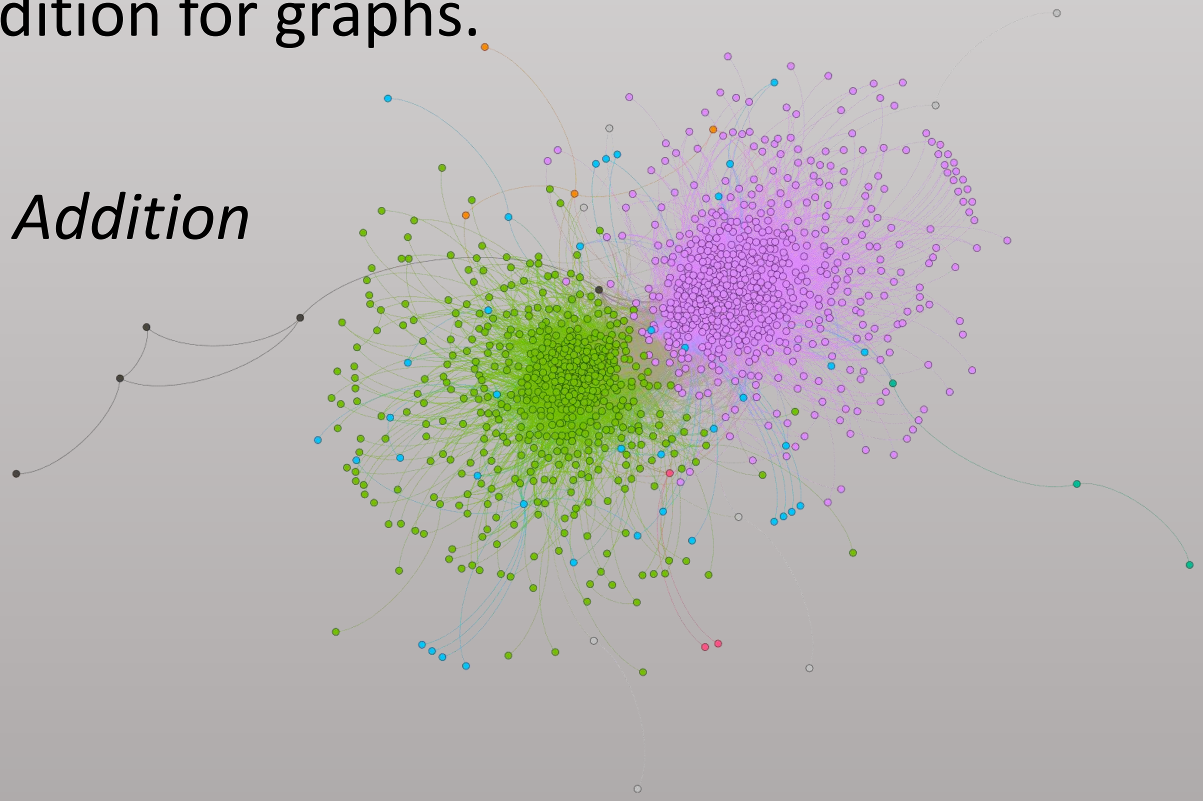
- *Protect against reidentification:* Modify the data so that intruders cannot find a record in the database. Add noise into the data.

Objective

- *Develop a sound approach for graph masking.*

Based on the analogy of noise addition for graphs.

We use *Random Graphs & Graph Addition*



Random Graphs

Basic models

- Gilbert model: $\mathcal{G}(n,p)$

n nodes and each edge is chosen with probability p.

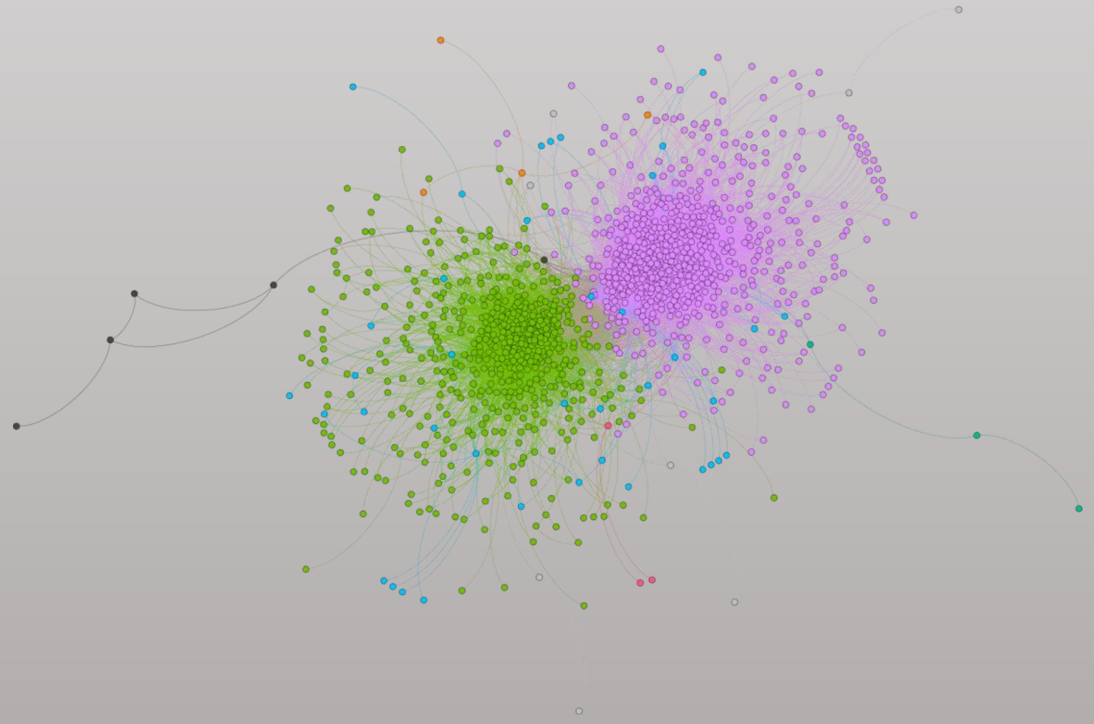
- Erdös-Renyi: $G(n,e)$

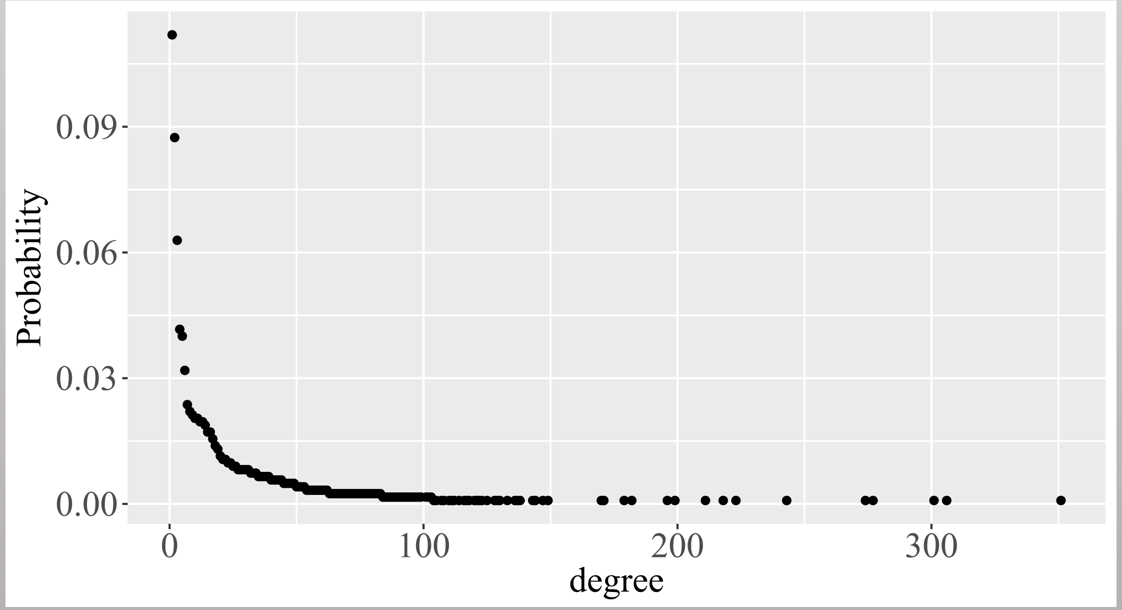
A uniform probability of all graphs with n nodes and e edges.

Both are asymptotically equivalent.

Online social networks

OSN are sparse & their degrees follow a power-law: $P(k) \sim k^{-\gamma}$





Random Graphs

Different models

- Models based on a given degree sequence. $\mathcal{D}(n, d^n)$

$\mathcal{D}(n, d^n)$ uniform probability of all graphs with n nodes, degree sequence d^n.

- Add constraints to graphs:

e.g., the degree sequence, spatial/ temporal constraints on the nodes.

Graph Addition

Formalization

Given two graphs $G_1(V, E_1)$ and $G_2(V', E_2)$ with $V \subseteq V'$; we define the addition of G_1 and G_2 as the graph $G(V', E)$ where:

$$E = \{e : e \in V \wedge e \notin V'\} \cup \{e : e \notin V \wedge e \in V'\}$$

$$G = G_1 \oplus G_2$$

Note that $\oplus$ is an *exclusive-or* of edges, most general definition is based on alignments.

Noise Graph Addition

Methods

For any graph G choose a noise-graph G' from $\mathcal{G}$ to add noise to G:

$$G \oplus G'$$

- Previous methods can be expressed in this way by adding constraints to the family of graphs $\mathcal{G}$.

Noise Graph Addition

Previous methods: examples

Changing m edges from the original graph.

Define: $\mathcal{G} = \{G' : |E(G')| = m\}$

- If we restrict $\mathcal{G}$ to be the family of graphs G such that $|E(G')| = 2m$ and $|E(G') \cap E(G)| = m$, then we are adding m edges and deleting m other edges.

Noise Graph Addition

Previous methods: examples

Random sparsification (for a probability p):

For each edge do independent Bernoulli trial. Leave the edge in case of success and remove otherwise.

Our method, use:

$\mathcal{G} = \mathcal{G}(n, 1-p) \cap G$

Add $G \oplus G'$ for some $G' \in \mathcal{G}$

Noise Graph Addition

Previous methods: examples

Degree preserving randomization

Define: $\mathcal{G} = \{G' : V(G') = i, j, k, l \subseteq V(G); ij, kl \in E(G')$ and $jk, li \notin E(G')\}$

$\mathcal{G}$ is the set of alternating 4-circuits of G.

$$G \oplus_{i=1}^{m} G'_i$$

Following this procedure for m large enough is equivalent to randomizing G to obtain all the graphs $\mathcal{D}(n, d^n)$.

Noise Graph Addition

New method

Local randomization

Define: $\mathcal{G} = \{G_u^t : V(G_u^t) = u, u_1, \ldots, u_t; E(G_u^t) = uu_1, \ldots, uu_t\}$

Then, $G \oplus G_u^t$ changes t-random edges incident to vertex $u \in V(G)$.

- So we can apply local t-randomization for all $u \in V(G)$ to obtain the graph $G^t = G \oplus_{u \in V(G)} G_u^t$

Local Randomization

Risk properties

Adversary's prior and posterior probabilities to predict whether there is a sensitive link between $i, j \in V(G)$ by exploiting the degree d_i and access to G^t

$\mathrm{P}(a_{ij} = 1)$ equals: $\frac{d_i}{n-1}$

$\mathrm{P}(a_{ij} = 1 | a_{ij}^t = 1)$ equals: $\frac{d_i(\bar{t}^2 + t^2)}{d_i(\bar{t}^2 + t^2) + 2\overline{d}_i(\bar{t}t)}$

$\mathrm{P}(a_{ij} = 1 | a_{ij}^t = 0)$ equals: $\frac{2\overline{d}_i(\bar{t}t)}{d_i(\bar{t}^2 + t^2) + 2\overline{d}_i(\bar{t}t)}$

The most general noise

From Gilbert model

Let $G_1(V, E_1)$ an arbitrary graph with $n_1 = |E_1|$ and $G_2(V, E_2)$ generated from a Gilbert model with $n_2 = |E_2|$.

Then $G = G_1 \oplus G_2$ will have on average: $\frac{n_2(t-n_1)+n_1(t-n_2)}{t}$ edges.

Where $t = |V|(|V|-1)/2$.

Summary

Different approaches

Noise addition method	Definition of $\mathcal{G}$	Additional requirements for $G' \in \mathcal{G}$	Properties of $G \oplus \mathcal{G}$
Random perturbation [20]	$\lvert E(G')\rvert = 2m$	$\lvert E(G') \cap E(G)\rvert = m$ $\lvert E(G') \cap E(\overline{G})\rvert = m$	G' adds m edges and removes m edges
Random sparsification [6]	$G' \in \mathcal{G}(n; 1-p) \cap G$	None	The edges of G remain with probability p, no added edges
Local t-randomization	$G' = G_u^t$	Applied to every node in G	Every node has t modified incident edges
Degree preserving randomization [5]	$G' \in \mathcal{S}_G$	$\mathcal{S}_G$ is the set of swaps of G	$G, G \oplus G' \in \mathcal{D}(n, d^n)$
Gilbert model	$G' \in \mathcal{G}(n; 1-p)$	None	Every edge is added or removed with probability p

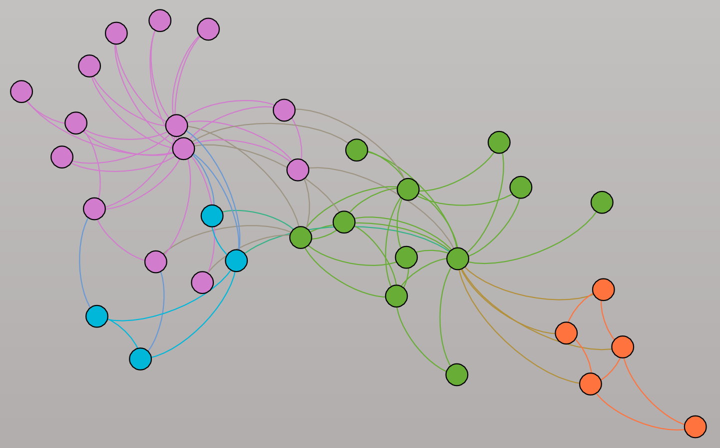
Conclusions

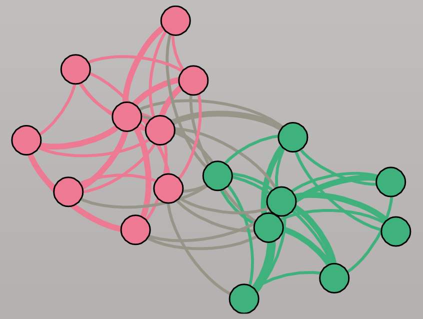
- We defined noise graph addition.

  Some existing methods can be seen from this perspective.

  Proven some properties.

- This approach permits a more systematic study of graph perturbation.





Thank you

Any questions?