PrivGraph: Differentially Private Graph Data Publication by Exploiting Community Information

<u>Quan Yuan¹</u>, Zhikun Zhang^{2,3}, Linkang Du¹, Min Chen³, Peng Cheng¹, and Mingyang Sun¹

¹Zhejiang University ²Stanford University

³CISPA Helmholtz Center For Information Security







- **D** Background
- **D** Problem Definition
- **D** Method
- **D** Evaluation
- Conclusion

D Background

- **D** Problem Definition
- **D** Method
- **D** Evaluation
- **D** Conclusion

Big Data Era

Data collection

Browsing history, communication records, ...

Data analysis

Improving user experience, recommendation, ...



Privacy Accidents



2017, Yahoo breached 3 billion user data



2021, Facebook 1.5 billion user data sold

5



2020, Microsoft exposed 250 million records



Laws for Privacy Protection

Most Recent Legislative Developments Regulation in place / due to come into force in key markets* Regulation under discussion Russia – Law on Personal China - Personal Information Data (152-FZ, 2006), Security Specification (May 2018) Japan - Act on the Switzerland - Revised Data Localisation Law and Amendments (entering into Canada - Digital Privacy Protection of Personal Data Protection Act (still (242-FZ, 2014), Delisting force in October 2020), Regulation Information (APPI) Act (November 2018), in drafting stage) Law (149-FZ, 2016) on the Protection of Children's reforming the Personal (last amended in Personal Information Online 2017). Amendments Information and Protection (October 2019), Draft Measures currently under and Electronic Documents (proposed in May and June 2019) consideration. Act (PIPEDA) EU – ePrivacy Regulation under discussion. (still in drafting stage) Korea – Personal Information Protection Act (2011), latest amendments India – Personal Data US - Competing entering into force in Nigeria – Data Protection California – Protection Bill 2019 data privacy bills August 2020. Regulation 2019 (date of Consumer Privacy Act issued in Congress (still in drafting stage) entry into force still to be (January 2020) Thailand - Personal Data determined) Protection Act (entering into force in May 2020) Brazil - General Data Singapore - Personal Data Protection Law (LGPD -Kenya – Data Law 13.709) (entering into Protection Act (PDPA) (2012) Protection Act force in August 2020) Ecuador - Data (November 2019) Protection Bill (still New Zealand - Privacy in drafting stage) Uruguay - Law on the Bill 34–2 (still in Protection of Personal Data drafting stage, entry and Habeas Data (Law into force planned for Chile – Proposal Data Uganda - Data Protection and 18.331/2008) amendments November 2020) Protection Law (still in Privacy Act (May 2019) (January 2019) and Decree Malaysia: drafting stage) (No. 64/020) Australia - Privacy Act 1988 and Personal Data amendments(last amended in Protection Act March 2014, including 13 South Africa – Protection Argentina – Proposal Data 2010 (in force of Personal Information Act Australian Privacy Principles Protection Bill (MEN-2018-147since 2013) (POPIA) (entry into force to APN-PTE) (still in drafting stage) be determined)

D Background

D Problem Definition

D Method

D Evaluation

D Conclusion



Social Contacts

Vote Network

Email Communication

The edges of a graph may contain users' sensitive information.

Related Tasks



Advertising

User Protrait

Epidemiological Study



An edge neighboring graph G'

Edge-DP: Limit the impact of any edge in the graph on the output







Our goal: Synthesize a graph under edge-DP while ensuring high utility

- **D** Background
- **D** Problem Definition
- **D** Method
- **D** Evaluation
- **D** Conclusion

Intuition



Sparse Connection



- Phase 1: Community Division (CD)
- Phase 2: Information Extraction (IE)
- > Phase 3: Graph Reconstruction (GR)



- Phase 1: Community Division (CD)
- Phase 2: Information Extraction (IE)
- > Phase 3: Graph Reconstruction (GR)



- Phase 1: Community Division (CD)
- Phase 2: Information Extraction (IE)
- > Phase 3: Graph Reconstruction (GR)



- Phase 1: Community Division (CD)
- Phase 2: Information Extraction (IE)
- Phase 3: Graph Reconstruction (GR)



Background

- **D** Problem Definition
- **D** Method
- **D** Evaluation

D Conclusion

Experiment Setup Dataset

> 6 real world datasets

Dataset	Nodes	Edges	Density	Туре
Chamelon [65]	2,277	31,421	0.01213	Web page
Facebook [46]	4,039	88,234	0.01082	Social
CA-HepPh [44]	12,008	118,521	0.00164	Collaboration
Enron [63]	33,696	180,811	0.00032	Email
Epinions [62]	75,879	405,740	0.00014	Trust
Gowalla [13]	196,591	950,327	0.00005	Social

Experiment Setup

Image: Metrics

- Community Discovery: Normalized Mutual Information
- > Node Information: Eigenvector Centrality Score
- > Degree Distribution
- Path Condition: Diameter
- Topology Structure: Clustering Coefficient, Modularity
 Competitors
 - > LDPGen^[1] > TmF^[2] > PrivHRG^[3] > DER^[4]

[1] 2017 CCS Generating Synthetic Decentralized Social Graphs with Local Differential Privacy
[2] 2015 ASONAM Differentially Private Publication of Social Graphs at Linear Cost
[3] 2015 SIGKDD Differentially Private Network Data Release via Structural Inference
[4] 2014 VLDBJ Correlated Network Data Publication via Differential Privacy

Performance



Performance



Comparison with Tailored Methods



PrivGraph achieves competitive performance on the degree distribution.

Preservation for Small Communities

Louvain adopted in phase 1 might miss the small communities^[1] during the modularity optimization process.



PrivGraph can compensate the shortcoming of Louvain since the information extraction and graph reconstruction processes help to recover the small communities.

[1] S. Fortunato and M. Barthelemy. Resolution limit in community detection. PNAS, 104(1):36–41, 2007.

- **D** Background
- **D** Problem Definition
- **D** Method
- **D** Evaluation

D Conclusion

Conclusion

- > A deep analysis of existing solutions on differentially private graph synthesis
- A practical method PrivGraph to generate a synthetic graph under DP
- An extensive evaluation on multiple datasets and metrics to illustrate the superiority of PrivGraph

Thank you for your attention Q & A

Email: yq21@zju.edu.cn