When Degree Anonymity Is Not Enough: Assessing Structural Re-identification Risk in Social Graphs

Christiano J. P. de Brito[1], André L. Vignatti[1], Sidgley C. de Andrade[2]

[1] Universidade Federal do Paraná (UFPR), Brazil
{christiano.brito, vignatti}@ufpr.br
[2] Universidade Tecnológica Federal do Paraná (UTFPR), Brazil
sidgleyandrade@utfpr.edu.br

Abstract. Removing identifiers and equalizing node degrees do not guarantee graph privacy, as network topology acts as a structural quasi-identifier. Evaluating rather than assuming privacy, this article frames structural re-identification risk assessment as a reproducible evaluation problem and presents a configuration-driven pipeline that anonymizes graphs, validates anonymity, and reports privacy-utility metrics. We adopt the structural k-anonymity paradigm as our core design choice, evaluating it on Facebook Ego-Nets and Email-Enron across $k \in \{2, 5, 10, 20\}$. Under a 1-hop subgraph adversary ($d = 1$), effective protection reduces to degree-based anonymity: the 1-hop re-identification rate exceeds the degree-based rate by approximately 6× on Facebook and 38× on Enron at $k = 2$, indicating that neighborhoods can remain structurally distinguishable. We position this assessment relative to differential privacy (a difference of premises: publishing the whole graph versus releasing queries) and to stronger learned de-anonymization adversaries. Furthermore, increasing k does not improve privacy monotonically, as aggressive anonymization distorts degree distributions and shifts vulnerability patterns. Since our scenarios model basic adversaries, these risk rates represent lower bounds, indicating that graphs anonymized by degree-based protection should undergo structural risk assessment before publication.

CCS Concepts: • **Security and privacy** → **Data anonymization and sanitization**; *Privacy protections*; • **Information systems** → **Data mining**; *Social networks*.

Keywords: data privacy, graph mining, k-anonymity, re-identification risk, social networks, structural anonymization.

1. INTRODUCTION

Removing identifiers does not anonymize a social network. Publishing its structure for research rests on that flawed premise: network topology is itself a structural quasi-identifier, and an adversary armed with fragments of the graph structure (e.g., node degrees or local neighborhood subgraphs) can re-identify nodes even in the absence of attributes [Backstrom et al. 2007; Liu and Terzi 2008; Zhou and Pei 2008]. Yet structural anonymization techniques can be trusted only as far as the adversarial model against which they are evaluated: without a rigorous, quantifiable adversary, the privacy axis of any privacy-utility trade-off has no operational meaning, since utility can be measured directly but privacy is defined only relative to a specific adversary. To ground the privacy axis in a quantifiable adversary, we adopt the structural k-anonymity model [Zhou et al. 2008] for two reasons. First, k-anonymity-based structural graph anonymization lacks a standardized, reproducible risk-assessment instrument: the gap this work addresses. Second, whereas differential privacy remains restricted to localized queries, structural anonymity protects the whole network, making it well suited to settings that require publishing the full graph. This model mitigates re-identification by altering the network topology so that every node becomes structurally identical to at least $k - 1$ peers. The robustness of this protection depends on the adversarial knowledge radius, denoted by d. While a baseline adversary ($d = 0$) knows only a target's node degree, an adversary in the $d = 1$ regime leverages the 1-hop neighborhood subgraph, exploiting the localized edge patterns among the victim's immediate peers. Taking the need for an operational adversarial model as its starting point, this article frames structural re-identification risk assessment as a *reproducible evaluation problem* within graph mining.

Our Results: We present an assessment framework, rather than a novel anonymization algorithm or an offensive exploit. This article delivers three contributions. First, it introduces a reproducible pipeline for evaluating structural re-identification risk in anonymized social graphs, wherein every experiment is governed by a versioned configuration file, seeds are deterministically managed, and all metrics are derived from structured execution logs. Second, it evaluates the empirical privacy-utility trade-offs of a structural k-anonymity technique [He et al. 2009] across two real-world network datasets. Third, it provides empirical evidence that, within the evaluated $d = 1$ regime, formally satisfying degree-based protection fails to prevent 1-hop structural distinguishability. As a secondary insight, we analyze how increasing k under aggressive anonymization can shift the most effective adversarial strategy from 1-hop neighborhood structures to degree-based signatures.

We employ two real-world networks with asymmetric roles: Facebook Ego-Nets as the primary dataset and Email-Enron as a secondary, cross-domain check. Our central finding is that, on graphs formally satisfying the adopted k-anonymity definition for a degree-based ($d = 0$) adversary, a 1-hop ($d = 1$) neighborhood attack re-identifies substantially more nodes than a degree-based attack: an approximate increase of 6 times on Facebook and 38 times on Enron at $k = 2$.

The term "re-identification" warrants an ethical boundary. The pipeline operates only on publicly available, source-de-identified datasets, uses no external auxiliary data, and reports solely aggregate rates verified against internal labels in a strictly closed environment. It evaluates anonymization robustness rather than targeting real individuals, and never executes the offensive exploit of a true de-anonymization attack: linking anonymized data to external registries to unmask identities.

2. STRUCTURAL ANONYMIZATION IN SOCIAL GRAPHS

k-anonymity does not transfer directly to graphs. Formalized by Sweeney [2002] for tabular microdata, it guarantees that each record is indistinguishable from at least $k - 1$ others on its quasi-identifier attributes; but in a network the structure itself acts as a quasi-identifier, with no single column to suppress that would remove the information carried by the connection pattern.

Backstrom et al. [2007] demonstrated this vulnerability in practice, showing that when social networks are published without identifiers, both active adversaries (who plant substructures before publication) and passive ones (who exploit pre-existing structural knowledge) can re-identify nodes with little structural knowledge. This risk motivated a family of structural anonymization techniques targeting richer structural quasi-identifiers. Liu and Terzi [2008] introduced degree k-anonymity, ensuring each degree is shared by at least k nodes. At the level of the 1-hop neighborhood, Zhou and Pei [2008] showed that immediate neighborhoods remain identifying even when degrees are uniform, and proposed neighborhood anonymization. He et al. [2009] extended protection to local structures by partitioning the graph and rendering each k-anonymous. This family remains active: recent work extends degree k-anonymity to multilevel schemes for directed graphs [Hao et al. 2024], so structural k-anonymity persists as a current line whose risk warrants assessment.

Another line of research focuses on large-scale offensive de-anonymization, crossing anonymized data with external auxiliary information to recover identities; the canonical Netflix attack [Narayanan and Shmatikov 2008] was later extended to social networks via an external auxiliary graph [Narayanan and Shmatikov 2009]. Because this depends entirely on external auxiliary data, it lies outside our closed-environment scope. More recently, de-anonymization has moved beyond exact structural matching: machine-learning methods—graph embeddings that learn latent node representations and re-associate nodes by similarity, tolerating the noise anonymization introduces and the variation across networks—raise the strength of the adversary [Wang et al. 2023]. Such a learned adversary is stronger than the deliberately basic structural scenarios (degree and 1-hop subgraph) assessed here, which reinforces reading the reported rates as a *lower bound* on risk. Much of the privacy community has moved to Differential Privacy (DP); the two differ in *premises*, not merit. DP on graphs typically protects

queries and aggregate statistics answered with calibrated noise under a semantic ε guarantee, without releasing the raw graph, with recent work spanning node-, edge-, and graph-level DP [Mueller et al. 2022; Mendonça et al. 2023]. Structural k-anonymity instead *publishes the whole graph* for reuse in graph mining, under a syntactic guarantee; since legitimate interest persists in releasing complete datasets, which query release does not serve directly and which the still-maturing DP graph-release line covers by another route [Yuan et al. 2023; Brito and Machado 2024], its risk assessment remains pertinent. Finally, Nettleton and Salas [2016] combined neighborhood preservation with t-closeness; since t-closeness constrains attribute distributions, it presupposes attributed graphs and an attribute-disclosure threat model, a regime orthogonal to the purely structural, attribute-free setting evaluated here rather than a comparable baseline. This article contributes a reproducible instrument to *assess* the structural risk that survives an existing technique, measuring rather than assuming the privacy it delivers. We use the structural k-anonymity of He et al. [2009] as a representative baseline to illustrate the utility of this assessment framework: a single canonical technique suffices to show that such assessment is necessary.

Preliminaries and Formal Definitions: Let $G = (V, E)$ be an undirected graph. A *structural quasi-identifier* is any topological feature that allows an adversary to uniquely distinguish a target node $v \in V$. We model this background knowledge using a radius $d \geq 0$, where the attacker knows the local subgraph $G_d(v)$ induced by all nodes within d hops from v. The radius d dictates the adversary's strength: under $d = 0$, the attacker only knows the target's node degree, while under $d = 1$, they leverage the 1-hop neighborhood subgraph $G_1(v)$ to exploit edge patterns among immediate peers. For instance, consider two nodes u and v of degree three. A $d = 0$ adversary cannot tell them apart; yet if the three neighbors of u are mutually connected (forming a triangle) while those of v share no edges, the 1-hop subgraphs $G_1(u)$ and $G_1(v)$ are non-isomorphic, so a $d = 1$ adversary distinguishes them, exactly the degree-equal but neighborhood-distinct gap our evaluation targets.

Definition 1 introduces the concept of structural k-anonymity.

Definition 1. *A graph $G = (V, E)$ satisfies* structural k-anonymity *for a knowledge radius d if, for each v, $|\{u \in V \mid G_d(u) \equiv G_d(v)\}| \geq k$.*

Definition 1 enforces that every node must be structurally identical to at least $k-1$ peers, restricting the re-identification risk to at most $1/k$. Our work evaluates the privacy gap that persists when a network satisfies this condition for a degree-based adversary ($d = 0$) but faces an attacker operating with 1-hop knowledge ($d = 1$).

3. METHODOLOGY

Our methodology is a three-stage pipeline. The first stage, *anonymization*, alters the input graph so that it satisfies structural k-anonymity. The second stage runs a *controlled re-identification scenario* against the anonymized graph, measuring how many nodes remain structurally distinguishable under a fixed adversary. The third stage, *evaluation*, reports the resulting privacy and utility metrics. The remainder of this section details each stage, the evaluated technique, the adversarial scenarios, the metrics, and the experimental design.

Each experiment is configuration-driven: seeds, parameters, validation verdicts, privacy rates, and utility metrics are logged and aggregated per k. All reported numbers derive from these logs, and the pipeline stages communicate only through graph artifacts[1].

Evaluated technique and the main-grid regime: The evaluated technique is the structural k-anonymity of He et al. [2009], under which the *local structure* of each node, a partition of size ℓ (in

[1]Implementation details, experiment configurations, and consolidated reports are available in the versioned project archive: `https://doi.org/10.5281/zenodo.20973364`.

nodes), must be structurally indistinguishable from at least $k-1$ others. We adopt it as the *canonical baseline* of the structural graph k-anonymity family: a widely cited reference whose adversary and defense are both specified in the original paper, a classic, interpretable starting point for a reproducible assessor. The implementation follows the original phases: k-way partitioning of neighborhoods into candidate groups, isomorphism-guided grouping with simplified frequent-subgraph mining (minimum support σ, maximum size $s_{\max}$), the isomorphization of each group's structures by edge operations, edge reconnection between groups, and independent k-anonymity validation of the output graph. Partitioning uses METIS [Karypis and Kumar 1998] on both datasets, a deliberate choice to unify the engine and make the grids directly comparable. Both datasets ran under the same effective anonymizer configuration, with the main grid fixing the anonymizer at $\ell = 1$ (Table I); the structural regime ($\ell > 1$) is probed separately by the local-structure-size sweep.

Controlled re-identification scenarios: Two controlled scenarios run against the anonymized graph and measure the fraction of nodes whose structural signature remains unique, with no external auxiliary information. In the *degree-based scenario*, the signature is the node degree (the weakest adversary, a baseline). In the *1-hop subgraph-based scenario*, the signature is the radius-1 neighborhood, and the scenario counts how many anonymized nodes have an isomorphic 1-hop neighborhood. In both, a node counts as re-identified when its candidate set is a singleton. Isomorphism uses two engines, one per dataset: VF2 [Cordella et al. 2004] on Facebook, with a per-target safety timeout; and, on Enron, a Weisfeiler-Lehman hash [Shervashidze et al. 2011] of the anonymized 1-hop neighborhoods, indexed into buckets and resolved by lookup. The WL hash is a necessary isomorphism invariant (isomorphic neighborhoods share the hash), so no true isomorph is missed and the only theoretical error (a hash collision) is conservative, biasing only toward *under*-counting uniqueness. This lets both datasets' subgraph rates be read as the same metric.

Metrics, datasets, and parameters: Each run yields four metrics: two for privacy and two for utility. The privacy metrics are the two per-scenario re-identification rates rr (degree-based and 1-hop subgraph-based; lower is more resistant). The utility metrics are the Kolmogorov–Smirnov statistic (KS-D) between the original and anonymized degree distributions and the relative variation of the mean clustering coefficient (Δclust). Before any scenario, the anonymized graph passes an independent k-anonymity verifier that does not reuse anonymizer code and operates on the output graph; a coverage deficit is accepted only if fully attributable to identified structural causes (residual incomplete groups). The design uses two real SNAP networks with asymmetric roles (Table II). Facebook is the primary dataset (ego-network 3437 from [Leskovec and McAuley 2012]; ego node removed, restricted to the largest connected component, size floor $10 \cdot k_{\max}$). Email-Enron is the secondary, cross-domain dataset (≈ 63 times larger), symmetrized by OR projection and restricted to the largest connected component. The networks differ in scale, density, and origin, so absolute risk and utility magnitudes are *not* directly comparable; Enron is not a symmetric replica of Facebook, nor a second equivalent point on a single privacy-utility curve, but a generalization (external-validity) test of whether the qualitative trends from the primary dataset hold in a network of different nature. The main grid per dataset is $k \in \{2, 5, 10, 20\}$ with three seeds, both re-identification scenarios (degree-based and 1-hop subgraph-based) enabled; a complementary local-structure-size sweep $\ell \in \{1, 2, 5, 10\}$ (48 runs per dataset) was executed on both datasets to exercise the structural property of the technique.

4. RESULTS

Facebook — k-anonymity achieved, structure exposed: The main grid, run under the unified METIS engine, reached k-anonymity at $k = 2$ and $k = 5$ (coverage 0.9850 and 0.9398; residual deficit fully attributable to incomplete groups, no isomorphism violations). At $k = 10$ and $k = 20$ coverage falls to 0.8647 and the verdict is failure by low coverage: the cause stays structural (METIS yields less balanced partitions on this dense ego-network, leaving more residual nodes as k grows), but the deficit crosses the acceptance threshold, so these two configurations do not satisfy the k-anonymity criterion

Table I. Experimental setup: anonymization and re-identification parameters (uniform across datasets). The anonymization depth (ℓ) and the adversary radius (d) are distinct.

Parameter	**Value**
k	2, 5, 10, 20
ℓ (anonymization local-structure size, nodes)	1 (main grid)
d (re-identification adversary radius)	0 (degree), 1 (1-hop)
σ (FSM minimum support)	0.5
$s_{\max}$ (FSM maximum subgraph size)	4
Isomorphization variant	edge operations
Seeds	42, 1337, 2718
Scenarios	degree-based (d=0); 1-hop subgraph-based (d=1)

Table II. Datasets and their roles. Facebook carries the full experimental grid (all k, seeds, and scenarios) and the primary inferences; Enron is a cross-domain replication (corporate e-mail vs. ego-network) probing the robustness of the observed trends.

Dataset	**Nodes**	**Edges**	**Role**
Facebook Ego-Net 3437	532	4,812	primary
Email-Enron (LCC)	33,696	180,811	secondary

Table III. Facebook ($\ell = 1$, unified METIS engine): aggregates by k (means over 3 seeds). $k = 2, 5$ reach k-anonymity; $k = 10, 20$ fail by low coverage (0.8647, residual incomplete groups, no isomorphism violation).

k	coverage	rr degree	rr subgraph	KS-D	Δclust
2	0.9850	0.0232	0.1454	0.0520	0.0529
5	0.9398	0.0094	0.0307	0.2744	0.2735
10	0.8647	0.2851	0.0370	0.8227	0.1301
20	0.8647	0.0883	0.0000	0.9361	0.4703

(Table III). This weakens, but does not invalidate, the finding: the shortfall is one of coverage against a strict threshold (no cell drops below ≈0.86), not an isomorphism violation. The privacy-utility curve (Figure 1) keeps the trade-off shape: the subgraph re-identification rate falls sharply from $k = 2$ to $k = 5$ ($0.1454 \to 0.0307$), stays low (≈ 0.037 at $k = 10$) and reaches zero at $k = 20$, while utility cost rises steeply (KS-D $0.0520 \to 0.9361$). The degree rate is low at $k = 2$ and $k = 5$ but spikes at $k = 10$ (0.2851): an artifact of the coverage failure, since residual nodes left outside complete groups carry singular degree signatures and count as re-identified.

Enron — k-anonymity achieved, structure exposed on a graph $\approx 63\times$ larger in node count: The secondary-dataset run completed all 12 executions with coverage ≥ 0.9960 and zero timeouts (Table IV). The qualitative trends hold: the subgraph rate falls monotonically with k ($0.1241 \to 0.0569$); the degree rate is residual (≤ 0.0033, ≈ 38 times below subgraph); and utility is far better preserved at scale (KS-D $\in [0.0273, 0.1303]$ versus up to 0.9361 on Facebook). The divergence is at the strongest k: on Facebook the subgraph rate collapses to zero at $k = 20$, whereas Enron retains 0.0569 of distinguishable local structure; a scale effect, not a mechanism difference. The theoretical bound $rr_{\text{subgraph}} \leq 1/k$, which presupposes k-anonymity of the inspected structure ($\ell \geq 2$), can be violated at $\ell = 1$, and is, on Enron at $k = 20$ ($0.0569 > 0.0500$): the expected empirical signature that $\ell = 1$ delivers degree, not structural, k-anonymity. Note that, under the unified engine, the $k = 2$ subgraph rates of the two datasets are now close; what distinguishes the ratios below is the degree rate, residual on Enron and larger on the small Facebook ego-network.

The central finding — equalizing degrees does not protect the neighborhood: At $\ell = 1$ the achieved k-anonymity is, in practice, degree k-anonymity: the $d = 1$ (1-hop) adversary inspects exactly the structure an $\ell = 1$ anonymizer does not touch, so the published graph formally satisfies the

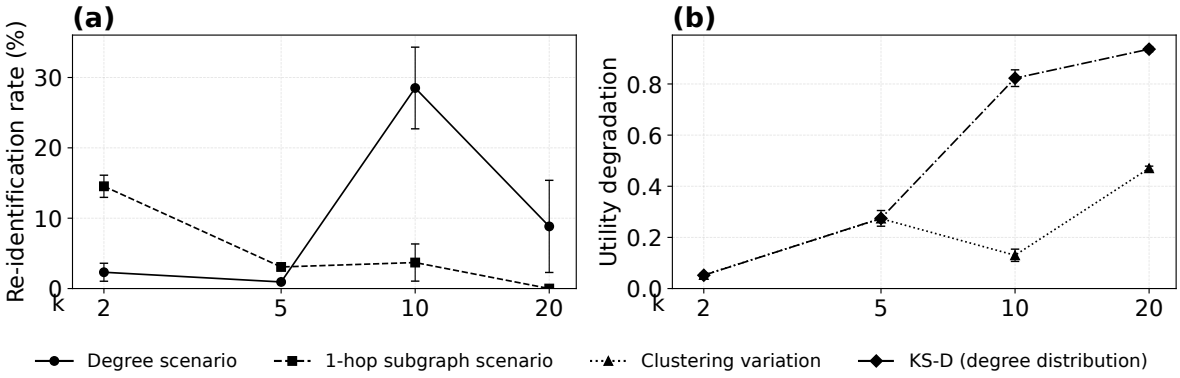

Fig. 1. Privacy–utility trade-off, Facebook ($\ell = 1$), means over 3 seeds with ± 1 std error bars. (a) privacy: re-identification rate (↓ better); (b) utility: degradation, clustering variation and KS-D (↓ better).

Table IV. Enron ($\ell = 1$): aggregates by k (means over 3 seeds).

k	coverage	rr degree	rr subgraph	KS-D	Δclust
2	0.9999	0.0033	0.1241	0.0382	0.0170
5	0.9994	0.0023	0.1024	0.0273	0.0514
10	0.9983	0.0027	0.0787	0.0387	0.0609
20	0.9960	0.0019	0.0569	0.1303	0.0929

Table V. The subgraph × degree gap at $k = 2$, under the unified METIS engine. Ratios round the exact quotients to the nearest integer: 6.3× (Facebook), 37.6× (Enron).

Dataset ($k = 2$, $\ell = 1$)	rr subgraph	rr degree	ratio
Facebook Ego-Net 3437	0.1454	0.0232	$\approx 6\times$
Email-Enron (LCC)	0.1241	0.0033	$\approx 38\times$

definition yet its neighborhoods stay distinguishable. The criterion for a general claim was fixed before writing: a wide *subgraph-over-degree gap* appearing in a single dataset would be a local observation. The gap holds in both networks (Table V, Figure 2): at $k = 2$ the 1-hop rate exceeds the degree rate by $\approx 6\times$ on Facebook (0.1454 vs. 0.0232) and $\approx 38\times$ on Enron, persisting through $k = 5$ in both; at higher k the picture diverges (on Facebook the vector shifts to degree and the subgraph rate collapses to zero by $k = 20$, whereas on Enron the gap persists). That immediate neighborhoods remain identifying under uniform degrees is an established principle [Zhou and Pei 2008]; the contribution is its reproducible quantification on graphs that formally certify degree-based k-anonymity, across networks of different scale, density, and origin and under a single partitioning engine. This is the central result: independent, reproducible evidence that anonymization alone is not sufficient.

Supporting experiment — the local-structure-size sweep: The complementary sweep ($k \in \{2, 5, 10, 20\} \times \ell \in \{1, 2, 5, 10\}$, 48 runs per dataset, METIS throughout) exercises the structural dimension. On Facebook it supports two readings. First, the two scenarios trend oppositely in k: the subgraph rate weakens while the degree rate, residual through $k = 5$, jumps at $k = 10$ as anonymization distorts the degree distribution (KS-D rising from ≈ 0.2–0.3 to ≈ 0.7–0.9) and residual nodes left outside complete groups keep singular degree signatures; increasing k therefore does not improve privacy monotonically but shifts the most effective scenario from subgraph to degree. Second, ℓ controls the target equivalence-group size ($\approx k \cdot \ell$): larger ℓ thickens groups and reduces structural re-identification at the cost of utility, the structural version of the same trade-off.

The same sweep on Enron (48 runs, coverage ≥ 0.987) tests whether this holds in a network $\approx 63\times$

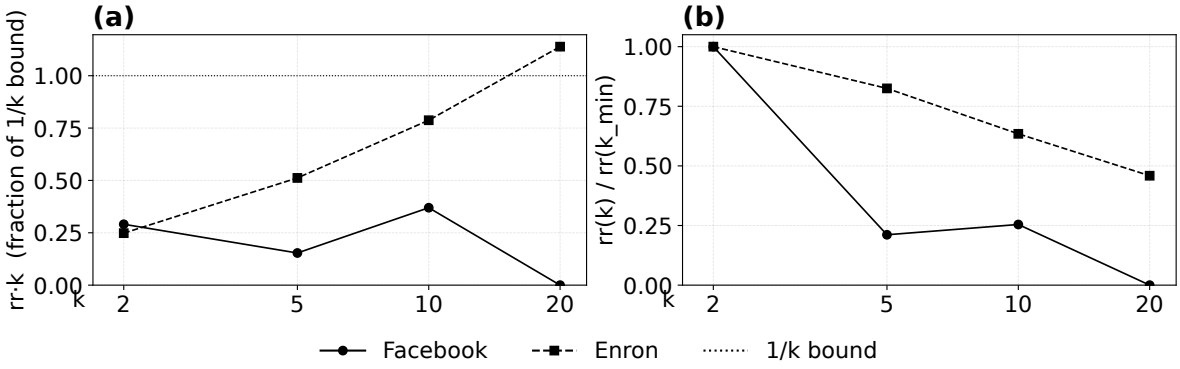

Fig. 2. Facebook vs. Enron, normalized comparison (1-hop subgraph-based scenario). (a) bound fraction $rr_{\text{subgraph}} \cdot k$ relative to the $1/k$ limit (reference line at 1.0); (b) relative decay $rr_{\text{subgraph}}(k)/rr_{\text{subgraph}}(k_{\min})$. Trends hold across networks of different scale, density, and origin; absolute magnitudes are not comparable between datasets.

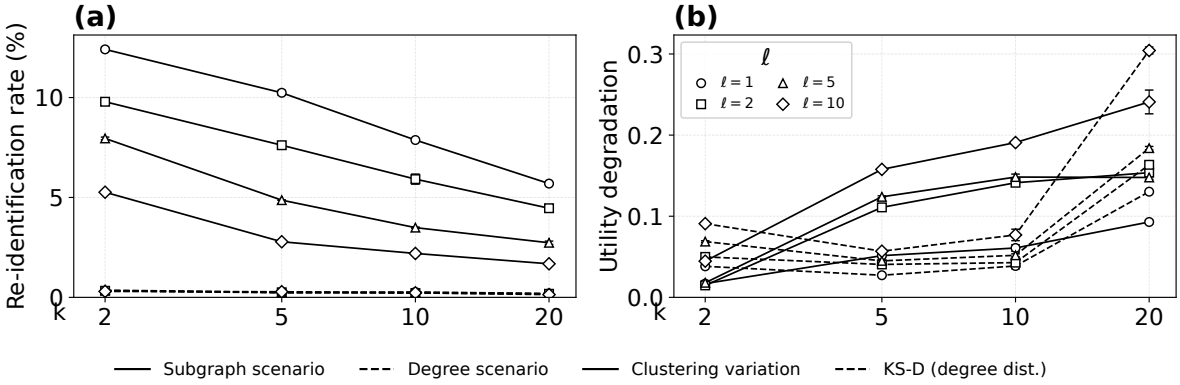

Fig. 3. Local-structure-size sweep on Enron: re-identification rates (subgraph and degree) and utility vs. k, one series per $\ell \in \{1, 2, 5, 10\}$ (means over 3 seeds, ± 1 std). The subgraph rate weakens monotonically with ℓ while the degree rate stays at the floor: no vector shift at scale.

larger, with a partial result (Figure 3). The structural half replicates more cleanly than on Facebook: the equivalence-group size tracks $\approx k \cdot \ell$ and the subgraph rate weakens monotonically with ℓ across all k, at rising utility cost. The vector shift does *not* replicate: the degree rate stays at the floor (≈ 0.002–0.003) across the grid, because in a 33.7k-node network degree collisions are abundant and anonymization rarely isolates a unique degree. There is thus no crossover (the subgraph attack stays dominant in every cell) so the shift is *scale-dependent*: thickening groups weakens the subgraph attack on both networks, but only the small ego-network promotes degree to the top vector.

5. DISCUSSION AND CONCLUSION

The second dataset supports the main qualitative claim: the subgraph-over-degree gap, the monotonic decay of the 1-hop rate with k, and the privacy–utility trade-off persist in Email-Enron, a larger and different network. The $\ell > 1$ sweep refines this claim: thicker equivalence groups reduce subgraph re-identification at a utility cost in both networks, whereas the high-k shift toward degree signatures occurs only in the smaller Facebook ego-network. Thus, the structural effect generalizes, but the dominant attack vector can be scale-dependent; generalization across multiple ego-networks remains open.

The scope is deliberately narrow: all measurements use one canonical anonymization technique

and basic closed-environment adversaries, so the rates are lower bounds rather than exhaustive risk estimates. Running both datasets with METIS removes a partitioning-engine confounder and makes the cross-dataset comparison directly comparable. The instrument exposes a parameterization gap: at $\ell = 1$, the evaluated configuration effectively delivers degree k-anonymity and does not exercise neighborhood protection.

This article presented a reproducible risk-assessment pipeline and applied it to two real social networks. At $k = 2$, the 1-hop subgraph re-identification rate exceeds the degree rate by $\approx 6\times$ on Facebook and $\approx 38\times$ on Enron, despite formal degree-based anonymity. Graphs anonymized by degree-based protection should therefore undergo structural risk assessment before publication. Future work includes multiple ego-networks, partition entropy as an adversarial scenario, attributed graphs (attribute-aware methods such as [Nettleton and Salas 2016]), and dynamic networks.

REFERENCES

Backstrom, L., Dwork, C., and Kleinberg, J. Wherefore art thou r3579x? anonymized social networks, hidden patterns, and structural steganography. In *Proc. of the 16th Int. Conf. on World Wide Web*. ACM, pp. 181–190, 2007.

Brito, F. T. and Machado, J. C. Differentially private release of count-weighted graphs. In *Anais Estendidos do XXXIX Simpósio Brasileiro de Bancos de Dados (SBBD 2024)*. SBC, Porto Alegre, pp. 183–189, 2024.

Cordella, L. P., Foggia, P., Sansone, C., and Vento, M. A (sub)graph isomorphism algorithm for matching large graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (10): 1367–1372, 2004.

Hao, Y., Li, L., Chang, L., and Gu, T. MLDA: A multi-level k-degree anonymity scheme on directed social network graphs. *Frontiers of Computer Science* 18 (2): 182814, 2024.

He, X., Vaidya, J., Shafiq, B., Adam, N., and Atluri, V. Preserving privacy in social networks: A structure-aware approach. In *Proceedings of the IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2009)*. IEEE, pp. 647–654, 2009.

Karypis, G. and Kumar, V. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on Scientific Computing* 20 (1): 359–392, 1998.

Leskovec, J. and McAuley, J. J. Learning to discover social circles in ego networks. In *Advances in Neural Information Processing Systems (NIPS 2012)*. Curran Associates, pp. 539–547, 2012.

Liu, K. and Terzi, E. Towards identity anonymization on graphs. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data (SIGMOD 2008)*. ACM, New York, NY, USA, pp. 93–106, 2008.

Mendonça, A. L. C., Brito, F. T., and Machado, J. C. Privacy-preserving techniques for social network analysis. In *Anais Estendidos do XXXVIII Simpósio Brasileiro de Bancos de Dados (SBBD 2023)*. SBC, Porto Alegre, pp. 174–178, 2023.

Mueller, T. T., Usynin, D., Paetzold, J. C., Rueckert, D., and Kaissis, G. SoK: Differential privacy on graph-structured data. arXiv preprint arXiv:2203.09205, 2022.

Narayanan, A. and Shmatikov, V. Robust de-anonymization of large sparse datasets. In *Proceedings of the 2008 IEEE Symposium on Security and Privacy (S&P 2008)*. IEEE, pp. 111–125, 2008.

Narayanan, A. and Shmatikov, V. De-anonymizing social networks. In *Proceedings of the 2009 30th IEEE Symposium on Security and Privacy (S&P 2009)*. IEEE, pp. 173–187, 2009.

Nettleton, D. F. and Salas, J. A data driven anonymization system for information rich online social network graphs. *Expert Systems with Applications* vol. 55, pp. 87–105, 2016.

Shervashidze, N., Schweitzer, P., van Leeuwen, E. J., Mehlhorn, K., and Borgwardt, K. M. Weisfeiler-lehman graph kernels. *Journal of Machine Learning Research* vol. 12, pp. 2539–2561, 2011.

Sweeney, L. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10 (5): 557–570, 2002.

Wang, H., Yang, W., Man, D., Wang, W., and Lv, J. Anchor link prediction for privacy leakage via de-anonymization in multiple social networks. *IEEE Transactions on Dependable and Secure Computing* 20 (6): 5197–5213, 2023.

Yuan, Q., Zhang, Z., Du, L., Chen, M., Cheng, P., and Sun, M. PrivGraph: Differentially private graph data publication by exploiting community information. In *Proceedings of the 32nd USENIX Security Symposium (USENIX Security 2023)*. USENIX Association, pp. 3241–3258, 2023.

Zhou, B. and Pei, J. Preserving privacy in social networks against neighborhood attacks. In *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering (ICDE 2008)*. IEEE, pp. 506–515, 2008.

Zhou, B., Pei, J., and Luk, W. A brief survey on anonymization techniques for privacy preserving publishing of social network data. *SIGKDD Explor. Newsl.* 10 (2): 12–22, 2008.