XColor: Protecting General Proximity Privacy

Ting Wang Ling Liu

College of Computing, Georgia Institute of Technology

{twang, lingliu}@cc.gatech.edu

***Abstract*— As a severe threat in anonymized data publication, proximity breach is gaining increasing attention. Such breach occurs when an attacker learns with high confidence that the sensitive information of a victim associates with a set of semantically proximate values, even though not sure about the exact one. Recently (ϵ, δ)-dissimilarity [14] has been proposed as an effective countermeasure against general proximity attack. In this paper, we present a detailed analytical study on the fulfillment of this principle, derive criteria to efficiently test its satisfiability for given microdata, and point to a novel anonymization model, XCOLOR, with theoretical guarantees on both operation efficiency and utility preservation.**

I. INTRODUCTION

Privacy preservation has become a paramount concern in numerous data dissemination applications that involve private personal information, e.g., medical data and census data. Typically, such *microdata* is stored in a relational table T: each record in T corresponds to an individual; the attributes of T are categorized as either *sensitive* or *non-sensitive*. In the setting of *central publication*, a publisher intends to release an *anonymized* version T^* of the microdata table T, such that no malicious user, called an *attacker*, can infer the sensitive information regarding any individual from T^*, whereas the statistical utility of T is still preserved in T^*.

Towards this end, a bulk of work has been done on anonymized data publication [1], [3], [7], [9], [10], [11], [12], [13], [14], [15], [16]. One of the major aims is to address *association attack*: the attacker possesses the exact non-sensitive (*quasi-identifier* (QI)) values of the victim, and attempts to discover his/her sensitive (SA) value from the published table T^*. A popular methodology of thwarting such attacks is *generalization* [12]: after partitioning the microdata table T into a set of disjoint subsets of tuples, called *QI-group*, generalization transforms the QI-values in each group to a uniform format such that all tuples belonging to the same group are indistinguishable in terms of their QI-values.

Example 1. Consider publishing the medical data as shown in Table I: *age* and *zip-code* are QI-attributes, while *syndrome* is a composite SA-attribute, each component indicating the severity of a patient's suffering the corresponding symptom. The generalization of the microdata produces two QI-groups, as indicated by the group identifiers (GID). An attacker who knows *Alice*'s QI-values can no longer uniquely identify her SA-value: any tuple in the first group may belong to her; without further information, the attacker can only conclude that *Alice* associates with each specific *syndrome* value with identical probability 20%.

Essentially, the attack above is performed by leveraging the association between quasi-identifier and sensitive attributes (QI-SA association) appearing in the published data. Generalization weakens such association by reducing the representation granularity of QI-values. The protection is sufficient if the weakened association is no longer informative enough for the attacker to infer individuals' SA-values with high confidence, even in a proximate sense, i.e., proximity privacy.

Example 2. Assume that the semantic distance between two SA-values, represented as vectors $P = \langle p_i \rangle_{i=1}^n$ and $Q = \langle q_i \rangle_{i=1}^n$, is defined as $\Delta(P, Q) = \min_i |p_i - q_i|$. Measuring the pairwise distance of the *syndrome* values in the first QI-group, one can notice that the first four tuples form a compact "neighborhood" structure, wherein the value of #3 is proximate to that of #1, #2, and #4, as shown in Fig. 1.

From the attacker's view, every tuple in this group belongs to *Alice* with equal possibility; she can thus conclude that *Alice* associates with the neighborhood structure with probability 80%. Moreover, she might choose the value of the center node (#3) as an estimation, and arrives at a privacy intruding claim that "*Alice*'s *syndrome* value is fairly close to $(0.7, 0.1, 0.1)$".

Existing privacy principles and definitions (e.g., [9], [10], [11], [12], [13], [15], [16]), however, are incapable of capturing this general form of breach because of their assumptions regarding the underlying data models. Recently, $(\epsilon, \delta)^k$-*dissimilarity* [14], a data-model-independent privacy principle, has been proposed as an effective countermeasure against general proximity breach. However, it is proved in [14] that even determining the satisfiability of $(\epsilon, \delta)^k$-*dissimilarity* for given microdata is NP-hard. In this paper, we intend to develop approximate solutions that allow flexible and intuitive tuning of multiple privacy parameters, and find high-quality anonymization of the microdata efficiently.

More concretely, we re-formulate the problem of finding an $(\epsilon, \delta)^k$-dissimilarity-satisfying partition in the framework of *defect graph coloring*; we map it to a novel *relaxed equitable coloring* problem that embeds all the privacy parameters. We then conduct an analytical study on the sufficient conditions (in terms of ϵ, δ, and k) for the existence of a valid coloring. The constructive nature of the proofs naturally leads to a novel anonymization model, XCOLOR, with guarantees on both operation efficiency and utility preservation.

II. $(\epsilon, \delta)^k$-DISSIMILARITY

Let T denote a microdata table intended to be published, which consists of d quasi-identifier (QI) attributes $\{A_i^{qi}\}_{i=1}^d$

	age	zip-code	syndrome			GID
			allergy	asthma	myocarditis	
Alice 1	[18, 30]	[12k, 17k]	**0.8**	**0.0**	**0.0**	1
2	[18, 30]	[12k, 17k]	0.6	0.4	0.4	1
3	[18, 30]	[12k, 17k]	**0.7**	**0.1**	**0.1**	1
4	[18, 30]	[12k, 17k]	1.0	0.2	0.2	1
5	[18, 30]	[12k, 17k]	0.1	0.9	0.9	1
6	[32, 40]	[22k, 30k]	0.2	0.5	0.2	2
7	[32, 40]	[22k, 30k]	0.8	0.1	0.9	2
8	[32, 40]	[22k, 30k]	0.4	0.3	0.5	2
9	[32, 40]	[22k, 30k]	0.6	0.9	0.3	2
10	[32, 40]	[22k, 30k]	1.0	0.7	0.7	2

TABLE I: Anonymized data publication.

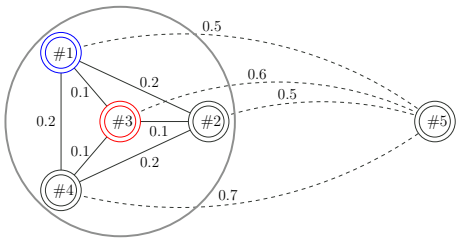

Fig. 1: General proximity breach.

and a sensitive (SA) attribute A^s. Particularly, 1) A^s can be of arbitrary data type, e.g., categorical, numeric, and customized defined type; 2) a semantic distance metric $\Delta(\cdot,\cdot)$ is defined over the domain of A^s, with $\Delta(x,y)$ denoting the distance between two SA-values x and y.

Definition 1 (QI GROUP/PARTITION)**.** *The microdata T is divided into m disjoint subsets of tuples $\mathcal{G}_T = \{G_i\}_{i=1}^m$, which satisfy (i) $\bigcup_{i=1}^m G_i = T$ and (ii) $G_i \cap G_j = \emptyset$ for $i \neq j$. Each G_i is called a QI-group, and $\mathcal{G}_T$ is referred to as a partition of T. The QI-values of G_i is transformed to a uniform format.*

Definition 2 (ϵ-NEIGHBORHOOD)**.** *In a QI-group G with SA-values as a multi-set $\mathcal{SV}_G$, the ϵ-neighborhood of a value $v \in \mathcal{SV}_G$, $\Phi_G(v,\epsilon)$, is defined as the subset of $\mathcal{SV}_G$ with their distance to v within ϵ.*

To remedy general proximity breach, $(\epsilon,\delta)^k$*-dissimilarity* [14] has been proposed as an effective countermeasure.

Definition 3 ($(\epsilon,\delta)^k$-DISSIMILARITY)**.** *A partition $\mathcal{G}_T$ is said to satisfy $(\epsilon,\delta)^k$-dissimilarity if for each $G \in \mathcal{G}_T$, (i) $|G| \geq k$, and (ii) every SA-value v in G has less than $(1-\delta)\cdot(|G|-1)$ ϵ-neighbors.*

Here, ϵ specifies the threshold of semantic proximity; while δ essentially controls the risk of potential proximity breach. It is shown in [14] that a partition $\mathcal{G}_T$ is free of general proximity breach if and only if it satisfies $(\epsilon,\delta)^k$-dissimilarity. Unfortunately, finding a $(\epsilon,\delta)^k$-dissimilarity-satisfying partition $\mathcal{G}^T$ is proved to be NP-hard.

III. THEORY

The key to anonymizing microdata T through generalization is to determine a partition $\mathcal{G}_T$. Instead of attempting to seek the exact answer to whether an $(\epsilon,\delta)^k$-dissimilarity-satisfying partition exists for given T, we are more interested in an approximate solution that allows intuitive and flexible tuning of multiple privacy parameters, and finds high-quality partitions with polynomial complexity.

A. Problem Re-formulation

Given a microdata table T and a proximity threshold ϵ, one can construct an abstract graph $\Psi^\epsilon = (\mathcal{V}^\epsilon, \mathcal{E}^\epsilon)$.

Definition 4 (ABSTRACT GRAPH)**.** *$\mathcal{V}_T^\epsilon$ denotes the set of vertices, each corresponding to a SA-value in T; $\mathcal{E}_T^\epsilon$ represents the set of edges over $\mathcal{V}_T^\epsilon$, and two vertices are adjacent if and only if their corresponding SA-values are ϵ-neighbors.*

A partition $\mathcal{G}$ of T corresponds to a m-coloring of Ψ^ϵ (may not be proper), which partitions the vertices $\mathcal{V}^\epsilon$ into m *color classes*, defined as below.

Definition 5 (COLOR CLASS)**.** *A m-coloring of a graph $\Psi = (\mathcal{V},\mathcal{E})$ partitions $\mathcal{V}$ into m disjoint subsets (color classes) $\{V_i\}_{i=1}^m$, each corresponding to one distinct color.*

Next, we re-formulate the problem of finding an $(\epsilon,\delta)^k$-dissimilarity-satisfying partition $\mathcal{G}$ in the framework of graph coloring. Sufficiently and necessarily, if $\mathcal{G} = \{G_1\}_{i=1}^m$ satisfies $(\epsilon,\delta)^k$-dissimilarity, then there must exist a corresponding coloring of Ψ^ϵ that satisfies the following conditions: 1) Ψ^ϵ is colored using m colors ($\{V_i\}_{i=1}^m$ represent the m color classes); 2) the size of every color class is at least k, i.e., $|V_i| \geq k$ $(1 \leq i \leq m)$; and 3) for any $v \in V_i$ $(1 \leq i \leq m)$, at most $(1-\delta)\cdot(|V_i|-1)$ vertices in V_i are adjacent to v.

We note that the coloring problem above can be considered as a "relaxed" version of the classic proper coloring, in the sense that it allows a constrained number of monochromatic edges (called *defects*). It however deviates from the conventional setting of *defect coloring* as studied in graph theory, e.g., [2], in the sense that it imposes constraints on the size of every color class.

Therefore, in developing our solution, we target the following *relaxed equitable coloring* problem.

Definition 6 (RELAXED EQUITABLE $(\lfloor \frac{n}{k} \rfloor, \delta)$-COLORING)**.** *A relaxed equitable $(\lfloor \frac{n}{k} \rfloor, \delta)$-coloring of a graph Ψ^ϵ satisfies the following conditions:*

(i) Ψ^ϵ is colored using $m = \lfloor \frac{n}{k} \rfloor$ colors, with the corresponding color classes denoted by $\{V_i\}_{i=1}^m$;

(ii) the sizes of any two color classes differ by at most 1;

(iii) for any $v \in V_i$ $(1 \leq i \leq m)$, at most $\lfloor (1-\delta)\cdot(|V_i|-1) \rfloor$ vertices in V_i are adjacent to v.

Clearly, this coloring scheme incorporates both k-anonymity (condition (ii)) and (ϵ,δ)-dissimilarity (condition (iii)). Note that for ease of presentation, here we limit the size of every color class to be either k or $k+1$ (i.e., equitable coloring); our results however can be readily extended to support different group sizes. The details are referred to our technical report due to the space constraint.

To the best of our knowledge, there is no previous study on such relaxed equitable coloring problem; therefore, the solution presented here is interesting in its own right from the perspective of graph theory.

notation	definition
ϵ	threshold of semantic proximity
δ	threshold of breach
k	parameter of k-anonymity
n	cardinality of microdata table T
m	number of color classes $\lfloor n/k \rfloor$
t	$\lfloor (1-\delta)\cdot(k-1) \rfloor$
$\mathcal{G}_T^\epsilon$	abstract graph for given ϵ and T
Θ_T^ϵ	maximum degree of $\mathcal{G}_T^\epsilon$
g	lower bound of the number of movable classes

TABLE II: List of symbols and notations.

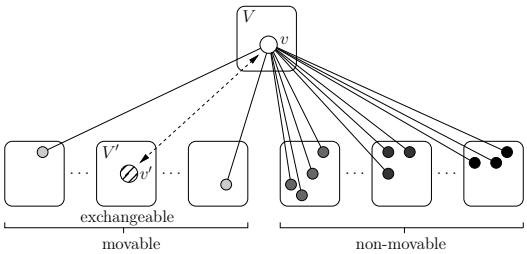

Fig. 2: Movable classes and exchangeable classes.

B. Rationale

Following, we present the theoretical rationale of our equitable $(\lfloor \frac{n}{k} \rfloor, \delta)$-coloring scheme. We begin with introducing the fundamental concepts. The complete list of notations used in the presentation can be found in Table II.

Among the properties of Ψ^ϵ, we are particularly interested in its maximum degree, which is formally defined as below.

Definition 7 (MAXIMUM DEGREE)**.** *The maximum degree Θ^ϵ of a graph Ψ^ϵ is defined as $\Theta^\epsilon = \max_{v \in \mathcal{V}^\epsilon} \mathcal{D}_{\Psi^\epsilon}(v)$, where $\mathcal{D}_{\Psi^\epsilon}(v)$ denotes the degree of v in Ψ^ϵ.*

For the sake of clarity, we assume that n is divisible by k; thus, every color class is of identical size k. We use $m = \frac{n}{k}$ to denote the number of color classes, and $t = \lfloor (1-\delta)\cdot(k-1) \rfloor$ to represent the maximum number of neighbors that a vertex is allowed to have in its self-colored class.

Let $\mathcal{D}_V(v)$ denote the number of neighbors of a vertex v in a color class V. Clearly, if v has overlarge $\mathcal{D}_V(v)$ in its self-colored class V, it violates (ϵ, δ)-dissimilarity, formally

Definition 8 (VIOLATION/MOVABLE CLASS)**.** *Given a color class V and a vertex v, if $v \in V$ and $\mathcal{D}_V(v) \geq (t+1)$, v is called a violation; if $v \notin V$ and $\mathcal{D}_{V'}(v) \leq t$, V is called a movable class for v, or v is movable to V.*

We have the following lemma that establishes a lower bound on the number of movable classes for any v.

Lemma 1. *Given a graph $\Psi^\epsilon = (\mathcal{V}^\epsilon, \mathcal{E}^\epsilon)$, and a m-coloring of Ψ^ϵ, $\mathcal{C} = \{V_i\}_{i=1}^m$, for any $v \in \mathcal{V}^\epsilon$, at least $g = m - \lfloor \Theta^\epsilon/(t+1) \rfloor$ color classes $V \in \mathcal{C}$ satisfy $\mathcal{D}_V(v) \leq t$.*

Proof: [Lemma 1] Summing the degrees of v over all the color classes, we have $\sum_{i=1}^m \mathcal{D}_{V_i}(v) \leq \Theta^\epsilon$. According to the *pigeon-hole* principle, one can derive that at most $\lfloor \Theta^\epsilon/(t+1) \rfloor$ classes contain more than t neighbors of v, from which follows this lemma.

Assume that an initial coloring $\mathcal{C} = \{V_i\}_{i=1}^m$ violates (ϵ, δ)-dissimilarity. The key idea of transforming $\mathcal{C}$ to an (ϵ, δ)-dissimilarity-satisfying coloring $\mathcal{C}' = \{V_i'\}_{i=1}^m$ is to move every violation $v \in V$ $(V \in \mathcal{C})$ to a movable class V' with $\mathcal{D}_{V'}(v) \leq t$. In order to satisfy the requirement that all color classes are of identical size, it is necessary to move a vertex $v' \in V'$ back to V, i.e., v and v' are exchanged. The movement of adding v' to V, however, could potentially create more violations in V, thereby making this transformation process never converge.

To remedy this, we introduce the concept of *global potential* as an indication of the convergence of this process. Specifically, the global potential of a coloring with respect to a graph is defined as the total number of monochromatic edges. Informally, if every move of the transformation makes the potential decrease, the process will converge in polynomial time (the maximum possible potential of a coloring with respect to Ψ^ϵ is $|\mathcal{E}^\epsilon|$). Now, we proceed to formulating the impact of each move over the global potential.

Definition 9 (POTENTIAL CHANGE)**.** *The change in potential resulted from moving a vertex v from color class V to V', $\Lambda(V \xrightarrow{v} V')$, is calculated as $\Lambda(V \xrightarrow{v} V') = \mathcal{D}_{V'}(v) - \mathcal{D}_V(v)$, i.e., the change in the number of monochromatic edges.*

We demand that the move of switching two vertices $v \in V$ and $v' \in V'$ is allowed only if the global potential is decreased, formally

Definition 10 (EXCHANGEABLE CLASS)**.** *A vertex v is exchangeable to a color class V' (or V' is an exchangeable class for v) only if (i) v is movable to V', and (ii) $\exists v' \in V'$, $\Lambda(V \xrightarrow{v} V') + \Lambda(V' \cup \{v\} \xrightarrow{v'} V \setminus \{v\}) < 0$.*

This scenario is illustrated in Fig. 2: among the family of movable classes exists an exchangeable class V' that contains a vertex v' such that switching v and v' results in the decrease of the global potential.

We are thus interested in investigating the existence of exchangeable class among the family of movable classes for v. In the following lemma, we show that if the maximum degree Θ^ϵ is bounded by certain threshold, there must exist at least one exchangeable class for v.

Lemma 2. *If $\Theta^\epsilon \leq \frac{m\cdot(t+1)}{2}$, for an arbitrary coloring $\mathcal{C} = \{V_i\}_{i=1}^m$ and any $v \in V$ with $\mathcal{D}_V(v) \geq (t+1)$, there exists at least one exchangeable class V' for v.*

Proof: [Lemma 2] Otherwise, assume that all the color classes of $\mathcal{C}$ are non-exchangeable for v. Consider the family of movable classes for v. Without loss of generality, assume that $\{V_i\}_{i=1}^g$ are movable for v. Applying the assumption of $\mathcal{D}_V(v) \geq (t+1)$, we have the following two facts:

1) $\sum_{i=1}^g \mathcal{D}_{V_i}(v) \leq \Theta^\epsilon - (m-g)\cdot(t+1)$, derived from Definition 8;

2) $\Lambda(V \xrightarrow{v} V_i) \leq \mathcal{D}_{V_i}(v) - (t+1)$ $(1 \leq i \leq g)$, derived from Definition 9.

According to the assumption, none of the movable classes are exchangeable for v; therefore, any vertex $v' \in V_i$ $(1 \leq i \leq$

g) should satisfy the next condition:

$$\Lambda(V_i \cup \{v\} \xrightarrow{v'} V \setminus \{v\}) \geq -\Lambda(V \xrightarrow{v} V_i)$$

We thus have the following inequality:

$$\begin{aligned}\mathcal{D}_{V\setminus\{v\}}(v') &= \Lambda(V_i \cup \{v\} \xrightarrow{v'} V \setminus \{v\}) + \mathcal{D}_{V_i\cup\{v\}}(v') \\ &\geq \Lambda(V_i \cup \{v\} \xrightarrow{v'} V \setminus \{v\}) \\ &\geq -\Lambda(V \xrightarrow{v} V_i) \\ &\geq (t+1) - \mathcal{D}_{V_i}(v)\end{aligned}$$

Summing $\mathcal{D}_{V\setminus\{v\}}(v')$ over all the vertices in the family of movable classes $\{V_i\}_{i=1}^{g}$ for v, we can obtain

$$\begin{aligned}\sum_{i=1}^{g}\sum_{v'\in V_i}\mathcal{D}_{V\setminus\{v\}}(v') &\geq \sum_{i=1}^{g}\sum_{v'\in V_i}[(t+1) - \mathcal{D}_{V_i}(v)] \\ &= g\cdot k\cdot(t+1) - k\sum_{i=1}^{g}\mathcal{D}_{V_i}(v) \\ &\geq g\cdot k\cdot(t+1) \\ &\quad -k\cdot[\Theta^{\epsilon} - (m-g)\cdot(t+1)] \\ &= m\cdot k\cdot(t+1) - k\cdot\Theta^{\epsilon} \\ &\geq k\cdot\Theta^{\epsilon}\end{aligned}$$

It is thus derived that the maximum degree of the vertices in $V \setminus \{v\}$ is at least $\frac{\sum_{i=1}^{g}\sum_{v'\in V_i}\mathcal{D}_{V\setminus\{v\}}(v')}{k-1} > \Theta^{\epsilon}$, which is a contradiction to the maximality of Θ^{ϵ}.

Based on Lemma 2, we are ready to introduce the following theorem, which can be considered as a significant extension of the classic result of Lovász [8] along the dimension of equitable coloring.

Theorem 1. *For given m, a graph $\Psi = (\mathcal{V}, \mathcal{E})$ with maximum degree Θ can be equitably colored using m colors, with each color class of degree at most $(\frac{2\Theta}{m} - 1)$, in time $O(|\mathcal{E}| \cdot |\mathcal{V}|)$.*

Proof: [Theorem 1] Start with an arbitrary initial coloring wherein all the color classes are of identical size. Consider a vertex v in a class V with more than $(\frac{2\Theta}{m} - 1)$ self-colored neighbors. As proved in Lemma 2, there must exist at least one vertex v' in an exchangeable class V' for v such that switching v and v' decreases the potential of the graph. We exchange the colors of v and v', thereby decreasing the overall number of monochromatic edges in the graph by at least 1. Repeat this process until all the violations are removed. This takes at most $|\mathcal{E}|$ steps, with the cost of each step at most $|\mathcal{V}|$, leading to the overall complexity of $O(|\mathcal{E}| \cdot |\mathcal{V}|)$. We entitle this scheme XCOLOR.

IV. RELATED WORK

Graph coloring has been a prominent topic in graph theory for a long history. An (ordinary vertex) coloring is a partition of the vertices of a graph into independent sets. It is known that determining if a general graph can be colored with less than k colors (its chromatic number) is NP-Hard [5]. Many variants and generalizations have been considered, particularly in relation to practical applications. Cowen et al. [2] considered a relaxation of coloring in which the color classes partition the vertices into subgraphs of degree at most d, called (k, d)-coloring, following the classic work of Lovász [8]. In [4], Erdös considered the problem of equitable coloring, imposing the constraint that each color class should be of identical size, and made the famous conjecture that the chromatic number of a graph with maximum degree Θ is at most $(\Theta + 1)$, which was later proved in [6]. However, to the best of our knowledge, no previous work exists on the problem of equitable coloring with defect, as discussed in this paper.

V. CONCLUSION

This work represents a detailed analytical study on the fulfillment of $(\epsilon, \delta)^k$-dissimilarity, a general proximity privacy definition. We derived the criteria that enable to efficiently check its satisfiability for given microdata table, and developed a novel anonymization model, XCOLOR, with guarantees on both operation efficiency and utility preservation. One of our ongoing research directions is to bridge the gap between theoretical methodology and practical generalization algorithm, by addressing key challenges including parameter setting, utility optimization, and algorithm efficiency.

ACKNOWLEDGEMENT

This work is partially supported by grants from NSF CyberTrust, NSF NetSE, and IBM faculty award, IBN SUR grant and a grant from Intel research council.

REFERENCES

[1] C. Aggarwal. "On k-anonymity and the curse of dimensionality". In *VLDB*, 2005.

[2] L. Cowen, W. Goddard and C. Jesurum. "Coloring with defect". In *SODA*, 1997.

[3] B. Chen, R. Ramakrishnan and K. LeFevre. "Privacy skyline: privacy with multidimensional adversial knowledge". In *VLDB*, 2007.

[4] P. Erdös. "Some applications of probability of graph theory and combinatorial problems". Theory of Graphs and its Applications, Publ. House Czechoslovak Acad. Sci., Prague, 1964.

[5] M. Garey and D. Johnson. "Computers and intractability: a guide to the theory of NP-completeness". Freeman, San Francisco, CA, 1981.

[6] A. Hajnal and E. Szemerédi. "Proof of a conjecture of P. Erdös". Combinatorial theory and its application, II. North-Holland, Amsterdam, 1970.

[7] D. Kifer and J. Gehrke. "Injecting utility into anonymization databases". In *SIGMOD*, 2006.

[8] L. Lovász. "On decompositions of graphs". *Studia Sci. Math. Hungar.*, 1:237-238, 1966.

[9] K. LeFevre, D. DeWitt and R. Ramakrishnan. "Workload-aware anonymization". In *SIGKDD*, 2006.

[10] J. Li, Y. Tao and X. Xiao. "Preservation of proximity privacy in publishing numerical sensitive data". In *SIGMOD*, 2008.

[11] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. "l-diversity: privacy beyond k-anonymity". In *ACM TKDD*, 1(1), 2007.

[12] L. Sweeney. "k-anonymity: a model for protecting privacy". In *International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems*, 10(5), 2002.

[13] R. Wong, J. Li, A. Fu and K. Wang. "(alpha, k)-anonymity: an enhanced k-anonymity model for privacy preserving data publishing". In *SIGKDD*, 2006.

[14] T. Wang, S. Meng, B. Bamba, L. Liu and C. Pu. "A General Proximity Privacy Principle". In *ICDE*, 2009.

[15] X. Xiao and Y. Tao. "m-invariance: towards privacy preserving republication of dynamic datasets". In *SIGMOD*, 2007.

[16] Q. Zhang, N. Koudas, D. Srivastava and T. Yu. "Aggregate query answering on anonymized tables". In *ICDE*, 2007.