frontiers in Big Data

ORIGINAL RESEARCH
published: 31 May 2019
doi: 10.3389/fdata.2019.00007

You Can't See Me: Anonymizing Graphs Using the Szemerédi Regularity Lemma

Daniele Foffano[1], *Luca Rossi*[2*] *and Andrea Torsello*[1]

[1] *Dipartimento di Scienze Ambientali, Informatica e Statistica, Università Ca' Foscari Venezia, Venezia, Italy,* [2] *Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China*

Complex networks gathered from our online interactions provide a rich source of information that can be used to try to model and predict our behavior. While this has very tangible benefits that we have all grown accustomed to, there is a concrete privacy risk in sharing potentially sensitive data about ourselves and the people we interact with, especially when this data is publicly available online and unprotected from malicious attacks. *k*-anonymity is a technique aimed at reducing this risk by obfuscating the topological information of a graph that can be used to infer the nodes' identity. In this paper we propose a novel algorithm to enforce *k*-anonymity based on a well-known result in extremal graph theory, the Szemerédi regularity lemma. Given a graph, we start by computing a regular partition of its nodes. The Szemerédi regularity lemma ensures that such a partition exists and that the edges between the sets of nodes behave almost randomly. With this partition, we anonymize the graph by randomizing the edges within each set, obtaining a graph that is structurally similar to the original one yet the nodes within each set are structurally indistinguishable. We test the proposed approach on real-world networks extracted from Facebook. Our experimental results show that the proposed approach is able to anonymize a graph while retaining most of its structural information.

Keywords: privacy, anonymity, social networks, graph, regularity lemma

OPEN ACCESS

Edited by:
Roberto Interdonato, Télédétection et Information Spatiale (TETIS), France

Reviewed by:
Ruggero Gaetano Pensa, University of Turin, Italy
Matteo Zignani, University of Milan, Italy

***Correspondence:**
Luca Rossi
rossil@sustech.edu.cn

Specialty section:
This article was submitted to Data Mining and Management, a section of the journal Frontiers in Big Data

Received: *25 March 2019*
Accepted: *13 May 2019*
Published: *31 May 2019*

Citation:
Foffano D, Rossi L and Torsello A (2019) You Can't See Me: Anonymizing Graphs Using the Szemerédi Regularity Lemma. Front. Big Data 2:7. doi: 10.3389/fdata.2019.00007

1. INTRODUCTION

The beginning of the twenty-first century has been characterized by the rise of online social media and data-hungry artificial intelligence (AI). In this context, sophisticated machine learning algorithms feed off massive amounts of data produced by our digital personas to perfect the way they model and predict our behavior, both online and offline. However, the comforts of an increasingly AI-assisted life are overshadowed by the threat it poses to our privacy and freedom (Fung et al., 2010; Rossi and Musolesi, 2014; Rossi et al., 2015b; Qian et al., 2016). At the same time, the digital traces we produce, particularly interactions between users in an online social network, are often abstracted using a graph representation and made available in the form of public datasets, as they offer a unique opportunity for researchers to study real-world complex networks of interactions (Kwak et al., 2010; Chorley et al., 2016).

A common practice to protect the identity of the users whose interactions are captured by the graph is that of stripping the nodes of sensitive information (e.g., the users names), generating a random identifier to label the graph nodes. However, it has been shown that this does not guarantee that the user's privacy is preserved (Backstrom et al., 2007). Indeed, it is possible to disclose the identity of an individual participating in the network with minimal external background information. One common example is that of a user for which the number of connections in the network is known (i.e., the number of friends on Facebook) and this number happens to be unique for that individual. In other words, this piece of information alone would be sufficient to identify that user among the rest of the nodes. Most importantly, once the identity is revealed, other potentially sensitive pieces of information can be inferred. For instance, the individual may turn out to belong to a group of nodes labeled with a certain sensitive attribute, e.g., health condition.

For these reasons, the problem of anonymizing graph data is becoming an increasingly studied one (Hay et al., 2008; Liu and Terzi, 2008; Rossi et al., 2015a; Qian et al., 2016). A common anonymity model is k-anonymity, which aims to ensure that each node in a network is structurally indistinguishable from at least other k nodes. Different works have focused on different definitions of "structurally indistinguishable." Liu and Terzi (2008) considered the case of k-degree anonymous graphs, where k-degree anonymity guarantees that each node of the graph shares the same degree of at least k other nodes. Successive works attempted to reduce the total running time of Liu and Terzi (2008) to make it feasible to scale up to large networks (Hay et al., 2008). Rossi et al. (2015a), on the other hand, extended the concept of k-degree anonymity to multi-layer and time-varying graphs. Other researchers considered different structural distinguishability criteria where the attacker has increasing levels of information available to deanomymize the nodes (Hay et al., 2008; Cheng et al., 2010; Zhou and Pei, 2011), however the main issue with these approaches lies in the need to add increasing amounts of noise as increasingly complex structural information needs to be obfuscated. More recently Rousseau et al. (2018) considered the problem of anonymizing a graph maximizing the amount of preserved community information. Finally, Qian et al. (2016) and Ma et al. (2018) looked at the complementary problem of deanonymizing a graph in the case where the attacker has access to richer features as well as structural information.

While most of the previous k-anonymity approaches assume that the attacker has access only to a certain level of structural information (from the degree of a node, to its immediate neighborhood or even the whole graph), in this paper we propose a method that creates k-anonymous groups of nodes where no degree of structural information can help to break the anonymity guarantee. Our approach is based on the Szemerédi regularity lemma (Diestel, 2012), a well-known result of extremal graph theory. The Szemerédi regularity lemma has been successfully applied to several problems, from graph theory (Komlós and Simonovits, 1996) to computer vision and pattern recognition (Sperotto and Pelillo, 2007; Pelillo et al., 2017). The lemma roughly states that every sufficiently large and dense graph[1] can be approximated by the union of random-like bipartite graphs called regular pairs. Our observation is that the groups of graph nodes that form these regular pairs can be anonymized by rewiring the intra-group edges according to an Erdös-Rényi process (Erdős, 1960). Thanks to the theoretical guarantees of the Szemerédi regularity lemma, this has minimal effect on the overall graph structure and, together with the random-like behavior of the inter-group connections, ensures that the each group is anonymous.

The reminder of the paper is organized as follows. We start by reviewing the key graph theoretical concepts underpinning our work in section 2. In section 3 we propose our anonymization method based on the Szemerédi regularity lemma and in section 4 we evaluate it on three different networks abstracted from Facebook. Finally, section 5 concludes the paper.

2. SZEMERÉDI REGULARITY LEMMA

Let $G = (V, E)$ be an undirected graph with no self-loops, where V is the set of nodes and E is the set of edges. If X and Y are disjoint subsets of V, the *edge density* of this pair (X, Y) is defined as $d(X, Y) = \frac{|E(X,Y)|}{|X||Y|}$, where $E(X, Y)$ is the set of edges connecting nodes in X to nodes in Y. The edge density satisfies $0 \leq d(X, Y) \leq 1$.

Given a positive real $\varepsilon > 0$, a pair of node sets X and Y is called *ε-regular* if for all subsets $A \subseteq X$ and $B \subseteq Y$ satisfying $|A| \geq \varepsilon|X|$ and $|B| \geq \varepsilon|Y|$ we have $|d(X, Y) - d(A, B)| \leq \varepsilon$. Stated otherwise, the distribution of the edges between an ε-regular pair is almost uniform, i.e., the graph over $X \cup Y$ behaves like a random bipartite graph.

Let the node set V be divided into a partition $\mathcal{P}$ of l sets $V_1, \cdots, V_l$. $\mathcal{P}$ is an *ε-regular partition* if: (1) $||V_i| - |V_j|| \leq 1$, for $1 \leq i < j \leq l$ and (2) all except at most εl^2 pairs (V_i, V_j) $(1 \leq i < j \leq l)$, are ε-regular. With these definitions in hand, we can finally state the following.

Lemma 2.1 (Szemerédi regularity lemma)**.** *For every positive real $\varepsilon > 0$ and every positive integer m, there exist positive integers $N = N(\varepsilon, m)$ and $M = M(\varepsilon, m)$ such that, if $G = (V, E)$ is a graph with $|V| \geq N$ nodes, there is an ε-regular partition of V into l groups with sizes that differ at most by 1, where $m \leq l \leq M$.*

In other words, the Szemerédi regularity lemma states that a graph can be seen as a collection of groups of nodes such that the edges between these groups are almost uniformly distributed. More generally, as stated by Komlós and Simonovits (1996), the regularity lemma states that every graph can be approximated by generalized random graphs. Note that the lemma also states that there may be a number of ε-irregular pairs that do not behave like random bipartite graphs. However, for a sufficiently small ε, the number of such pairs will be low (i.e., smaller than εl^2).

Given a graph G and an ε-regular partition of its nodes, a reduced graph can be constructed by replacing each pair of ε-regular groups with two nodes connected by an edge. As shown

[1] Note that the lemma has been extended to sparse graphs as well (Gerke and Steger, 2005).

by the Key lemma (Komlós and Simonovits, 1996), the reduced graph inherits many of the fundamental structural properties of the original graph, to the point that the graph obtained by simply replacing each pair of connected nodes of the reduced graph with a complete bipartite graph over $2t$ nodes yields a new graph that can be used as a surrogate of the original one, where $t \geq 1$ is an integer.

Recall that the aim of this paper is to anonymize a graph $G = (V, E)$ by grouping V into sets of k-anonymous nodes. The Szemerédi regularity lemma states that the node set of each graph can be rearranged to reveal a random-like structure, where pairs of groups of k nodes are connected in an almost uniform (in other words, random) way. That is, for the purpose of graph de-anonymization, the edge information between the groups of nodes is unusable. Unfortunately, the intra-group connections can be still exploited to deanonymize the nodes. However, the Szemerédi regularity lemma and the fact that the reduced graph (where the intra-group connections are lost) preserves the fundamental structural properties of the original graph imply that these intra-group connections are small in number and structurally negligible.

3. ANONYMIZATION FRAMEWORK

In the previous section we introduced the Szemerédi regularity lemma and we showed how this can be seen as a first step toward obtaining a k-anonymous graph. To achieve full k-anonymity, however, we need to obfuscate the structural information contained in the intra-group connections of the ε-regular partition. Our solution involves rewiring these connections using the Erdös-Rényi model (Erdős, 1960), effectively replacing each subgraph (i.e., each group of the ε-regular partition) with an Erdös-Rényi graph over the same set of nodes. Crucially, for each subgraph, we set the parameter p, which governs the probability of adding/deleting an edge, equal to the density of the original subgraph. More specifically, our approach follows three steps: (1) we first find a regular partition using the regularity lemma; (2) then, we randomize the groups' intra-connections; and (3) finally, we randomize the edges connecting irregular pairs.

In the **first step** we apply the algorithm implemented by Fiorucci et al. (2019)[2]. This extends the previous algorithm of Fiorucci et al. (2017) by proposing a novel heuristic procedure where the node set is first partitioned into two groups of nodes and then these are recursively split into smaller groups until a desired cardinality is met and certain conditions that measure quality of the ε-regularity of the partition are satisfied (Pelillo et al., 2017). In particular Fiorucci et al. propose two different heuristics to split the groups, one called *degree based*, which groups together nodes with similar degrees (Fiorucci et al., 2017), and a second one called *indeg guided*, which splits a sparse (dense) partition into two sparse (dense) partitions. Note that using this method we can only get a number of ε-regular groups which is a power of 2.

[2]Code available at: https://github.com/MarcoFiorucci/graph-summarization-using-regular-partitions.

The **second step** involves randomly rewiring the connections within each group of vertices. To this end, we add or delete an edge with a probability p equal to the density of the subgraph H spanned by the group of nodes we are trying to anonymize. Note that we only change the internal connections of H, so we are not altering the ε-regularity relations. The resulting subgraph H' will have the same density of H, however its structural information will not be of any use when trying to deanonymize its nodes.

Recall that each ε-regular partition allows up to εl^2 irregular pairs, where l is the number of sets of the ε-regular partition. So far we ensured that the connections within and between ε-regular pairs are anonymous, however we have not yet dealt with irregular pairs. The **third step** addresses this and requires rewiring the connections between groups forming an ε-irregular pair. Let (V_i, V_j) be one such pair, with total number of nodes n. Consider the bipartite subgraph $H = (V_i \cup V_j, E_{ij})$ where we only consider the set of edges E_{ij} connecting nodes in V_i with nodes in V_j. In order to render the structural information contained in these edges unusable for deanonymization purposes, we randomly rewire each pair of nodes (u, v), with $u \in V_i$ and $v \in V_j$, by adding/deleting an edge to E_{ij} with probability p equal to $|E_{ij}|/(V_i \times V_j)$.

In this framework ε can be interpreted as a measure of the error made by the Szemerédi regularity lemma approximation, i.e., the smaller ε the better the anonymized graph approximates the original graph. In fact, the amount of structural information preserved is inversely proportional to the number of edges we need to rewire. The Szemerédi regularity lemma allows us to safely rewire intra-group connections, knowing that these are small in number and structurally negligible. So the key to preserving the structural information of the original graph is to minimize the number of ε-irregular pairs. This becomes particularly relevant when anonymizing real-world complex networks, which often display a scale-free structure (Barabási and Albert, 1999). In these networks a small number of nodes (i.e., hubs) has a very large degree. If an irregular pair contains a hub we will end up rewiring a large number of edges, potentially compromising the structural information for the sake of anonymity. Therefore, minimizing the number of ε-irregular pairs is of fundamental importance. Also, recall that the method of Fiorucci et al. is based on heuristics, and in general different runs of their algorithm can result in different ε-regular partitions. For this reason, we repeat the computation of the ε-regular partition `max_iter` times and we choose the partition with the minimum ε and number of ε-irregular pairs. Note that each iteration of the algorithm of Fiorucci et al. has computational cost $O(n^{2.376})$, and this cost dominates in the overall anonymization complexity.

4. EXPERIMENTAL RESULTS

We test the proposed method on three real-world networks abstracted from Facebook. Note that all the graphs are sparse, as shown in **Table 1**. *Facebook Combined* represents circles (or friend lists) from Facebook. It was introduced for the first time by Mcauley and Leskovec in Leskovec and Mcauley (2012). The

two remaining networks, *Tv Shows* and *Politicians* describe blue verified pages of different kinds, where edges represent mutual likes among them (Rozemberczki et al., 2018).

With these graphs in hand, we compute their anonymized versions and we measure the amount of structural information lost with respect to the original graphs. In particular, we track the changes in number of edges, degree distribution, average clustering coefficient (Watts and Strogatz, 1998), and page rank vector (Page et al., 1999). We compute these changes for different levels of k-anonymity, which in turn correspond to different choices of the partition cardinality l. Recall in fact that k and l are related by the fact that in a graph with n nodes an ε-regular partition groups the vertices into l sets of cardinality $k \approx \frac{n}{l}$.

Note also that larger values of l also imply larger values of εl^2, the maximum number of ε-irregular pairs we can find in the network. Irregular pairs force us to randomly rewire connections that are not guaranteed to be structurally negligible by the Szemerédi regularity lemma (like the intra-group connections), so in general for large values of l more effort has to go into finding an ε-regular partition with minimum value of ε (in these experiments we vary ε from 0.01 to 0.2, with steps of 0.025). This is also the reason why we were only able to compute the ε-regular partitions for a small range of values of l. In fact, for some combinations of dataset and l, the algorithm of Fiorucci et al. was unable to find an optimal partition within `max_iter` $= 100$ iterations. In our experiments, the runtime to compute an ε-regular partition varies between approximately 10 and 80 s, on a machine with an 8-core 3.6 GHz CPU and 16GB of RAM.

We start by comparing the degree distributions of the original graphs and the anonymized ones, using both the *degree based* and the *indeg guided* heuristics. **Figure 1** shows the log-log plots of the results. Note that larger values of l tend to correspond to more accurate approximations of the original degree distribution. This is confirmed by looking at the Jensen-Shannon (JS) divergence Lin (1991) between the degree distributions, which for the *degree guided* heuristic and the *Politicians* dataset goes from 0.062 (with $l = 4$) to 0.011

TABLE 1 | Summary of the main structural characteristics of the original graphs.

Dataset	Nodes	Density	Edges	Avg. clustering coefficient
Facebook Combined	4,039	0.011	88,234	0.606
Politicians	3,892	0.002	41,729	0.385
Tv shows	5,908	0.002	17,262	0.374

TABLE 2 | Average variation in the number of edges (average clustering coefficient) between the original graph G and the anonymized graph $\tilde{G}$, calculated as $|s_G - s_{\tilde{G}}|/s_G$, where s_G and $s_{\tilde{G}}$ are the statistics considered.

Dataset	$l = 4$	$l = 8$	$l = 16$	$l = 32$	$l = 64$
Facebook Combined	0.0012 (0.7162)	0.0012 (0.6310)	0.0010 (0.5696)	0.0010 (0.5302)	0.0010 (0.4822)
Politicians	0.0021 (0.6983)	0.0020 (0.6415)	0.0015 (0.5261)	0.09 (0.2395)	n.a.
Tv shows	0.0034 (0.6553)	0.0036 (0.5064)	0.0013 (0.3158)	n.a.	n.a.

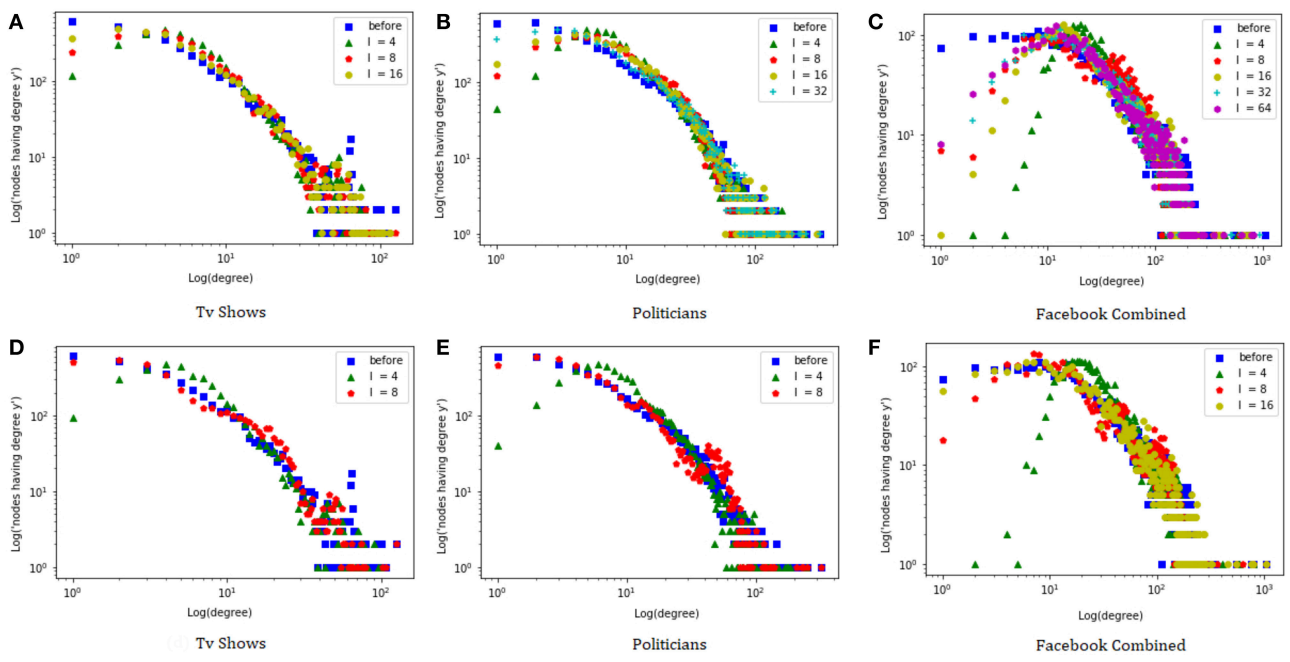

FIGURE 1 | Degree distribution of the graphs with *degree based* **(A–C)** and *indeg guided* **(D–F)** heuristics.

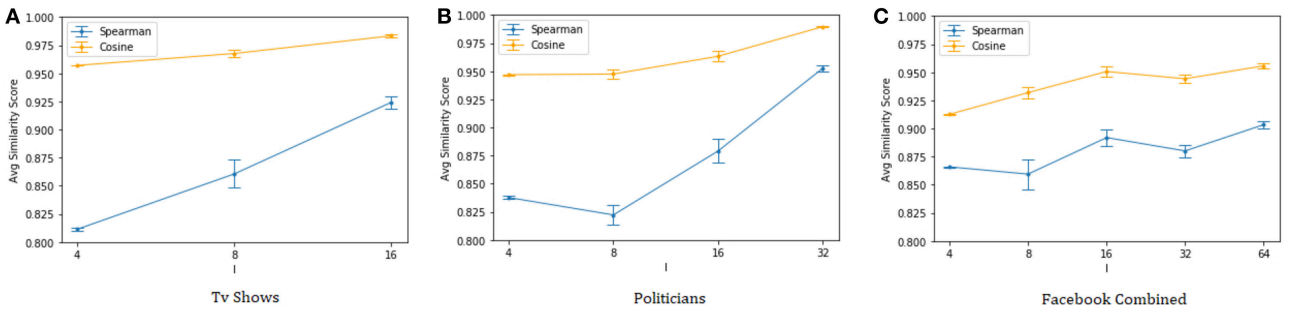

FIGURE 2 | Cosine similarity and Spearman's correlation of the page rank vectors (*indeg guided* heuristic). **(A)** Tv shows, **(B)** Politicians, and **(C)** Facebook Combined.

(with $l = 32$)[3]. Interestingly, the *indeg guided* heuristic seems to yield the best approximations. This could be because the degree-based heuristic struggles to create groups of nodes with similar degree when there are hubs among them. Indeed, for the *indeg guided* heuristic the JS divergence goes from 0.066 (with $l = 4$) to 0.016 ($l = 8$), whereas for $l = 8$ the *degree guided* heuristic achieves a JS divergence of 0.034[4]. In the remainder of the experiments we focus only on the *indeg guided* heuristic.

Table 2 shows the variation in the number of edges and average clustering coefficient with respect to the original graph. More precisely, we report $|s_G - s_{\tilde{G}}|/s_G$, where s_G and $s_{\tilde{G}}$ are statistics computed on the original and anonymized graphs, respectively (averaged over 10 anonymizations). We first note that the number of edges of the graphs changes only very slightly. Indeed, when we alter the structure of a group of vertices we do it by adding/deleting edges with a probability equal to the original edge density of the group. This in turn has the effect of keeping the number of edges approximately the same, regardless of the size k of the anonymity sets.

We then check the effect of the anonymization on the average clustering coefficient of the graph. **Table 2** shows that these statistics change significantly. Recall that the average clustering coefficient is proportional to the number of triangles in a network (Watts and Strogatz, 1998), however the Erdös-Rényi rewiring used to anonymize the vertex groups and the ε-irregular pairs is likely to break these triangles. While the Szemerédi regularity lemma ensures that the vertex groups are sufficiently sparse that we can ignore their inner structure, this clearly does not hold for ε-irregular pairs, which we also need to anonymize. This is particularly an issue when hubs fall within such an irregular pair. However, note that increasing l (i.e., reducing the size k of the anonymity sets) allows us to preserve the average clustering coefficient better. In general, a low value of l implies larger anonymity groups, but it also forces the heuristic procedure used to approximate the ε-regular partition to bring more edges (and triangles) inside the groups, which are then affected by the Erdös-Rényi rewiring. Indeed, high anonymity demands several more structural modifications. In practice it is common to look for smaller k-anonymity groups (i.e., larger l), and for these values we are better able to preserve the average clustering coefficient information.

Finally, **Figure 2** shows the cosine similarity and the Spearman's rank correlation between the page rank vectors (Page et al., 1999) of the original and anonymized graphs. The results confirm that the proposed anonymization procedure is able to preserve well the centrality information of the nodes, once again with the quality of the approximation generally improving as we reduce the size of the anonymity groups.

5. CONCLUSION

We considered the problem of protecting the identity of the nodes of a network from an attacker with background structural knowledge. We proposed to use the Szemerédi regularity lemma to compute an ε-regular partition of the original graph which is then anonymized by injecting Erdös-Rényi at selected locations. This creates a k-anonymous graph where the loss of structural information is minimized. We validated our method on three real-world networks abstracted from Facebook. Future work should perform a more extensive evaluation of the proposed method on larger graphs, with a wider range of values, and compare our method with alternative anonymization approaches.

DATA AVAILABILITY

Publicly available datasets were analyzed in this study. This data can be found here: a https://snap.stanford.edu/data/index.html.

AUTHOR CONTRIBUTIONS

AT: conceptualization. LR and AT: methodology. DF: software. DF, LR, and AT: investigation, writing–review, and editing. LR: writing–original draft preparation.

[3] The JS divergence takes a value between 0 and 1, with 0 indicating identical distributions. Results on other datasets are omitted due to space constraints.

[4] Note, however, that the value of the JS divergence is biased by the fact that most of the probability mass is on low-degree nodes.

REFERENCES

Backstrom, L., Dwork, C., and Kleinberg, J. (2007). "Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography," in *Proceedings of the 16th International Conference on World Wide Web (WWW '07)* (Banff, AB), 181–190. doi: 10.1145/1242572.1242598

Barabási, A.-L., and Albert, R. (1999). Emergence of scaling in random networks. *Science* 286, 509–512. doi: 10.1126/science.286.5439.509

Cheng, J., Fu, A. W.-C., and Liu, J. (2010). "K-isomorphism: privacy preserving network publication against structural attacks," in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data (SIGMOD '10)* (Indianapolis, IN), 459–470. doi: 10.1145/1807167.1807218

Chorley, M. J., Rossi, L., Tyson, G., and Williams, M. J. (2016). "Pub crawling at scale: tapping untappd to explore social drinking," in *Tenth International AAAI Conference on Web and Social Media* (Cologne). Available at: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13048 (accessed May 20, 2019).

Diestel, R. (2012). *Graph Theory*. Graduate Texts in Mathematics, Vol. 173. Available online at: https://www.springer.com/gp/book/9783662536216

Erdős, P. (1960). Graphs with prescribed degrees of vertices (hungarian). *Mat. Lapok* 11, 264–274.

Fiorucci, M., Pelosin, F., and Pelillo, M. (2019). Separating structure from noise in large graphs using the regularity lemma. *CoRR* abs/1905.06917.

Fiorucci, M., Torcinovich, A., Curado, M., Escolano, F., and Pelillo, M. (2017). "On the interplay between strong regularity and graph densification," in *11th IAPR-TC-15 International Workshop, GbRPR 2017* (Anacapri), 165–174.

Fung, B., Wang, K., Chen, R., and Yu, P. S. (2010). Privacy-preserving data publishing: a survey of recent developments. *ACM Comput. Surveys* 42:14. doi: 10.1201/9781420091502

Gerke, S., and Steger, A. (2005). The sparse regularity lemma and its applications. *Surveys Combin.* 327, 227–258. doi: 10.1017/CBO9780511734885.010

Hay, M., Miklau, G., Jensen, D., Towsley, D., and Weis, P. (2008). Resisting structural re-identification in anonymized social networks. *Proc. VLDB Endow.* 1, 102–114. doi: 10.14778/1453856.1453873

Komlós, J., and Simonovits, M. (1996). Szemerédi's regularity lemma and its applications in graph theory. *Combinatorics* 2, 295–352.

Kwak, H., Lee, C., Park, H., and Moon, S. (2010). "What is twitter, a social network or a news media?," in *Proceedings of the 19th International Conference on World Wide Web (WWW '10)* (Raleigh, NC), 591–600. doi: 10.1145/1772690.1772751

Leskovec, J., and Mcauley, J. J. (2012). "Learning to discover social circles in ego networks," in *Advances in Neural Information Processing Systems*, 539–547.

Lin, J. (1991). Divergence measures based on the shannon entropy. *IEEE Trans. Inform. Theor.* 37, 145–151. doi: 10.1109/18.61115

Liu, K., and Terzi, E. (2008). "Towards identity anonymization on graphs," in *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data (SIGMOD '08)* (Vancouver, BC), 93–106.

Ma, J., Qiao, Y., Hu, G., Huang, Y., Sangaiah, A. K., Zhang, C., et al. (2018). De-anonymizing social networks with random forest classifier. *IEEE Access* 6, 10139–10150. doi: 10.1109/ACCESS.2017.2756904

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). *The Pagerank Citation Ranking: Bringing Order to the Web*. Technical Report 1999-66, Stanford InfoLab.

Pelillo, M., Elezi, I., and Fiorucci, M. (2017). Revealing structure in large graphs: Szemerédi's regularity lemma and its use in pattern recognition. *Pattern Recogn. Lett.* 87, 4–11. doi: 10.1016/j.patrec.2016.09.007

Qian, J., Li, X.-Y., Zhang, C., and Chen, L. (2016). "De-anonymizing social networks and inferring private attributes using knowledge graphs," in *IEEE INFOCOM 2016–The 35th Annual IEEE International Conference on Computer Communications* (San Francisco, CA), 1–9.

Rossi, L., and Musolesi, M. (2014). "It's the way you check-in: identifying users in location-based social networks," in *Proceedings of the Second ACM Conference on Online Social Networks (COSN '14)* (Dublin), 215–226. doi: 10.1145/2660460.2660485

Rossi, L., Musolesi, M., and Torsello, A. (2015a). "On the k-anonymization of time-varying and multi-layer social graphs," in *Ninth International AAAI Conference on Web and Social Media* (Oxford).

Rossi, L., Williams, M., Stich, C., and Musolesi, M. (2015b). "Privacy and the city: user identification and location semantics in location-based social networks," in *Ninth International AAAI Conference on Web and Social Media* (Oxford).

Rousseau, F., Casas-Roma, J., and Vazirgiannis, M. (2018). Community-preserving anonymization of graphs. *Knowl. Inform. Syst.* 54, 315–343. doi: 10.1007/s10115-017-1064-y

Rozemberczki, B., Davies, R., Sarkar, R., and Sutton, C. (2018). Gemsec: Graph embedding with self clustering. *arXiv preprint arXiv*:1802.03997.

Sperotto, A., and Pelillo, M. (2007). "Szemerédi's regularity lemma and its applications to pairwise clustering and segmentation," in *Proceedings of the 6th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR'07)* (Ezhou), 13–27.

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature* 393, 440. doi: 10.1038/30918

Zhou, B., and Pei, J. (2011). The k-anonymity and l-diversity approaches for privacy preservation in social networks against neighborhood attacks. *Knowl. Inform. Syst.* 28, 47–77. doi: 10.1007/s10115-010-0311-2

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.