(k, ℓ)-anonymity of networks via k-metric antidimension: bridging graph theory & mathematical optimization*

Elena Fernández
Departamento de Estadística de Investigación Operativa / Universidad de Cádiz

Joint work with:
Dorota Kuziak, Manuel Munoz-Márquez
Departamento de Estadística e Investigación Operativa / Universidad de Cádiz

Ismael G. Yero
Departamento de Matemática Aplicada / Universidad de Cádiz

June 6, 2023

1 Introduction

Social network analysis is the process of investigating social structures while making use of networks and graph theory methods. Such analysis is widely developed in our society, motivated by several factors, including the increasing need of advances in basic computing and information technologies. These advances aim at improving systems and services frequently integrated within a variety of important business and societal functions, like e-commerce, health care, education, manufacturing, and personal interactions, among others. The benefits derived from these investigations are, however, not cost-free, as the privacy of the users of a network would be compromised if some involved entity could deliver sensitive data such as e-mails, instant messages, or relationships. A basic solution to the above issue is to apply some anonymization process to the released social network. Unfortunately, naive approaches are usually not enough to guarantee the privacy of users personal information in a network. In fact, there is always some potential probability of disclosing any user. Therefore, it is highly desirable to provide any public social network with a measure of its *disclosing* probability.

The (k, ℓ)-anonymity, introduced in [6] and extended in [3], is a measure quantifies the above probability. Specifically, a social graph achieving (k, ℓ)-anonymity satisfies that the probability of disclosing any given user of that network, in the presence of at most ℓ attacker nodes in the network, is $1/k$. The (k, ℓ)-anonymity is theoretically supported by a graph theory parameter called k-metric antidimension, which was introduced in [6] as well. The integer k is used as a privacy threshold, whereas the value ℓ stands for an upper bound on the expected number of attacker vertices in a

*Partially supported by by the Spanish Agencia Estatal de Investigacid European Regional Development Funds (ERDF) through MINECO PID project MTM2019-105824GB-I00, and by the Plan Propio - UCA 2022-2023.

given network. Since an attacker entity cannot easily control many vertices of the network, it is usually accepted that the number of attacker nodes in a network is likely to be significantly smaller than the total number of vertices, so it is assumed that ℓ is a small integer number. Given the tight relationship between the above two concepts, a key point for stating the (k, ℓ)-anonymity of a given graph G is finding its k-metric antidimension.

The k-metric antidimension of some classical families of graphs like complete bipartite, cycles, and some others, was considered in the seminal paper [6]; that of some generalized Petersen graphs was studied in [2]; of some wheel-related social graphs in [4]; and of some trees and unicyclic graphs (for $k = 1$) in [5]. New results have been recently developed in [1] for graph classes for which this parameter remained unknown, namely cylinders, toruses, and 2-dimensional Hamming graphs. Still, for classes of graphs without a predefined structure, or with a more intricate structure than those above mentioned, the usefulness of graph theoretical approaches maybe be very limited thus enhancing the need for alternative approaches.

In this talk I will focus on a new perspective for addressing this topic, which is of particular interest for classes of graphs without a predefined structure. Specifically, we will see how the (k, ℓ)-anonymity of any given graph can be computed using mathematical optimization tools. For this I will present an integer programming formulation for the k-metric antidimension problem. I will also present and analyze the numerical results obtained in a series of computational experiments with several types of randomly generated graphs: trees, general sparse graphs and, dense graphs. The obtained results indicate that random trees and general sparse graphs achieve low privacy properties, whereas random general dense graphs exhibit higher privacy properties. Similarly to 2-dimensional Hamming graphs, this can be due to the fact that dense graphs have more "near-symmetrical" properties than the other ones, and also that they intuitively should have small diameters. As a general conclusion concerning the (k, ℓ)-anonymity, it seems that higher privacy properties appear in graphs with more symmetry, larger degrees and smaller diameters.

References

[1] Fernández, E., D. Kuziak, M. Munoz-Marquez, I.G. Yero. On the (k, ℓ)-anonymity of networks via their k-metric antidimension. `https://arxiv.org/abs/2304.00849`.

[2] Kratica, J., Kovačevič-Vujčić, V., and Čangalović, M. (2019). k-metric antidimension of some generalized Petersen graphs, Filomat 33(13), 4085–4093.

[3] Mauw, S., Ramírez-Cruz, Y., and Trujillo-Rasua, R. (2019). Conditional adjacency anonymity in social graphs under active attacks, Knowledge and Information Systems 61(1), 485–511.

[4] Tang, J. H., Noreen, T., Salman, M., Rehman, M. U., and Liu, J. B. (2021). (k, ℓ)-anonymity in wheel-related social graphs measured on the base of k-metric antidimension, Journal of Mathematics, vol 2021, Article ID 8038253.

[5] Trujillo-Rasua, R., and Yero, I. G. (2016). Characterizing 1-metric antidimensional trees and unicyclic graphs, Computer Journal 59(8), 1264–1273.

[6] Trujillo-Rasúa, R., and Yero, I. G. (2016). k-metric antidimension: A privacy measure for social graphs, Information Sciences 328, 403–417.