Discrete Applied Mathematics 188 (2015) 25–31

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Graphic sequences, distances and k-degree anonymity

Julián Salas [a,*], Vicenç Torra [a,b]

[a] IIIA-CSIC, Consejo Superior de Investigaciones Científicas, Institut d'Investigació en Intel·ligència Artificial, Campus Universitat Autònoma de Barcelona, Spain

[b] School of Informatics University of Skövde, Skövde, Sweden

ARTICLE INFO

Article history:
Received 26 September 2013
Received in revised form 26 September 2014
Accepted 10 March 2015
Available online 6 April 2015

Keywords:
Degree sequence
Edge rotation distance
Rectilinear distance
Regular graph
Graph distance
Graphic sequence
k-anonymity
k-degree anonymity

ABSTRACT

In this paper we study conditions to approximate a given graph by a regular one. We obtain optimal conditions for a few metrics such as the edge rotation distance for graphs, the rectilinear and the Euclidean distance over degree sequences. Then, we require the approximation to have at least k copies of each value in the degree sequence, this is a property proceeding from data privacy that is called k-degree anonymity.

We give a sufficient condition in order for a degree sequence to be graphic that depends only on its length and its maximum and minimum degrees. Using this condition we give an optimal solution of k-degree anonymity for the Euclidean distance when the sum of the degrees in the anonymized degree sequence is even. We present algorithms that may be used for obtaining all the mentioned anonymizations.

1. Introduction

For modelling applications, it is very important to know the distance between isomorphism classes of graphs. At present different distances have been defined and used in the literature. For example in organic chemistry, the maximum common subgraph distance was used to reflect the principle of minimal structural change, cf. [2,11,18]. Another measure used for graph similarity is the edge rotation distance defined by Chartrand, Saba and Zou [3]. This is also known as the reaction metric. In [5], the authors consider the rotation distance and present bounds on this distance between two graphs in terms of their greatest common subgraphs.

Three different metrics derived from edge manipulation are studied in [7]. These are the edge move distance, the edge slide distance and the edge rotation distance. It is also observed a relation between edge rotation distance between two graphs and the rectilinear distance between its degree sequences.

In this work we consider the related problem of graph approximation. That is given a graph G and a distance d find the graph G' at minimum distance from G that satisfies a set of constraints. More specifically we consider the problem of approximating a graph with regular graphs and we obtain the best approximations under the Euclidean and rectilinear metrics between degree sequences and also under the edge rotation distance for graphs.

A similar but more complicated problem is that of approximating with a graph in which the degrees of its vertices have frequency at least k, this is a property proceeding from data privacy that is called k-degree anonymity.

* Corresponding author.
E-mail addresses: julian.salas@iiia.csic.es (J. Salas), vtorra@his.se, vtorra@iiia.csic.es (V. Torra).

http://dx.doi.org/10.1016/j.dam.2015.03.005

In data privacy, a data satisfies k-anonymity with respect to a query when for any value in the domain of the result of the query, there are at least k individuals with that value, cf. [14,16].

The concept of k-anonymity comes from the need to publish data for research or practical purposes, while preserving the privacy of the individuals. It was a common practice for organizations to release person specific data after only removing all explicit identifiers, such as name, or social security number. But this has been proven insufficient, since the combinations of few characteristics that might have been obtained from different data sources, often can be used to link an individual to a data record and therefore disclose private information.

Hence, in order to avoid this re-identification, Samarati and Sweeney proposed the concept of k-anonymity, in which a record of an individual in a database is cloaked among $k-1$ other equal records.

After the proliferation of social networks the concept of anonymity was extended to graphs in order to model the privacy issues that arise from its publication.

In [1], a model of attacks focused on anonymized social networks is proposed to learn whether edges exist or not between given targeted pairs of nodes, i.e. to know if there is a relation between a pair of subjects.

In [9], the authors define a class of queries $\mathcal{H}_i(x)$, of increasing power i, which report on the local structure of the graph around a node. The authors state that these queries are inspired by iterative vertex refinement, a technique originally developed to efficiently test for the existence of graph isomorphisms. The queries $\mathcal{H}_i(x)$ correspond to the degree of the neighbours at distance $i-1$ from vertex x, hence $\mathcal{H}_1(x)$ correspond to the vertex degree.

In [12], the k-anonymization of the class of queries $\mathcal{H}_1(x)$ is addressed, and it is called k-degree anonymization. The authors define the graph anonymization problem as finding a k-anonymous supergraph of a given graph G, and relate the optimal cost of an anonymization to the rectilinear distance δ_1. Note that this is not the only type of k-anonymization considered in graphs, other definitions exist to formalize other types of attacks, e.g. [15].

In this work, we further study the problem of finding a k-degree anonymous graph with two other metrics for degree sequences and graphs. They are the Euclidean distance δ_2 and the edge rotation distance between graphs δ_{ER}.

The optimal solution for vectors can be solved in polynomial time, cf. [8], by using microaggregation, a widely used disclosure control technique, in which records are aggregated into groups and the mean of the group to which an observation belongs is released, see [13].

This solution is not always meaningful for degree sequences, since the result may not be a degree sequence of a graph (i.e. a graphic sequence). For example, when the sum of all the values in the sequence is odd it cannot be graphic.

We give a new condition on the maximum and minimum values of a degree sequence that implies that the sequence is graphic, this can be used to guarantee that an optimal k-anonymization with respect to the metric δ_2 can be obtained. It is worth noting that all these results may be applied to all kinds of graphs not only for social networks. In this paper we look for theoretical results to find conditions for the optimal k-degree anonymity. The literature in data privacy discusses some approximations that have been used in experiments with some available graphs. See [17,4] for details.

In this work, we review in Section 2 the main definitions needed in the rest of the paper. In Section 3, we consider the problem of approximating a graph with regular graphs and we obtain the best approximations under the Euclidean and rectilinear metrics between degree sequences and also under the edge rotation distance for graphs. In Section 4, we consider the problem of approximating with degrees of frequencies at least k, that is the case when k-degree anonymity is required. This section contains the main results, a sufficient condition in order for a degree sequence to be graphic that depends only on its length and its maximum and minimum degrees. An approximation of the k-degree anonymity problem when the subsets are all of size k except one. An optimal approximation for subsets of any size greater than k, or at most an approximation at distance $4k^2$ to the optimal with the metric δ_2 when the degree sequence for the optimal partition is not graphic.

In the last section we present sketches of algorithms based on our results that can be used to obtain k-degree anonymous sequences. The paper finishes with a summary.

2. Definitions

Let G denote a simple graph, i.e. without loops or multiple edges, $V(G)$ its vertex set and $E(G)$ its edge set.

We consider that all degree sequences are in non-increasing order, i.e. if $s=(d_1, d_2, \dots, d_n)$ is the degree sequence of a graph G, then $d_1 \geq d_2 \geq \cdots \geq d_n$, and we will denote by v_i the vertex of degree d_i, that is $d(v_i)=d_i$. By adding a superindex we will denote the *frequency* of the degree, e.g. $(d_1^{n_1}, d_2^{n_2}, \dots, d_l^{n_l})$ is a sequence in which each degree d_i has n_i repetitions. We will denote $d_{G'}(x)$ as $d'(x)$, and $d_{G''}(x)$ as $d''(x)$.

By considering the degree sequences of graphs with n vertices as vectors in the discrete Euclidean space $\mathbb{N}^n \subset \mathbb{R}^n$, they inherit the *Euclidean distance* $\delta_2(s, s') = \sum_{i=1}^{n}(d_i - d_i')^2$, where $s=(d_1, d_2, \dots, d_n)$ is the degree sequence of a graph G and $s'=(d_1', d_2', \dots, d_n')$ the degree sequence of a graph G'. By measuring the distance coordinate by coordinate, the *rectilinear distance* is obtained $\delta_1(s, s') = \sum_{i=1}^{n}|d_i - d_i'|$.

The distance between two graphs G and G' can also be measured by the number of edge operations that must be done to G in order to obtain the graph G', if the operation is a rotation then we refer to the edge rotation distance δ_{ER}.

Let G and G' be two graphs having the same order $n=|V(G)|=|V(G')|$ and size $q=|E(G)|$, $q'=|E(G')|$, not necessarily equal. The *edge rotation distance* $\delta_{ER}(G, G')$ is defined as 0 if $G \cong G'$ and, otherwise, as the smallest positive integer r for which