# The Graph Clustering Problem has a Perfect Zero-Knowledge Proof

Oded Goldreich
Department of Computer Science and Applied Mathematics
Weizmann Institute of Science, Rehovot, Israel.
E-mail: oded@wisdom.weizmann.ac.il

November 3, 1996

**Abstract**

The Graph Clustering Problem is parameterized by a sequence of positive integers, $m_1, ..., m_t$. The input is a sequence of $\sum_{i=1}^{t} m_i$ graphs, and the question is whether the equivalence classes under the graph isomorphism relation have sizes which match the sequence of parameters. In this note we show that this problem has a (perfect) zero-knowledge interactive proof system.

**Keywords:** Graph Isomorphism, Zero-Knowledge Interactive Proofs.

## 1 Introduction

For many years, the Graph Clustering Problem (defined below), has been my favorite example for a concrete problem having low (but non-zero) knowledge-complexity (cf., [4, 3]). However, reconsidering the problem a few weeks ago, I've realized that current "state of the art" (specifically, the paper of De-Santis et. al. [1]) yields that this problem does have zero knowledge-complexity.

## 2 The Graph Clustering Problem

The Graph Clustering Problem (GCP) is parameterized by a sequence of positive integers, $m_1, ..., m_t$. Let $m \overset{\text{def}}{=} \sum_{i=1}^{t} m_i$. Fixing these parameters the problem is specified as follows:

**input:** $m$ Graphs, denoted $G_1, ..., G_m$.

Without loss of generality we may assume all have $[n] \overset{\text{def}}{=} \{1, ..., n\}$ as their vertex set.

**question:** Does there exist a partition, $C_1, ..., C_t$, of $[m]$ so that $|C_i| = m_i$ for $i = 1, .., t$ and

- For every $i \in [t]$ and every $j, k \in C_i$, the graphs $G_j$ and $G_k$ are isomorphic.
- For every $i \neq j \in [t]$ and every $k \in C_i$ and $h \in C_j$, the graphs $G_k$ and $G_h$ are not isomorphic.

That is, $C_1, ..., C_t$ are the equivalent classes under the graph-isomorphism relation and their sizes match the $m_i$'s.

Let us denote this problem by $\text{GCP}_{m_1,...,m_t}$. Note that $\text{GCP}_2$ and $\text{GCP}_{1,1}$ correspond to the Graph Isomorphism and Graph Non-Isomorphism problems, respectively. Both are known to have perfect zero-knowledge proof systems [2].

## 3 The Zero-Knowledge Proof

The main tools we use are two results due to De-Santis et. al. [1]. In their paper the following problem parameterized by a Boolean formula $\Psi$ and a language $L$ is considered, where $k$ denotes the number of variables in $\Psi$:

**input:** $k$ instances, denoted $x_1, ..., x_k$.

**question:** Does $\Psi(\chi_L(x_1), ..., \chi_L(x_k)) = 1$ hold, where $\chi_L$ is the Characteristic function of $L$ (i.e., $\chi_L(x) \stackrel{\rm def}{=} 1$ if $x \in L$ and 0 otherwise).

Let us denote the above problem by $\mathcal{CL}_{L,\Psi}$. Also, let GI denote the set of pairs of isomorphic graphs. We use two of the results of [1]:

1. For every monotone formulae $\Psi$, the language $\mathcal{CL}_{\mathrm{GI},\Psi}$ has a (perfect) zero-knowledge proof system.

2. For every integer $u$, the language $\mathcal{CL}_{\mathrm{GI},T_u}$ has a (perfect) zero-knowledge proof system, where $T_u$ is the threshold function which is 1 iff there are at most $u$ 1's in the input.

Our (perfect) zero-knowledge proof for $\mathrm{GCP}_{m_1,...,m_t}$ follows by the observation that this problem is reduced to the AND of two $\mathcal{CL}_{\mathrm{GI},\cdot}$ problems, one of Type (1) and the other of Type (2). Specifically, let $k = \binom{m}{2}$ and consider a standard enumeration of all $k$ (unordered) pairs of distinct integers in $[m]$. Let $\{i_1, i_2\}$ be the $i^{\rm th}$ pair in this enumeration and define $x_i = (G_{i_1}, G_{i_2})$. Then

$$\mathrm{GCP}_{m_1,...,m_t}(G_1, ..., G_m) \;=\; \mathcal{CL}_{\mathrm{GI},\Psi}(x_1, ..., x_k) \wedge \mathcal{CL}_{\mathrm{GI},T_u}(x_1, ..., x_k)$$

where $u = \sum_{i=1}^{t} \binom{m_i}{2}$ and $\Psi$ is an adequate monotone formulae. The obvious question is whether the adequate $\Psi$ does exist. The answer is indeed in the affirmative: $\Psi$ is the disjunction of formulae $\Psi_{C_1,...,C_t}$, for all partitions $C_1, ..., C_t$ of $[m]$ which satisfy $|C_i| = m_i$ for all $i = 1, .., t$. The formulae $\Psi_{C_1,...,C_t}$ is true if the instances corresponding to pairs in any cluster are indeed in the Graph-Isomorphism language. That is

$$\Psi_{C_1,...,C_t}(\sigma_1, ..., \sigma_k) \;=\; \bigwedge_{j \in [t]} \bigwedge_{i_1,i_2 \in C_j} \sigma_{\{i_1,i_2\}}$$

The threshold formula $T_u$ makes sure that there are no additional pairs of isomorphic graphs.

**Comments:** Reduction to Threshold formulae suffices as long as $m \leq 5$ (since each partition of such $m$'s into $m_i$'s has a distinct value for $\sum_i \binom{m_i}{2}$). But for $k = 6$ both $6 = 2 + 2 + 2$ and $6 = 3 + 1 + 1 + 1$ have the same value for $\sum_i \binom{m_i}{2}$ (i.e., 3). On the other hand, our result can be proven using other tools in [1]; for example, the analogous proof systems for closures of Graph Non-Isomorphism.

# References

[1] A. De-Santis, G. Di-Crescenzo, G. Persiano and M. Yung, "On Monotone Formula Closure of SZK", *35th FOCS*, pp. 454–465, 1994.

[2] O. Goldreich, S. Micali and A. Wigderson, "Proofs that Yield Nothing But Their Validity or All Languages in NP Have Zero-Knowledge Proof Systems", *JACM*, Vol. 38, No. 1, pp. 691–729, 1991.

[3] O. Goldreich, and E. Petrank, "Quantifying Knowledge Complexity", *32nd FOCS*, pp. 59–68, 1991. A full version is available from `http://theory.lcs.mit.edu/~oded/`.

[4] S. Goldwasser, S. Micali and C. Rackoff, "The Knowledge Complexity of Interactive Proof Systems", *SIAM J. Comput.*, Vol. 18, No. 1, pp. 186–208, 1989.