Available online at www.sciencedirect.com
ScienceDirect
Discrete Applied Mathematics 156 (2008) 1399–1406
DISCRETE APPLIED MATHEMATICS
www.elsevier.com/locate/dam

# Reconstruction of a graph from 2-vicinities of its vertices☆

Vladimir Levenshtein[a], Elena Konstantinova[b], Eugene Konstantinov[c], Sergey Molodtsov[d]

[a] *Keldysh Institute for Applied Mathematics, Russian Academy of Sciences, Miusskaya sq. 4, 125047 Moscow, Russia*
[b] *Sobolev Institute of Mathematics, Siberian Branch of Russian Academy of Sciences, Pr. Koptuyga 4, 630090 Novosibirsk,Russia*
[c] *Budker Institute of Nuclear Physics, Siberian Branch of Russian Academy of Sciences, Pr. Lavrent'eva 11, 630090 Novosibirsk, Russia*
[d] *Vorozhtsov Institute of Organic Chemistry, Siberian Branch of Russian Academy of Sciences, Pr. Lavrent'eva 9, 630090 Novosibirsk, Russia*

Received 21 January 2004; received in revised form 24 September 2004; accepted 2 November 2006
Available online 30 August 2007

**Abstract**

We prove that a connected graph of diameter at least 4 and of girth 7 or more (in particular, a tree) can be exactly reconstructed from metric balls of radius 2 of all its vertices. On the other hand, there exist graphs of diameter 3 and of girth 6 which are not reconstructible. This new graph theory problem is motivated by reconstruction of chemical compounds.


*Keywords:* Graph reconstruction; 2-Vicinity; Girth; Neighbourhood; Graph of neighbourhood; Reconstruction algorithm

## 1. Introduction

What do we know about reconstruction of graphs? There are two famous conjectures which are based on the decks of a graph. For a finite simple graph $G$, the deletion of an edge produces an edge-deleted subgraph of $G$ and the multiset of the edge-deleted subgraphs of $G$ is called the edge deck of $G$. The vertex-deleted subgraphs and the vertex deck of a graph are defined similarly taking into account that a vertex is deleted together with all incident edges. Kelly's [6,7] and Ulam's [15] vertex reconstruction conjecture states that a graph of order at least 3 is uniquely determined, up to isomorphism, by its vertex deck and Harary's [5] edge reconstruction conjecture states that a graph with at least four edges is uniquely determined by its edge deck. For details see the survey paper by Bondy [3].

Representation of distance matrices by trees and realisation of degree sequences give examples of the solved problems of graph reconstruction. The question "when can entries of a given matrix be realised as distances between the terminal vertices of some tree?" was formulated by Smolenski [14] and Zaretski [16]. They gave necessary and sufficient conditions for the unique reconstruction of trees. The problem "when is a given sequence of integers realizable as a degree sequence of some graph?" was stated by Erdös and Gallai in 1960 [4]. They obtained the necessary and sufficient conditions for the existence of a graph with a given degree sequence.

A new problem to restore an unknown sequence from a sufficiently large number of its erroneous samples was recently introduced and investigated by Levenshtein in [10]. The corresponding reconstruction problem can be treated as the following graph theoretical problem. Given a graph $G = (V, E)$ with a vertex set $V$ and an edge set $E$, and an

☆ This work was supported by the RFBR Grant 04-01-00122 and partially supported by the RFBR Grant 06-01-00694.
*E-mail address:* e_konsta@mail.ru (E. Konstantinova).

0166-218X/$ - see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.dam.2006.11.016

integer $t$, what is the minimum number $N(G, t)$ of vertices in the metric balls $B_t(x, G)$ of radius $t$ centred at $x \in V$ which is sufficient to uniquely reconstruct an arbitrary $x$? What is a simple algorithm which allows one to reconstruct this unknown $x$ from $N(G, t)$ different vertices of $B_t(x, G)$? These problems were solved in [9–11] for graphs whose path metric is induced by some sort of errors of essential interest in coding theory such as substitutions, transpositions, deletions and insertions of symbols.

In this paper we consider another new graph reconstruction problem which has a quite different nature as compared with the problem described above. We are interested in reconstruction of an unknown graph $G$ from local information on metric balls of all vertices of this graph. The problem can be formulated for a family $F$ of graphs with the same vertex set $V$ as follows. For an unknown graph $G = (V, E) \in F$ one knows the set of all metric balls $B_t(x)$, $x \in V$. Can this graph $G$ be reconstructed, exactly or up to isomorphism, from these sets $B_t(x)$, $x \in V$? What is a simple algorithm which allows one to reconstruct this unknown graph $G$ if it is possible?

This problem was motivated by the problem arising in chemistry: the structure elucidation of unknown compounds from nuclear magnetic resonance (NMR) spectroscopy data. NMR is an electronic method of research based on spin precession in magnetic field [12]. Spin is a vector value of the internal momentum of atomic nuclei and it precesses in applied external magnetic fields. Fortunately, carbon and hydrogen atoms possess non-zero spin thus it is possible to detect a signal coming from organic compounds. The frequency of this precession is defined by magnetic field value and belongs to the radio region. Spin movements in molecules depend on atomic bonds. A molecule appears to be a set of bells in this model so all these bells could be excited in a moment and the picture of the oscillations decay contains the molecule's structural information. NMR spectrometer is equipped with a pulse transmitter to excite the sample molecules, and a sensitive radio receiver to obtain the oscillation spectrum. In order to systematise NMR spectra the precession frequencies are normalised to the magnetic field so that there is a correspondence between every atom of carbon and hydrogen and some chemical shift (normalised position of the corresponding signal) in a NMR spectrum. Chemical shift is determined by what the corresponding atom is bonded to: carbon, hydrogen or a group of atoms. Correlations of shifts allow to find the atoms positioned a short distance (in number of bonds) from each other and the actual distance between them. It is possible to specify the atoms connected to each other in some cases. But in most cases the specified atoms are at a distance not longer than $k$ bonds. Further reconstruction of all structural formulae of chemical compounds follows the rule that the calculated distance between atoms does not contradict the given NMR data. As it is known, the structural formula can be represented by a molecular graph (labelled multigraph) whose vertices correspond to atoms, and multiple edges to bonds of a chemical compound. Thus, the problem of reconstruction of molecular graphs with the set of possible distances between some vertices arises [2,8].

Our main goal is to find necessary and sufficient conditions for exact reconstruction of an unknown graph by metric balls of a fixed radius $t$ with centres in all vertices. It is clear that any graph can be reconstructed by metric balls of radius 1. However, in general, it is not the case for $t = 2$. The main task of the paper is to prove that any connected graph of diameter at least 4 and of girth at least 7 (in particular, a tree) can be exactly reconstructed from metric balls of radius 2 of all vertices. We also give a simple algorithm for reconstruction of such a graph and show that the conditions above are in general necessary for exact reconstruction.

## 2. Definitions and notations

Let $G = (V, E)$ be a connected graph with the vertex set $V = \{1, 2, \ldots, n\}$ and edge set $E$. The *distance* $d_G(x, y)$ between vertices $x$, $y$ in a graph $G$ is the length of the shortest path that connects vertices $x$ and $y$. The *diameter* $d(G)$ of a graph $G$ is the maximum distance between vertices of $G$. The *girth* $g(G)$ of $G$ is the length of the shortest cycle in $G$ if G contains cycles. We put $g(G) = \infty$, if $G$ has no cycles (i.e., $G$ is a tree). To avoid trivial cases we will assume that $d(G) \geqslant 2$ and $g(G) \geqslant 3$ and hence $G$ does not contain loops and parallel edges. The *subgraph of $G$ induced by a subset $A \subseteq V$* is the graph with the vertex set $A$ whose edges are the edges of $G$ joining vertices of $A$. We denote this subgraph by $G(A)$. The cardinality of the set of all edges incident with a vertex $x$ is called *the degree* of the vertex $x$. A vertex of degree one is called a *terminal vertex*. A connected graph $G$ having only one vertex $x$ which is not terminal is called a *star centred at $x$*. The subgraph $G(V')$ induced by the subset $V'$ of all non-terminal vertices of $G$ will be denoted by $G'$ and called *the frame* of $G$. The frame of a connected graph is a connected graph. For any $x \in V$ and non-negative integer $t$ put

$$B_t(x, G) = \{y \in V : d_G(x, y) \leqslant t\}. \tag{1}$$