



Fifty years of graph matching, network alignment and network comparison



Frank Emmert-Streib^{a,*}, Matthias Dehmer^{b,*}, Yongtang Shi^c

^a Predictive Medicine and Analytics Lab, Department of Signal Processing, Tampere University of Technology, 33720 Tampere, Finland

^b Department of Computer Science, Universität der Bundeswehr München, 85577 Neubiberg, Germany

^c Center for Combinatorics and LPMC-TJKLC, Nankai University, Tianjin, PR China

ARTICLE INFO

Article history:

Received 5 February 2015

Revised 16 January 2016

Accepted 23 January 2016

Available online 4 February 2016

Keywords:

Network comparison

Network similarity

Graph matching

Quantitative graph theory

Biological networks

Computational graph theory

ABSTRACT

In this paper we survey methods for performing a comparative graph analysis and explain the history, foundations and differences of such techniques of the last 50 years. While surveying these methods, we introduce a novel classification scheme by distinguishing between methods for deterministic and random graphs. We believe that this scheme is useful for a better understanding of the methods, their challenges and, finally, for applying the methods efficiently in an interdisciplinary setting of data science to solve a particular problem involving comparative network analysis.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

The comparison of graphs has its roots in the 60s and 70s marked by the seminal work of Sussenguth [108], Vizing [120] and Zelinka [125]. Surprisingly, Sussenguth [108] already tackled the problem of comparing (chemical) graphs computationally in 1964 by developing an algorithm to determine the graph isomorphism studying small graphs. In contrast, Vizing [120] in 1968 did not define quantitative measures for comparing graphs but just raised the importance of the graph comparison problem, which was in 1975 analytically investigated by Zelinka [125] who was the first to quantify the distance between graphs [125], based on determining a graph isomorphism [53], assuming the two graphs have the same number of vertices. Later, Sobik [101] and Kaden [62] generalized Zelinka's work in various directions.

Studying and comparing the structural representations in the form of complex networks has been a very fruitful and fascinating research area [3,8,25,101]. Other examples relate to studying the growth mechanism of the world wide web graph by using stochastic and random graph models [1]. In particular, random graphs have been proven useful when investigating their mathematical growth properties, including phase transitions [42] or as a tool to predict structural properties of real world networks [78]. Another research topic is the investigation of network classes such as *small world* and scale-free networks and combinations thereof [78]. To date, many techniques have been developed to compare real-world graph patterns structurally; concrete examples thereof are methods applied in linguistics [26], web mining [24], chemoinformatics [123], and computational biology [38].

* Corresponding authors. Tel.: +358 33115 5353.

E-mail addresses: v@bio-complexity.com (F. Emmert-Streib), matthias.dehmer@unibw.de (M. Dehmer).

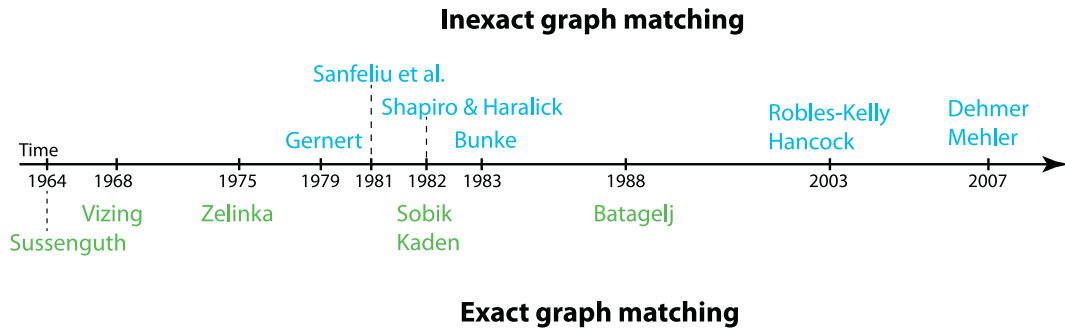


Fig. 1. Time line of important contributions to inexact and exact graph matching methods for deterministic networks.

When analyzing complex networks on a structural basis, comparative methods using similarity (or dissimilarity) measures [94,115] are crucial [8,27,101,125]. Depending on the discipline, this problem is referred to as graph matching, network comparison or network alignment and relates to comparing graphs by either finding isomorphic relations (*exact* graph matching) [12,101,125] or by using *inexact* (approximative) techniques for matching graphs in an error-tolerant way [12,15]. In Fig. 1 we show a time line highlighting important contributions in either paradigm. All of these methods were developed for deterministic networks—see Section 3 for a discussion.

In general, measuring the structural similarity of large graphs is a challenging undertaking because of the often unfavorable computational complexity of the analysis methods. This is an important issue, because graph matching techniques [12] have found ample applications in various scientific disciplines, e.g., artificial intelligence and pattern recognition [22], biology [36,61], chemoinformatics [86,87,117], neuroscience [105,106], computational linguistics [73], image recognition [58,110], machine learning [6,57] and web mining [23,27]. The importance of graph matching, network comparison and network alignment methods stems from the fact that such considerably different phenomena can be represented with the same mathematical concept forming part of what is nowadays called *network science*. Furthermore, by quantifying differences in networks the application of many standard machine learning and statistics methods are enabled relying on the availability of such metrics.

A major contribution of this paper is to survey graph matching, network comparison and network alignment methods together side-by-side, because, usually, methods from these topics are reviewed in separation due to interdisciplinary boundaries of the respective communities. For instance, graph matching methods have been reviewed in [12,22,46] and network alignment methods are discussed in [99]. However the mentioned reviews about graph matching are either narrowly focused on a particular class of methods or a discipline. In this sense, Bunke [12] puts the emphasis on inexact graph matching methods applied to image and video indexing but, for example, does not discuss classical contributions dealing with graph isomorphism (exact graph matching). Conte et al. [22] discussed graph matching techniques for image, document and video analysis and the presented taxonomy in [22] is solely focused on algorithms from these areas and also does not discuss the complexity of graph matching. Gao et al. [46] focused on surveying results for the graph edit distance only. However, we think that a reader may profit tremendously from a boundary-free, interdisciplinary presentation receiving in this way the most comprehensive overview possible that allows a cross fertilization of ideas and concepts for comparing networks. Another major difference to previous reviews is that in our paper we assume a data-centric view that means we highlight problem domain specific characteristics of networks in its most basic way by distinguishing between deterministic and random networks, as explained in detail in Section 3.

This paper is organized as follows. In the next section we provide a clear description what we mean by a ‘comparative network analysis’ and in Section 3 we give a general overview of the methods discussed in this paper. In Section 4 we review methods for deterministic and in Section 5 methods for random networks. The paper finishes in Section 6 with a conclusion.

2. Comparative network analysis of bivariate measures

In order to be precise what we mean by a ‘comparative network analysis’ we provide some definitions that clarify the scope of the following discussions.

Definition 1. Given two networks $G_1 \in \mathcal{G}_1$ and $G_2 \in \mathcal{G}_2$. A measure ‘M’ having two arguments, denoted by $M(G_1, G_2)$, performing a mapping from $\mathcal{G}_1 \times \mathcal{G}_2 \rightarrow \mathbb{R}$, is called a ‘bivariate network measure’. $M(G_1, G_2)$ is either the similarity or the dissimilarity (distance) between two graphs, hence, we use the symbol $s(G_1, G_2)$ to indicate the similarity and $d(G_1, G_2)$ for the distance.

Definition 2. Investigating a bivariate network measure for finite or infinite sets \mathcal{G}_1 and \mathcal{G}_2 is called a ‘comparative network analysis’ and computing $s(G_1, G_2)$ or $d(G_1, G_2)$ is called network comparison.

Download English Version:

<https://daneshyari.com/en/article/392420>

Download Persian Version:

<https://daneshyari.com/article/392420>

[Daneshyari.com](https://daneshyari.com)