Efficient Algorithms for Attributed Graph Alignment with Vanishing Edge Correlation

Ziao Wang
Department of Electrical and Computer Engineering, The University of British Columbia
ZIAOW@ECE.UBC.CA

Weina Wang
Department of Computer Science, Carnegie Mellon University
WEINAW@CS.CMU.EDU

Lele Wang
Department of Electrical and Computer Engineering, The University of British Columbia
LELEWANG@ECE.UBC.CA

Abstract

The graph alignment problem, also referred to as the graph matching or noisy graph isomorphism problem, is the problem of finding the correspondence between the vertices of two correlated graphs. This problem has garnered significant attention due to its widespread use in various real-world applications (Singh et al., 2007; Cho and Lee, 2012). Extensive study has been done on polynomial-time algorithms for the graph alignment problem under the Erdős–Rényi graph pair model $G(n, q, \rho)$, where the two graphs are Erdős–Rényi graphs with $n$ vertices and edge probability $q$, correlated under certain vertex correspondence. The correlation coefficient between the corresponding edges in the two graphs is given by $\rho$. Given a pair of graphs generated from the model, the goal is to exactly recover the vertex correspondence between the two graphs. Polynomial-time algorithms for graph alignment have been studied in the line of work (Dai et al., 2019; Fan et al., 2020; Ding et al., 2021; Mao et al., 2023a, 2021, 2023b; Ding and Li, 2023). To achieve exact recovery of the correspondence, all existing algorithms at least require the edge correlation coefficient $\rho$ between the two graphs to be non-vanishing as $n \to \infty$. Moreover, it is conjectured that no polynomial-time algorithm can achieve exact recovery under vanishing edge correlation $\rho < 1/\text{polylog}(n)$.

In this work, we study the attributed graph alignment problem (Zhang et al., 2024), where additional side information about the vertices, termed attributes, is assumed to be publicly available. The consideration of attribute information is motivated by real-world applications such as LinkedIn and Twitter, where user attributes like birthplace and educational background can aid alignment. We show that with a vanishing amount of additional attribute information, exact recovery of the vertex correspondence is polynomial-time feasible under vanishing edge correlation $\rho \geq n^{-\Theta(1)}$. The key step in the proposed algorithm is a tree-counting process. We identify a family of local tree structures to count, which incorporates one layer of user information and one layer of attribute information. We construct a feature vector for each vertex by counting the trees in the family. The vertices are then matched based on the similarity of their feature vectors. Compared to existing polynomial-time algorithms for attributed graph alignment by Wang et al. (2024), the proposed algorithm requires strictly less attribute information and tolerates weaker edge correlation. When specialized to the problem of seeded graph alignment, the proposed algorithm strictly improves the best-known feasible region by polynomial-time algorithms (Mossel and Xu, 2020).\footnote{Extended abstract. Full version appears as [arXiv 2308.09210, v2 (Wang et al., 2023)]}

Keywords: Graph matching, statistical inference, efficient algorithms
References


