Journal of King Saud University – Computer and Information Sciences (2016) **28**, 230–246

King Saud University

Journal of King Saud University – Computer and Information Sciences

www.ksu.edu.sa
www.sciencedirect.com

Maximal assortative matching for complex network graphs

Natarajan Meghanathan

Jackson State University, Jackson, MS 39217, United States

Received 27 May 2015; revised 12 October 2015; accepted 30 October 2015
Available online 2 December 2015

KEYWORDS

Maximal matching;
Assortative matching;
Dissortative matching;
Assortativity index;
Complex networks;
Node similarity

Abstract We define the problem of maximal assortativity matching (MAM) for a complex network graph as the problem of maximizing the similarity of the end vertices (with respect to some measure of node weight) constituting the matching. In this pursuit, we introduce a metric called the assortativity weight of an edge, defined as the product of the number of uncovered adjacent edges and the absolute value of the difference in the weights of the end vertices. The MAM algorithm prefers to include edges that have the smallest assortativity weight in each iteration (one edge per iteration) until all edges are covered. The MAM algorithm can also be adapted to determine a maximal dissortative matching (MDM) to maximize the dissimilarity between the end vertices that are part of a matching as well as to determine a maximal node matching (MNM) that simply maximizes the number of vertices that are part of the matching. We run the MAM, MNM and MDM algorithms on real-world network graphs as well as on the theoretical model-based random network graphs and scale-free network graphs and analyze the tradeoffs between the % of node matches and assortativity index (targeted optimal values: 1 for MAM and −1 for MDM).

1. Introduction

A matching M for a graph $G = (V, E)$ is a subset of the edges E such that no two edges in M have a common vertex. A maximal matching is a set of independent edges such that the inclusion of any additional edge to the set violates the property of matching (no common vertex between any two edges of the set). A matching for a graph is said to be maximum if every vertex in the graph could be matched with another vertex of the graph through a set of edges such that no two edges in the set have a common vertex. There may exist maximal matching of various sizes for the vertices of a graph; but, every maximal matching need not be a maximum matching; on the other hand, a maximum matching of the vertices in a graph is the largest possible maximal matching for the vertices of the graph. Accordingly, we refer to the maximum matching problem as a problem of finding the largest set of independent edges whose end vertices form the non-overlapping node pairs such that the maximum number of node pairs is $\frac{V}{2}$ if the number of vertices V is even and is $\frac{V}{2} - 1$ if the number of vertices V is odd.

A well-known algorithm for finding the maximum set of independent edges for maximum node matching in arbitrary

E-mail address: natarajan.meghanathan@jsums.edu
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jksuci.2015.10.004

network graphs is the Blossom algorithm (Edmonds, 1965) of time-complexity $O(V^4)$ on a graph of V vertices. Several improvements (e.g., Micali and Vazirani) to the Blossom algorithm have been proposed in the literature. A weakness of all these algorithms is that in pursuit of maximum node matching, little consideration is given to the similarity between the vertices that are matched. As observed in the simulations of this paper, a maximum or maximal node matching of the vertices in a complex network graph need not match vertices of comparable node weight (for e.g., node degree). The motivation for the research presented in the paper stems from this observation. We want to determine a maximal matching (need not be the maximum matching, but close enough to the maximum matching) of the vertices that are very similar to each other (or very dissimilar from each other). This amounts to maximizing (or minimizing) a metric called the assortativity index of the edges that constitute the matching. Until now in the literature for complex network graphs, assortativity has been considered only at the network level (Newman, 2002) and node level (Piraveenan et al., 2008, 2012), but not with respect to the matching of the vertices. Ours is the first paper in this direction.

The assortativity index of a set of edges (with respect to any particular measure of node weight – like the node degree) is a quantitative measure of the similarity between the end vertices of the edges that are part of the set (Newman, 2010). The assortativity index values can range from -1 to 1. If the assortativity index of a set of edges calculated with respect to a particular measure of node weight is close to 1, then it implies the end vertices of the edges that form the set are very similar to each other with respect to the particular measure of node weight (for example, a high degree vertex matched to another high degree vertex, a low degree vertex matched to another low degree vertex, etc). If the assortativity index is close to 0, then the pairing of the vertices in the edge set is arbitrary with respect to the node weight. On the other hand, if the assortativity index of the set of the edges with respect to a measure of node weight is close to -1, then it implies that most of the node pairs constituting the edge set are not similar to each other with respect to the node weight (for example, if node degree is used as the node weight, then an assortativity index of -1 of a set of edges implies that most of the node pairings in this set involve a high degree vertex matched to a low degree vertex and vice versa).

For social networks and other complex real-world networks where peer-to-peer interaction and collaboration are preferred, it might be useful to pair vertices that are very similar (or very dissimilar) to each other as part of a maximal matching of the vertices in the network. A maximal matching that is arbitrary with respect to the weight of the vertices being matched need not be preferred in social networks. For example, a researcher who already has some accomplishments to his/her credit may want to pair with another researcher who also has a similar research profile (say quantified in terms of the number of peer-reviewed publications in a research area) so that they can mutually collaborate and benefit from each other. On the other hand, a newly joining researcher to a social forum (like researchgate.net or linkedin.com) may want to pair with an accomplished researcher. If each node in a social network can be matched with only one another node at a time, then it is imperative to match the nodes that are either dissimilar to each other or similar to each other (depending on the application of interest); an arbitrary matching of the vertices in a social network may not be of any practical benefit. To the best of our knowledge, we have not come across a maximal matching algorithm that maximizes the assortativity index (for matching nodes that are similar to each other) or minimizes the assortativity index (for matching nodes that are very different from each other) in complex network graphs.

In this paper, we propose a maximal matching algorithm that can be used to maximize or minimize the assortativity index of the edges constituting the matching determined in complex network graphs where the nodes have weights (the smaller the difference in the node weights, the more similar are the nodes and vice versa). An edge that is part of a matching is said to cover itself as well as cover the edges adjacent to it in the original graph and these edges cannot be part of the matching. We define a metric called the assortativity weight of an edge as the product of the number of uncovered edges adjacent to the edge in the graph and the absolute value of the difference in the weights of the end vertices constituting the edge. The maximal matching algorithm for maximizing the assortativity index (hereafter, referred to as the maximal assortative matching algorithm, MAM) prefers to include edges that have lower assortativity weight as part of the matching. The algorithm runs in iterations. In each iteration, we determine a ranking of the uncovered edges in the graph based on the assortativity weight metric defined above and choose the edge with the smallest value for the assortativity weight metric and include it among the edges constituting the matching. We continue the iterations until all edges in the graph are covered. An edge with the smallest value for the assortativity weight is likely to have fewer adjacent edges as well as comprise end vertices with close-enough node weights. Our hypothesis is that by choosing such edges with smaller values for the assortativity weight, for graphs that are sufficiently dense, we can simultaneously maximize the assortativity index of the matching as well as maximize the number of edges chosen as part of the matching. The proposed algorithm would be very useful for matching vertices in social networks and other real-world networks for peer-to-peer interaction and collaboration.

Ours will be the first such algorithm to determine a maximal matching of the vertices based on the notion of assortativity weight of the edges and does not use the notion of augmenting paths (Cormen et al., 2009), as used by most of the existing matching algorithms. We evaluate the performance of the proposed maximal assortative matching (MAM) algorithm on six real-world network graphs whose degree distribution ranges from Poisson (random networks) Strang, 2005 to Power-law (scale-free networks) (Caldarelli, 2007) as well as run the algorithm on complex networks simulated from theoretical models such as the Erdos–Renyi model (for random networks) (Erdos and Renyi, 1959) and Barabasi–Albert model (for scale-free networks) (Barabasi and Albert, 1999). We observe the MAM algorithm to determine a maximal matching of the nodes (the end vertices of each node pair are similar to each other) and the overall assortativity index of the matching is significantly larger than a matching of the nodes determined with the objective of just maximizing the number of nodes matched.

The focus of the paper is on presenting the proposed maximal assortative matching algorithm for maximizing the assortativity index of the matching. Toward, the end of the paper,