Spiders can be Recognized by Counting Their Legs

Sarah J. Berkemer · Ricardo R. C. Chaves · Adrian Fritz · Marc Hellmuth · Maribel Hernandez-Rosales · Peter F. Stadler

Received: 8 November 2014 / Accepted: 30 April 2015 / Published online: 19 May 2015 © Springer Basel 2015

Abstract Spiders are arthropods that can be distinguished from their closest relatives, the insects, by counting their legs. Spiders have eight, insects just six. Spider graphs are a very restricted class of graphs that naturally appear in the context of cograph editing. The vertex set of a spider (or its complement) is naturally partitioned into a clique (the body), an independent set (the legs), and a rest (serving as the head). Here we show that spiders can be recognized directly from their degree sequences through the number of their legs (vertices with degree 1). Furthermore, we completely characterize the degree sequences of spiders.

Keywords Phylogenetics · Cograph · P4-sparse · Spider · Degree sequence

Mathematics Subject Classification Primary 05C07; Secondary 05C75

To all arachnophobic mathematicians.

This work was funded in part by the German Research Foundation (DFG) (Proj. No. MI439/14-1). MHR is supported by the program “Cátedras Jóvenes Investigadores, CONACYT, México”.

S. J. Berkemer · P. F. Stadler
Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig, Germany
S. J. Berkemer · P. F. Stadler
Bioinformatics Group, Department of Computer Science, Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstrasse 16-18, 04107 Leipzig, Germany
e-mail: bsarah@bioinf.uni-leipzig.de

P. F. Stadler
e-mail: studla@bioinf.uni-leipzig.de

R. R. C. Chaves · M. Hernandez-Rosales
Departamento de Ciência da Computação (CIC), Instituto de Ciências Exatas, Universidade de Brasília, Campus Universitário-Asa Norte, Brasília, DF, CEP 70910-900, Brazil
e-mail: rrcchaves@gmail.com

A. Fritz · M. Hellmuth
Center for Bioinformatics, Saarland University, Building E 2.1, P.O. Box 15 11 50, 66041 Saarbrücken, Germany
e-mail: s9adfrit@stud.uni-saarland.de
1 Introduction

Determining the evolutionary history of species based on sequence information is one of the main challenges in phylogenomics. Recent advances in phylogenetics have shown that cographs play a crucial role in order to clean up orthology (or paralogy) data, i.e., estimates of pairs of genes that arose from a speciation (or duplication) event in the gene tree [4–6]. However, the (decision version of the) problem to edit a given graph into a cograph is NP-complete [10,11]. Nevertheless, the cograph editing problem can be solved in polynomial time, whenever the graph under investigation is so-called \( P_4 \)-sparse [10,11]. The structure of \( P_4 \)-sparse graphs can be characterized in terms of so-called spiders [8,9]. Hence, the cograph editing problem can be rephrased as a “spider editing problem”. To address the spider editing problem in future work, we give in this contribution a full characterization of spiders in terms of their degrees and respective degree sequences.

2 Preliminaries

We consider only simple graphs \( G = (V, E) \). The degree \( \text{deg}_G(v) \) of a vertex \( v \in V \) is defined as the number of edges that contain \( v \). If there is no risk of confusion we write simply \( \text{deg}(v) \) instead of \( \text{deg}_G(v) \). The degree sequence \( (n_0, n_1, n_2, \ldots, n_w) \) is a list of non-negative integers. A degree sequence is \emph{graphical}, if it can be realized by a graph, i.e., if there is a graph \( G \) s.t. there are exactly \( n_k \) vertices with degree \( k \) in \( G \). Of course, for simple graphs we have \( n_k = 0 \) for \( k \geq |V| \). It is well-known that one can verify in polynomial time, whether a degree sequence is graphical, either by means of the Havel [3] and Hakimi [2] algorithm or by using the Erdős and Gallai [1] theorem.

A graph \( H \) is a \emph{subgraph} of a graph \( G \), in symbols \( H \subseteq G \), if \( V(H) \subseteq V(G) \) and \( E(H) \subseteq E(G) \). A subgraph \( H \subseteq G \) is called \emph{induced}, when \( xy \in E(H) \) if and only if \( xy \in E(G) \) for all \( x, y \in V(H) \). The \emph{complement} \( \overline{G} \) of a graph \( G = (V, E) \) has vertex set \( V \) and edge set \( \overline{E} = \{xy \mid xy \notin E \text{ for all distinct } x, y \in V \} \).

A graph \( G \) is called \emph{\( P_4 \)-sparse} if every set of five vertices induces at most one \( P_4 \) (chordless path on four vertices) [7]. The efficient recognition of \( P_4 \)-sparse graphs is intimately connected to spider graphs, as the next lemma shows.

**Lemma 2.1** [8,9] A graph \( G \) is \( P_4 \)-sparse if and only if exactly one of the following three alternatives is true for every induced subgraph \( H \) of \( G \): (i) \( H \) is not connected, (ii) \( \overline{H} \) is not connected, or (iii) \( H \) is a spider.

Spiders come in two sub-types, called \emph{thin} and \emph{thick}.
A graph \( G \) has vertex partition \( K = \{0, 1, 2\}, S = \{x, y, z\} \) and \( R = \{a, b, c, d\} \) has degree sequence \( (n_0, n_1, \ldots, n_{|V|-1}) = (0, 3, 0, 0, 2, 0, 3, 0, 0) \). In this example there is another spider hidden in the head \( R \), namely \( K' = \{b, c\} \), \( S' = \{a, d\} \), and \( R' = \emptyset \). The thick spider \( \overline{G} \) has vertex partition \( K = \{x, y, z\} \), \( S = \{0, 1, 2\} \) and \( R = \{a, b, c, d\} \) and degree sequence \( (\overline{n}_0, \overline{n}_1, \ldots, \overline{n}_{|V|-1}) = (n_{|V|-1}, \ldots, n_1, n_0) = (0, 0, 3, 0, 2, 0, 0, 3, 0) \). a Thin spider. b Thick spider

**Definition 2.2** [9, 12] A graph \( G \) is a thin spider if its vertex set can be partitioned into three sets \( K, S, \) and \( R \) so that (i) \( K \) is a clique; (ii) \( S \) is a stable set; (iii) \( |K| = |S| \geq 2 \); (iv) every vertex in \( R \) is adjacent to all vertices of \( K \) and none of the vertices of \( S \); and (v) each vertex in \( K \) is connected to exactly one vertex in \( S \) by an edge and vice versa.

A graph \( G \) is a thick spider if its complement \( \overline{G} \) is thin spider.

The sets \( K, S, \) and \( R \) are usually referred to as the body, the set of legs, and head, resp., of a thin spider, see Fig. 1. The path \( P_4 \) is the only graph that is both a thin and thick spider.

### 3 Results

Spider graphs have characteristic degree sequences. Suppose \( G \) is a thin spider. It follows directly from the definition that \( \deg(x) = 1 \) for all \( x \in S \); the leg \( x \) is adjacent to a single body vertex and to none of the head vertices [Conditions (iv) and (v)]. Each body vertex \( x \in K \) has \( \deg(x) = |K| + |R| \), since it is adjacent to 1 leg, all \( |K| - 1 \) other body vertices, and all \( |R| \) head vertices. For a head vertex \( x \in R \) we have \(|K| \leq \deg(x) \leq |K| + |R| - 1 \) since it is adjacent at least to all body vertices and at most to all body and head vertices except itself. As \(|K| + |R| = |V| - |S| \) we can rephrase these observations in the following form, which have in parts already been discovered in [9, Obs. 2.8 & 2.9].

**Lemma 3.1** If \( G \) is a thin spider, then it has \(|S| \geq 2 \) legs with degree 1, \(|S| \) body vertices with degree \(|V| - |S| \), and \(|V| - 2|S| \) head vertices \( x \in R \) with \( 2 \leq |S| \leq \deg(x) \leq |V| - |S| - 1 \).

Our main result is that this condition is also sufficient to identify thin spiders:

**Theorem 3.2** A graph \( G = (V, E) \) is a thin spider if and only if there is an integer \( s \geq 2 \) so that \( G \) has exactly \( s \) vertices of degree 1 and exactly \( s \) vertices of degree \(|V| - s \).

**Proof** The “only if” part of this statement is a consequence of Lemma 3.1.

Hence, consider a graph \( G \) with \( s \geq 2 \) nodes with degree 1 and \( s \) nodes of degree \(|V| - s \). Define \( K := \{x \in V | \deg(x) = |V| - s\} \) as the set of body and \( S := \{x \in V | \deg(x) = 1\} \) as the legs.
Claim 1. For each body vertex $x \in K$ there is a unique leg $y \in S$ so that $xy \in E$.
If $x$ is not adjacent to a leg $y \in S$ then it has at most $|V| - |S| - 1$ neighbors, a contradiction. Thus, every body vertex $x \in K$ has at least one adjacent leg $y \in S$. Since $|K| = |S|$ by assumption, every body node is adjacent to exactly one leg.
This establishes property (v).

Claim 2. Every body vertex $x \in K$ is adjacent to all non-leg vertices $y \in V \setminus S$, except itself.
Since $\deg(x) = |V| - |S|$ and one of the neighbors of $x$ is a leg, it is adjacent to $|V| - |S| - 1$ non-leg vertices.
There are in total $|V| - |S|$ non-leg vertices, i.e., $|V| - |S| - 1$ non-leg vertices other than $x$ itself. Since $x$ cannot be adjacent to itself, it is adjacent to all other non-leg vertices.
It follows that $K$ forms a clique in $G$, i.e., property (i) holds. Furthermore every head node $z \in R := V\setminus(K \cup S)$ is connected to every body node and to none of the legs, i.e., property (iv) holds. Conditions (ii) and (iii) are satisfied by construction, thus $G$ is a thin spider.

Since the complement of a thick spider is a thin spider and vice versa, the characterization of thin spiders immediately implies a characterization of thick spiders because $\deg_G(x) + \deg_{\overline{G}}(x) = |V| - 1$. It follows that a thick spider has $|S|$ body vertices of degree $|V| - 2$ and $|S|$ leg vertices of degree $|S| - 1$.

Corollary 3.3 $G$ is a thick spider if and only if there is an integer $s \geq 2$ so that $G$ has exactly $s$ vertices with degree $s - 1$ and $s$ vertices with degree $|V| - 2$.

Theorem 3.4 Let $\Pi = (n_0, n_1, \ldots, n_w)$ be an arbitrary degree sequence. Then, $\Pi$ is realizable by a thin spider if and only if there are integers $s \geq 2$, $v \geq 4$, $s < v$ so that

(i) $n_1 = s$, $n_{v-1} = s$, and

(ii) $\Pi' = (m_0, m_1, \ldots, m_w)$ is a graphical degree sequence where

$$m_k = \begin{cases} 
  n_k = 0, & \text{if } 2 \leq k < s \text{ or } k \geq v - s \\
  n_k - s, & \text{otherwise}.
\end{cases}$$

A sequence $(n_0, n_1, \ldots, n_{|V|-1})$ is the degree sequence of a thick spider if and only if $(n_{|V|-1}, \ldots, n_{|V|-1-j}, \ldots, n_1, n_0)$ is the degree sequence of a thin spider.

Proof Let $\Pi$ be realizable by a thin spider. Theorem 3.2 implies Condition (i). As already observed, the $|V| - 2|S|$ head vertices of a thin spider have degrees in the range from $|S|$ to $|V| - |S| - 1$, that is reflected by the given degrees of $\Pi'$ in Condition (ii). Moreover, since $\Pi$ is the degree sequence of a thin spider $G$, $R$ induces a subgraph of $G$ and each vertex in $R$ has $s$ neighbors in $K$, we can conclude that $\Pi'$ is a graphical degree sequence. Note, from the algorithmic point of view, Condition (ii) can easily be checked by the Erdős and Gallai [1] theorem.

Now assume that there are integers $s \geq 2$, $v \geq 4$, $s < v$ so that Conditions (i) and (ii) are fulfilled for $\Pi$. We construct a spider from the graph $(V, E)$ with $|V| = v$ and $E = \emptyset$. Choose a subset $K \subseteq V$ with $|K| = s$, $S \subseteq V\setminus K$ with $|S| = s$, and set $R = V\setminus(K \cup S)$. Now, add edge $xy$ to $E$ for all distinct $x, y \in K$, choose a bijection $f : K \rightarrow S$ and add $xf(x)$ to $E$ for all $x \in K$ and finally add edge $ab$ to $E$ for all $a \in K$, $b \in R$. Clearly, the construction is feasible in terms of Conditions (i) and (ii), and fulfills the properties of a thin spider. Finally, Condition (ii) also ensures that $R$ induces itself a graph, that can be constructed with the Havel [3] and Hakimi [2] algorithm.

The result for thick spiders follows by observing that the degree sequence of $\overline{G}$ is related to the degree sequence of $G$ via $\overline{n}_k = n_{|V|-1-k}$ for $0 \leq k \leq |V| - 1$.

References

1. Erdős, P., Gallai, T.: Gráfok előírt fokszámú pontokkal. Mat. Lapok 11, 264–274 (1960)
2. Hakimi, S.L.: On realizability of a set of integers as degrees of the vertices of a linear graph. J. Soc. Ind. Appl. Math. 10, 496–506 (1962)
3. Havel, V.: A remark on the existence of finite graphs. Časopis Pro pěstování Matematiky (in Czech) 80, 477–480 (1955)
4. Hellmuth, M., Hernandez-Rosales, M., Huber, K.T., Moulton, V., Stadler, P.F., Wieseke, N.: Orthology relations, symbolic ultrametrics, and cographs. J. Math. Biol. 66(1–2), 399–420 (2013)
5. Hellmuth, M., Wieseke, N., Lechner, M., Lenzhof, H.-P., Middendorf, M., Stadler, P.F.: Phylogenomics with paralogs. PNAS 112(7), 2058–2063 (2015)
6. Hernandez-Rosales, M., Hellmuth, M., Wieseke, N., Huber, K.T., Moulton, P.F., Stadler, V.: From event-labeled gene trees to species trees. BMC Bioinform. 13(Suppl 19), S6 (2012)
7. Hoàng, C.T.: Perfect graphs. PhD thesis, School of Computer Science, McGill University, Montreal (1985)
8. Jamison, B., Olariu, S.: P4-reducible-graphs—a class of uniquely tree of uniquely tree-representable graphs. Stud. Appl. Math. 81, 79–87 (1989)
9. Jamison, B., Olariu, S.: Recognizing P4-sparse graphs in linear time. SIAM J. Comput. 21, 381–406 (1992)
10. Liu, Y., Wang, J., Guo, J., Chen, J.: Cograph editing: complexity and parameterized algorithms. In: Fu, B., Du, D.Z. (eds.) COCOON 2011. Lect. Notes Comp. Sci., vol. 6842, pp. 110–121. Springer, Berlin (2011)
11. Liu, Y., Wang, J., Guo, J., Chen, J.: Complexity and parameterized algorithms for cograph editing. Theor. Comput. Sci. 461(0), 45–54 (2012)
12. Nastos, J., Gao, Y.: Bounded search tree algorithms for parameterized cograph deletion: efficient branching rules by exploiting structures of special graph classes. Discret. Math. Algorithms Appl. 4, 1250008 (2012)