G Model
SON-691; No. of Pages 5
ARTICLE IN PRESS
Social Networks xxx (2011) xxx–xxx
Contents lists available at SciVerse ScienceDirect
Social Networks
journal homepage: www.elsevier.com/locate/socnet

A second look at Krackhardt's graph theoretical dimensions of informal organizations

Martin G. Everett[a,*], David Krackhardt[b]

[a] Mitchell Centre for Social Network Analysis, School of Social Sciences, University of Manchester, Arthur Lewis Building, Bridgeford Street, Manchester M13 9PL, UK
[b] Carnegie Mellon University, United States

ARTICLE INFO

Keywords:
Arborescence
Out-tree
Graph-theoretic dimensions
Connectedness
Graph hierarchic
Graph efficient
Least upper boundedness

ABSTRACT

Krackhardt (1994) proposed four dimensions to describe and measure the amount of hierarchy in networks of informal organizations. We examine these conditions, suggest some relaxations and prove that they are both necessary and sufficient to guarantee an arborescence (or out-tree). In addition we suggest situations some of which are outside of informal organizations in which fewer of the conditions can be used to capture the hierarchical tree structure.

1. Introduction and mathematical preliminaries

Krackhardt (1994) proposed four dimensions to capture the extent to which a network formed a hierarchy. His stated aim was to allow researchers to have a set of ways to characterize and measure hierarchy in informal organizations (networks) in order to make comparisons across complex network structures. Such measures would enable a researcher, for example, to test propositions that relate the shape of a network to the organization's ability to operate successfully in uncertain environments (Lawrence and Lorsch, 1967; Thompson, 1967; Lincoln and Kalleberg, 1985), to deal with major crises (Krackhardt and Stern, 1988), or even to be profitable (Sarkar et al., 2010).

His argument started with the fact that traditional *formal* organizational structures (organizational charts) can be represented in graph-theoretic terms as an out-tree also called an arborescence (Berge, 1962; Tutte, 1984), a term which we prefer and will use for the rest of this paper. Further, he claimed that an arborescence can be defined as a directed graph that has four necessary and sufficient properties–specifically, that the directed graph:

1. Is connected
2. Is graph hierarchic
3. Is graph efficient
4. Meets the least upper boundedness condition.

Before we formally define these terms we will need to establish some basic notation. Let $D=(V, A)$ be a digraph with a set of vertices V and arcs A. An arc that connects a vertex u to a vertex v will be called a uv arc. The book by Bang-Jensen and Gutin (2007) is a good source of information on digraphs. A graph $G=(V, E)$ consists of a set of vertices V and edges E, the edges are unordered pairs as opposed to the ordered pairs in a digraph. The underlying graph of a digraph is the simple (that is no multiple edges) undirected graph which results when the arcs are replaced by non-directed edges. A path from vertex u to vertex z in a digraph (graph) is a succession of arcs (edges) of the form uv, vw, wx, . . ., yz where no vertices are repeated, we shall sometimes use the term directed path to emphasize that it is a path in a digraph. If in a digraph or graph there is a path from vertex u to vertex z then we say that z is reachable from u or equivalently that u can reach z. The set $R(u)$ is the set of all vertices reachable from u. A cycle is a succession of arcs of the form uv, vw, wx, . . ., yz, zu in which u, v, w, . . ., y, z are all different. A semipath and a semicycle are defined similarly except we ignore the directions of the arcs, so that any arc xy in the sequence can be replaced by yx. A digraph is weakly connected if there is a semipath connecting every pair of vertices, a graph is connected if there is a path connecting every pair of vertices; it follows that a digraph is weakly connected if the underlying graph is connected. A component of a graph is a maximal connected subgraph, a weak component of a digraph is a maximal weakly connected subgraph. We can now give formal definitions of the Krackhardt dimensions.

1 *The digraph is connected*. Connectedness here is weakly connected so that the digraph has only one weak component. That is every pair of vertices is joined by a path in the underlying graph.
2 *The digraph is graph hierarchic*. Graph hierarchy means that for every pair of distinct vertices x and y, if x can reach y then y cannot reach x.

* Corresponding author. Tel.: +44 0161 275 2515.
E-mail address: martin.everett@manchester.ac.uk (M.G. Everett).

0378-8733/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.socnet.2011.10.006

3 *The digraph is graph efficient*. Efficiency here relates to the number of edges in the underlying graph. A graph is efficient if each component of the underlying graph has exactly $q-1$ edges where q is the number of nodes in that component.
4 *The least upper boundedness condition*. This condition states that every pair of vertices in the same weak component in the digraph has a least upper bound. An upper bound for a pair of vertices x and y is a vertex which can reach both. A least upper bound is an upper bound that is included on at least one directed path from every other upper bound to each of x and y. Note the condition is for pairs of vertices in the same weak component and does not apply to all pairs.

Krackhardt then goes on to propose a set of measures based on these conditions. He measures the extent to which the observed digraph of informal relations in an organization conforms to each of these four properties by counting the number of violations of these it has and then producing a normalized score. On each scale, a value of 1 indicates there are no violations to that particular property in the digraph; a value of 0 indicates that the digraph contains the maximum possible number of violations to the property for a fixed number of vertices.

Whilst Krackhardt states that these four conditions are necessary and sufficient he does not provide a formal proof but instead uses these as a definition of an arborescence. We note that they cannot be both necessary and sufficient conditions and the definition of an arborescence at the same time. Krackhardt justifies this implicitly by using graph diagrams of arborescences. We shall first formally prove this result. Here we propose to use the following common definition of an arborescence: an arborescence is a digraph with $n-1$ arcs and with the property that all vertices are reachable from a single vertex, called the root. As noted by Krackhardt (see also Bang-Jensen and Gutin, 2007) there are a number of alternative and equivalent definitions.

Theorem 1. *A digraph is an arborescence if and only if it is weakly connected, graph hierarchic, graph efficient and every pair of vertices in the same weak component has a least upper bound.*

Proof. We first show an arborescence has these properties. By definition it is weakly connected (since every vertex is reachable from the root) and graph efficient. Since the digraph has $n-1$ arcs and it is weakly connected it cannot contain semi-cycles and hence must be graph hierarchic. Let x and y be any pair of vertices, by definition the root is an upper bound. The path from the root, r, to x and the path from r to y are both unique (a property of arborescences) let these be $rp_1p_2\ldots x$ and $rq_1q_2\ldots y$. We note that it is possible for p_i to equal q_j only when $i=j$ otherwise we would induce a semi-cycle and this would contradict the fact that the arborescence has $n-1$ arcs. Let s be the largest value such that $p_s=q_s$ (this must exist since both paths start at r) then p_s is a least upper bound.

Conversely suppose a digraph $D(V,A)$ has these properties. Since D is graph hierarchic it has no reciprocated arcs and as the underlying graph has $n-1$ edges it follows that D must have $n-1$ arcs. We need to show D has a root. Select any vertex x, if $R(x)=V$ then x is a root and the result follows. If x is not a root then $V-R(x)$ is non-empty and we can select a vertex $y \varepsilon V-R(x)$. The pair of vertices x and y have a least upper bound z and so z can reach both x and y and it follows that $R(x)$ is a proper subset of $R(z)$. Now either $R(z)=V$ or we can continually repeat the same construction obtaining strictly larger reachable sets until we eventually find a vertex s such that $R(s)=V$ and the result follows. □

2. Relaxing the conditions

In the paper Krackhardt states that the four conditions are independent. Unfortunately this is not precisely true. Krackhardt gives examples where each of the measures is zero and the other measures are not zero and these are correct. But if they were completely independent then any combination of values would be possible and this is not the case. For example if connectivity is zero this can only happen if we have the null graph (that is all actors are isolates) and hence the other measures are trivially all one. This dependency means that the conditions can be weakened and the other conditions used in conjunction with them to still produce an arborescence.

For example we can weaken the hierarchy condition to be simply no reciprocity. It is easy to see that a graph with no reciprocity and the other properties will have hierarchy. In fact we prove a stronger result as we do not need upper boundedness.

Theorem 2. *A digraph that is graph efficient, weakly connected and has no reciprocated arcs is graph hierarchic.*

Proof. Suppose there is a path P from x to y ($x \neq y$) then we need to show there is no path from y to x. Suppose there is a path P' from y to x then since none of the arcs in P are reciprocated then P union P' must induce a cycle in the underlying graph. Since the graph is weakly connected and graph efficient then the underlying graph must be a tree, contradicting the existence of a cycle. □

In the same way we can relax the *least* upper bound condition to require that each pair of vertices simply have an upper bound.

Theorem 3. *Let D be a digraph that is graph efficient and has the property that every pair of vertices in the same weak component has an upper bound then every pair of vertices in the same weak component has a least upper bound.*

Proof. Suppose x and y are in the same weak component but do not have a least upper bound. Let S be the set of upper bounds for x and y. We know by definition S is non-empty. If S only has a single element then this element will, by definition, be a least upper bound and so S must have more than one element. If S only contained x and y then again by definition these must be least upper bounds and so S must contain a vertex different from x and y. Let p be a member of S that is different from x and y. By definition p is not a least upper bound and hence there exists a vertex q in S such that p is not on any q–x path or p is not on any q–y path. Without loss of generality suppose p is not on any q–x path. If q is different from x then let z be an upper bound of p and q. It follows that $zpxq$ must form a semi-cycle and hence induce a cycle in the underlying graph contradicting the fact that it is graph efficient. If q is the same as x then there is a path from every member of S to x as they are all upper bounds, in which case x is a least upper bound, a contradiction. □

Combining these results we conclude that we could use the four weaker conditions

1. The digraph is connected (i.e. weakly connected).
2. The digraph has no reciprocated arcs.
3. The digraph is graph efficient.
4. Every pair of vertices in the same weak component has an upper bound.

However whilst these may satisfy a mathematical nicety it must be remembered that each measure proposed by Krackhardt had a justification in terms of informal organizations and the advantages of the arborescence structure. The individual measures may well capture something that is important structurally but is missed in the relaxations presented here. It should be noted that in the original paper there was no direct discussion about the exact nature of the relations involved. There is discussion about communication and a clear concept of some kind of authority relation, particularly in the discussion about upper boundedness, but the absence of a formal definition of the type of relation involved is an issue when trying to interpret the measures. We now look more closely at each of these four conditions.

2.1. Connectedness

Krackhardt justifies this by stating that connectedness is about the division (or not) of the organization. He argues (p. 97) that for complex tasks it would be necessary to have "a set of established communication and advice relations that incorporates all the actors, at least indirectly, would be essential." At first sight it is difficult to imagine that ignoring the direction of the relations would capture this. However, it is true that the directed relations usually considered are line management or advice relations at least this is what the paper implies. Hence directionality implies something about the differentiation in status of the actors in the relation; it implies nothing about which way information travels. So, for example, the fact that *X* goes to *Y* for advice does not at all imply that information only travels from *X* to *Y*. Indeed, it could imply just the opposite. When *X* is imparting information to *Y* about the nature of a problem that *X* (or perhaps the department or work group) is experiencing then *Y* is imparting information to *X* about solutions, advice, other resources or where *X* should go for help, etc. The advice-relation has a direction; *X* is deferring to *Y* because *X* is approaching *Y* for advice. Beyond that, it establishes only that *X* and *Y* are *interacting*, and whilst the direction of the arrow carries with it important information about the nature of the relation it does not, in this case, reflect the direction of communication which we can assume is two way. It is difficult to understand why directionality, which is fundamental to the arborescence, is ignored in two of the four measures unless one has the interpretation above in which the direction of the tie does not affect the underlying flow of communication. We could have modelled this more formally and have communication as a separate relation from the advice or line management relation. It is quite possible, and many would argue usual, that communication does not follow the advice or line management relations. However, since we expect these relations to be one source of two way communication they define a minimum of what would be required.

There may be situations in which the relation does reflect the direction of communication. One such example would be in the military where the relation involves giving orders, in this case we may want a definition that reflects this (Krackhardt's original paper clearly relates his work to informal organizations but we will suggest here that it can also be applied to more formal structures). In addition there may be other very different types of networks which we want to assess for arborescenceness and direction is fundamental. As an example we may be interested in animal networks and dominance type relations (Iverson and Sade, 1990). Clearly in both these cases the assumptions of the previous section would not apply and we would need a different measure of connectedness. One possible alternative, which takes account of directionality, would be the smallest number of actors required such that all actors in the network are reachable from this set of actors. Clearly for an arborescence this would be one (the root) and for a graph consisting of isolates this would be n. We can use this to provide a measure in the same vein as in the original paper. If *V* is the size of such a smallest set then we can define the degree of reachable connectedness as:

$$\text{reachable connectedness} = 1 - \frac{V-1}{n-1} \qquad (1)$$

However, it should be noted that if we accept that communication can occur regardless of the direction of the relation then the original definition and measure should be used.

Returning to the original connectedness definition having dealt with the fact that we can in most cases use the underlying graph then we do need to look at the fact that there should be just one weak component. That is there should be indirect ties between all of the actors in the network. We believe that it is essential for an organization to be coordinated. Information, sentiment, support, beliefs, etc., all propagate through the network. It is the successful diffusion of these beliefs, attitudes, etc., that define the organization's culture, orients its participants towards a common goal, coordinates it, allows it to deal with difficult and unanticipated problems. If the network is disconnected, then this propagation of information, beliefs, goals, etc., breaks down and it becomes more difficult to move the organization in a coordinated way. In the extreme, if everyone is disconnected from everyone else (connectedness = 0), then trying to organize, orient, change this organization would be an impossible task. The important issue (made in the original paper) is that connectedness is related to the manager's ability to mobilize *change* in the organization.

2.2. Hierarchy

As we have seen in the earlier section hierarchy can be relaxed and deduced from the simpler conditions such as reciprocity. We could therefore use reciprocity as an alternative and clearly this would be a radical departure from the original measure. Reciprocity measures the consistency of dyadic orientation; an organization that is characterized by a prevalence of non-symmetric ties might indeed be well worth watching. The original hierarchy measure captures the overarching status orientation of the organization as a whole, as if there were a magnetic force that oriented everyone towards some (high status) north pole. The difference is exemplified in a simple triad. The original measure would insist on transitivity to demonstrate hierarchy, implying a consistent status measure that allows each actor to ascertain exactly who has more status than whom within the triad. Reciprocity (at least lack of it) would allow the cyclical triad, which may relay something about constraints on individual dyadic relations (no symmetry allowed), but does not imply an overall status ordering of the actors. On balance hierarchy seems to be a fundamental property and for this reason the substitution of this condition by reciprocity seems a little perverse in most circumstances. As such we do not propose any alternative but usually include this measure in the set of dimensions.

2.3. Efficiency

In his original paper Krackhardt notes that if the underlying graph has more than $n-1$ edges then the digraph will have multiple paths or cycles. "They disrupt the bare bones structure of the outtree" (p. 98). It is however difficult to understand why the same logic is not applied to the 2-cycles that are allowed as a consequence of the reciprocated ties. The fact that a digraph with $n-1$ arcs is perfectly efficient and yet it is possible to double the number of arcs to $2n-2$ and this has the same perfect efficiency seems at first sight quite simply inconsistent. However, if we apply similar logic as to the discussion on connectedness we see this is not inconsistent. Efficiency is not about the number of arcs but the number of pairs of actors and as such a reciprocated relation requires the same effort and time commitment as does a non-reciprocated. Clearly this is an assumption and again there may be situations in which this is not true. In the days before electronic communication if the two actors were remotely located (so that telephone communication was also not possible) and communication was by mail

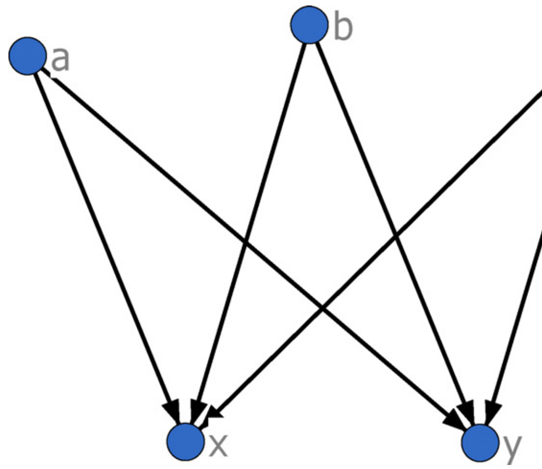

Fig. 1. Two actors with upper bounds but no least upper bounds.

then a reciprocated tie would have twice the efficiency cost as an un-reciprocated tie. This can also happen in modern day communication when communication is across large time zones or to remote locations. Another example is when the information exchanged is large or complex. For example amendments to large documents or when what is communicated cannot be sent electronically. The latter could include important originals, physical objects or highly sensitive materials for example. In these cases we count the number of arcs in the digraph rather than the underlying graph to get our efficiency measure. We shall call this digraph efficiency as opposed to graph efficiency.

2.4. Least upper bound

This is the most problematic of all the measures. On the face of it it is entirely natural. An upper bound is a common superior a least upper bound is the closest boss who has formal authority. This is how it is justified in Krackhardt's original paper. However the formal graph theoretic definition does not capture this concept but a different one. An upper bound of two actors in the digraph is any actor who can reach them both. Clearly there can be many upper bounds. Informally Krackhardt discusses the idea that a least upper bound (often referred to as a LUB) is someone to whom they can both appeal. The first issue is that a pair of actors can have upper bounds but no least upper bound as shown in Fig. 1. In Fig. 1a–c are all upper bounds for *x* and *y* but none of them are least upper bounds.

This is even more compounded when it can be seen that all the upper bounds could be least upper bounds and hence if a least upper bound exists it need not be unique. For example a component consisting of an isolated reciprocal dyad would mean that both the actors are least upper bounds and so both can appeal to themselves, and it is not clear how anything in this case is better resolved. This is not the only case when all upper bounds are least upper bounds the example in Fig. 2 has a–c as least upper bounds for *x* and *y*.

Alternatively a set of upper bounds could be linked by a very long cycle. Since they are on a cycle they would all be least upper bounds but the cycle length may mean that communication is not likely or even possible. In such cases it is difficult to see how having a least upper bound would help. Fig. 3 has this case as an example. The dotted lines represent long paths of length 20 (say). Both *a* and *b* are upper bounds for *x* and *y* but *a* is directly connected to both *x* and *y* and as such is the closest upper bound. But there is also a path from *a* to *b* to *x* and a path from *a* to *b* to *y* of length 20. Hence *b* is a least upper bound but is at a distance 20 from both *x* and *y*. Furthermore *a* is not even a least upper bound.

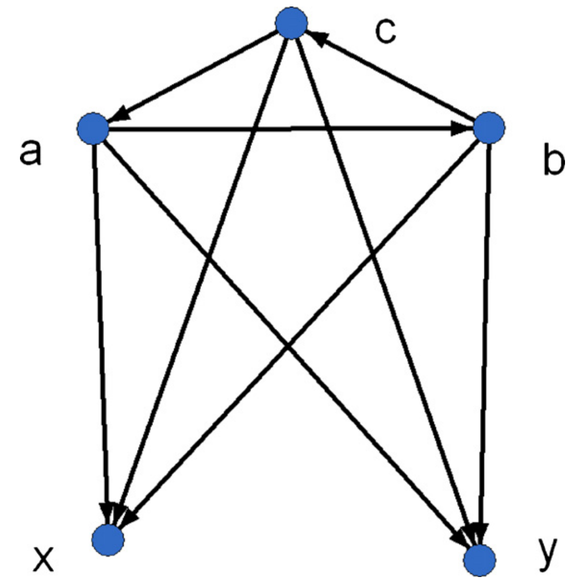

Fig. 2. A graph where all upper bounds are least upper bounds.

As already mentioned in the paper Krackhardt states that "In a formal organizational chart, the least upperbound of two employees is the closest boss who has formal authority over both of them" (p. 99). But as we see above the formal definition does not agree with this. What is interesting about the least upper bound measure is that it is an existence criteria and we do not need to know who the least upper bound is. But actually it is not clear why having a least upper bound is more advantageous than just having someone whom the parties can agree to appeal too. It would seem prudent to drop the least condition of a least upper bound so that we just require the existence of an upper bound.

3. Discussion

One of the issues that was important in Krackhardt's original paper was that the four dimensions he proposed had some measure of independence. We have already mentioned that they are not completely independent, but as they stand it is not possible to deduce one from the others. As such he was able to call them graph dimensions as each of them captured something which could not be wholly captured by the others. Whilst this is mathematically nice it may not always be the best approach.

However it really is not possible to separate out the pure mathematical structure from the modelling that is implicit in the

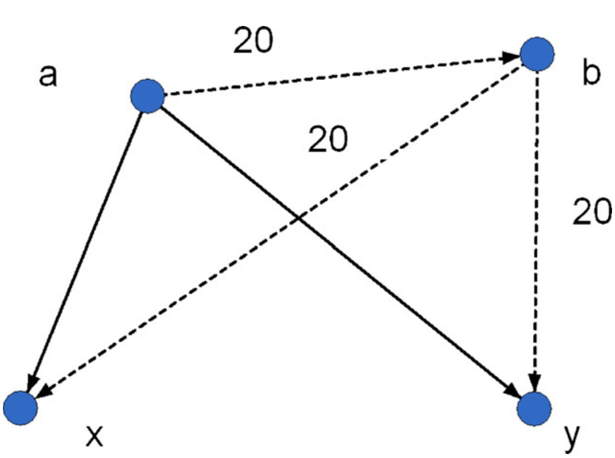

Fig. 3. A graph with a distant least upper bound.

measures. We must take some account of the nature of the relations involved to make sure the underlying assumptions which sit behind the measures are valid.

We therefore suggest different situations would require slightly different dimensions as follows.

To measure the extent to which a directed relation is an arborescence. In this instance we are not relating this to organizations and the relation can be any directed relation. We cannot assume that something flows against the direction nor can we assume that maintaining a reciprocal relation has the same efficiency as a non-reciprocated tie. In this case we would use the definition of the arborescence directly and would measure departure from these. That is

1. Reachable connectedness (i.e. there exists at least one root) measured as in Eq. (1).
2. Digraph efficiency (each weak component has $n_q - 1$ edges where n_q is the component size).

This is measured in the same way as graph efficiency that is

$$\text{digraph efficiency} = 1 - \frac{V}{\max V} \tag{2}$$

where V is the number of edges in excess of $n_q - 1$ summed over all components and $\max V$ is the maximum possible number of edges in excess of $n_q - 1$ summed over all components.

The fact that we recommend just two dimensions rather than the four in the original paper reflects the fact that we are capturing the raw definition and are not actually considering the nature of the relations involved.

These measures could be used on a formal organizational chart (who you report to) or on relations that we expect to reflect the formal structure, for example the relation who monitors your work. But we now have no real requirement to restrict these measures to organizations and we could use these to look at any networks in which we wish to see the extent to which they form an arborescence. A good and very different example as previously mentioned would be to examine the hierarchy of animal networks where dominance type relations are important.

Alternatively suppose we are in the situation where the nature of the relationship means that communication does flow against the direction of the arcs but this communication has a cost to maintain both ways. We would now require that our network is weakly connected (as in the original definitions given in Section 1), digraph efficient and satisfies the upper boundedness condition (not least upper boundedness). Note that weakly connected with digraph efficiency guarantees that we have the required $n - 1$ arcs and that the upper boundedness implies we have a root (as in the proof of Theorem 1). In summary if communication occurs against the flow but reciprocating a tie has a cost then we have three conditions and measures namely

1. Connectedness
2. Digraph efficiency
3. Upper boundedness.

Note we measure connectedness as in Krackhardt's original paper, digraph efficiency as in Eq. (2) and upper boundedness again as in the original paper but we just drop the least condition.

Situations in which this may occur are when there is a cost in making the communication in each direction as opposed to the original assumption that the cost is the same regardless of whether the communication is reciprocated or not. Examples of this are when the communication involves material that has to be physically transported or couriered (legal or sensitive documents) or when actions have to be separately recorded (for example conversations put into writing). Outside of organizations and again looking at animal networks then grooming amongst primates would be another example (Sade, 1972).

It is also possible that we have the situation where the relation is such that communication does not always flow both ways but maintaining two way communication has no significant extra cost. We would now require reachable connectedness and would relax digraph efficiency to graph efficiency. To make sure we get an arborescence all that would be required would be that in addition we had no reciprocity but as previously mentioned we prefer the stronger condition of graph hierarchy. Hence in this situation we suggest the three dimensions should be

1. Reachable connectedness
2. Graph efficiency
3. Graph hierarchy.

Again it is easy to see that these will give an arborescence.

This would be the case in highly authoritarian organizations such as the paramilitary or where there is a large status difference between those communicating for example senior consultants and junior doctors.

Finally we look again at the conditions implicit in Krackhardt's original paper. That is maintaining efficiency is about pairs of actors and it is not dependent on the direction of the arcs and that communication flows both ways in the relations. In which case we suggest a slight modification of the original dimensions and that we replace the least upper boundedness with just upper boundedness and so the dimensions are

1. Connectedness
2. Graph hierarchy
3. Graph efficiency
4. Upper boundedness.

The fact that these are necessary and sufficient for an arborescence follows directly from Theorems 1 and 3.

It should be noted that we could combine the measures to get an overall score of arborescenceness by say multiplying the measures. This would mean that we would have a value that would range from zero to one. Whilst this is a possibility we merely include it here as a suggestion for a possible direction for further work.

References

Bang-Jensen, J., Gutin, G., 2007. Digraphs: Theory, Algorithms and Applications. Springer-Verlag, New York.
Berge, C., 1962. The Theory of Graphs and its Applications. Mathuen, London.
Iverson, G.J., Sade, D.S., 1990. Statistical issues in the analysis of dominance hierarchies in animal societies. Journal of Quantitative Anthropology 2, 61–83.
Krackhardt, D., 1994. Graph theoretical dimensions of informal organizations. In: Carley, K., Prietula, M. (Eds.), Computational Organizational Theory. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 89–111.
Krackhardt, D., Stern, R., 1988. Informal networks and organizational crises: an experimental simulation. Social Psychology Quarterly 51, 123–140.
Lawrence, P.R., Lorsch, J.W., 1967. Organization and environment. Harvard University, Graduate School of Business Administration, Boston.
Lincoln, J.R., Kalleberg, A.L., 1985. Work organization and work-force commitment: a study of plants and employees in the U.S. and Japan. American Sociological Review 50, 738–760.
Sarkar, A., Fienberg, S.E., Krackhardt, D., 2010. Predicting profitability using advice branch bank networks. Statistical Methodology 7, 429–444.
Sade, D.S., 1972. Sociometrics of Macaca-Mulatta: linkages and cliques in grooming matrices. Folia Primatologica 18, 196–223.
Tutte, W.T., 1984. Graph Theory. Addison-Wesley, Reading, MA.
Thompson, J.D., 1967. Organizations in Action: Social Science Bases of Administrative Theory. McGraw-Hill, New York.