

USING NETWORK ANALYSIS TO CHARACTERIZE FOREST STRUCTURE

MICHAEL M. FULLER*
Department of Biology
University of New Mexico
Albuquerque, NM 87131
E-mail: mm.fuller@utoronto.ca

ANDREAS WAGNER
Department of Biochemistry
Winterthurerstrasse 190
CH-8057 Zurich Switzerland

BRIAN J. ENQUIST
Department of Ecology and
Evolutionary Biology
University of Arizona
Tucson, AZ 87519

ABSTRACT. Network analysis quantifies different structural properties of systems of interrelated parts using a single analytical framework. Many ecological phenomena have network-like properties, such as the trophic relationships of food webs, geographic structure of metapopulations, and species interactions in communities. Therefore, our ability to understand and manage such systems may benefit from the use of network-analysis techniques. But network analysis has not been applied extensively to ecological problems, and its suitability for ecological studies is uncertain. Here, we investigate the ability of network analysis to detect spatial patterns of species association in a tropical forest. We use three common graph-theoretic measures of network structure to quantify the effect of understory tree size on the spatial association of understory species with trees in the canopy: the node degree distribution (NDD), characteristic path length (CPL), and clustering coefficient (CC). We compute the NDD, CPL, and CC for each of seven size classes of understory trees. For significance testing, we compare the observed values to frequency distributions of each statistic computed from randomized data. We find that the ability of network analysis to distinguish observed patterns from those representing randomized data strongly depends on which aspects of structure are investigated. Analysis of NDD finds no significant difference between random and observed networks. However, analysis of

**Current Address:* Michael M. Fuller, Ph.D., Faculty of Forestry, University of Toronto, 33 Willcocks Street, Toronto, ON M5S 3B3

CPL and CC detected nonrandom patterns in three and one of the seven size classes, respectively. Network analysis is a very flexible approach that holds promise for ecological studies, but more research is needed to better understand its advantages and limitations.

KEY WORDS: Graph theory, network analysis, tropical trees, species association, community structure.

1 Introduction. Network analysis is used to study a wide range of natural and artificial systems, including social networks (Wasserman and Faust [1994], Scott [2000]), technological networks such as the World Wide Web (Albert et al. [1999]), and biological networks such as foodwebs (Williams and Martinez [2000]) and metabolic networks (Fell and Wagner [2000], Jeong et al. [2000]). Mathematical graph theory provides the tools needed for analyzing the structure of networks (which are therefore referred to as graphs). In the above examples, graph-theoretic approaches proved useful in detecting and understanding the causes of observed patterns. Given the network-like quality of many ecological phenomena, network analysis may be ideally suited for understanding ecological systems. For example, it may help to identify the processes by which species are able to coexist in communities (Dale [1985]). However, the use of graph-theoretic approaches by ecologists is limited to a few specific contexts. Most ecological studies involving graph-theoretic approaches focus on the structure of food webs (Briand and Cohen [1984], Pimm [1984], Cohen and Palka [1990], Williams and Martinez [2000], Williams et al. [2002], Garlaschelli et al. [2003]), with occasional forays into landscape ecology (Cantwell and Forman [1993], Urban and Keitt [2001]) and nearest-neighbor analysis in plants (Dale and Powell [1994]). Here, we extend the use of graph theory in ecology by applying it to the analysis of a single-trophic-level species community. We use three measures of graph topology to characterize the spatial organization of a tropical tree community.

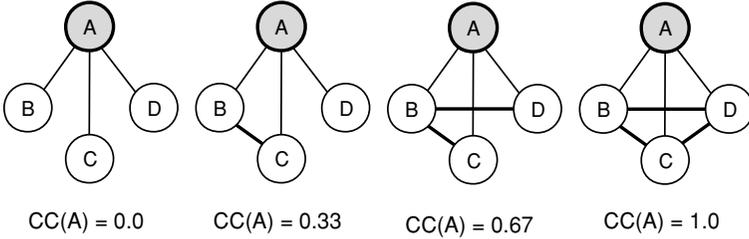
Although graph theorists have developed many different metrics for quantifying network structure (see review of popular approaches in Newman [2003]), most graph-based studies in ecology use one of two approaches: minimum spanning trees or polygon analysis (i.e., Dirichlet domains, Delaunay networks, Voronoi polygons). These approaches are useful, and they have been advocated by some ecologists for the

analysis of spatial patterns (Dale [1977, 1985, 1999], Urban and Keitt [2001], Fortin and Dale [2005]). By contrast, ecologists who study food webs tend to ignore graph-based measures of structure and instead invent indices to quantify specific food web properties, such as food chain length (Cohen et al. [1986]) and intervality (Cohen and Palka [1990]). Although there are exceptions to this approach (Dunne and Martinez [2002], Williams et al. [2002], Garlaschelli et al. [2003]), overall, ecologists have made little use of the diversity of graph-theoretic metrics that are commonly employed in other fields. Here, we present an analysis of species patterns in a tropical forest that employs three graph-theoretic measures of structure used in other fields. Specifically, we use the node degree distribution (NDD), clustering coefficient (CC), and characteristic path length (CPL; Wasserman and Faust [1994], Newman [2003]). These metrics are used to investigate a wide range of phenomena (Watts and Strogatz [1998], Barabasi et al. [2000], Jeong et al. [2000], Strogatz [2001], Dunne and Martinez [2002]). We use them to investigate how the spatial association of tree species changes with increasing tree size. We compare networks representing the observed pattern of species association to those representing trees whose spatial positions are randomized. Our purpose is to evaluate the usefulness of graph-theoretic metrics as measures of community structure.

1.1 Three measures of network organization. Here we describe the measures of graph topology we chose for our analysis. We assume the reader is familiar with the basic principles of graph theory as can be found in such texts as Harary [1969], Gross and Yellen [1999], and Newman [2003].

1.1.1 Node degree distribution. The degree of a node is the number of edges connected to it. The degree distribution characterizes the likelihood that a randomly chosen node will have a degree of a given value. In a directed graph, node degree represents two components: the outgoing edges (out degree) and incoming edges (in degree). Here, we used the simpler construction of undirected graphs to calculate node degree. In our species-association networks (SANs), node degree represents the number of heterospecifics that each species interacts with in local neighborhoods. In other words, node degree is a measure of

Clustering Coefficient



Characteristic Path Length

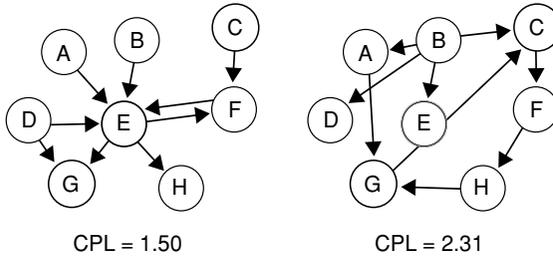


FIGURE 1. Measures of network structure. Top: $CC(A)$ = the clustering coefficient of node A. $CC(A)$ depends on the number of edges that connect the *neighbors* of node A (nodes B, C, and D). In the figure, CC increases from left to right with increasing connections between the neighbors of A. Bottom panel: CPL = characteristic path length. Both networks contain the same number of nodes and edges, but differ in CPL because of differences in how the nodes are connected.

the diversity of local neighborhoods and is proportional to the probability that a neighbor of a given tree is a different species. Negative interactions (i.e., competition) and strong differences in abiotic preferences among heterospecifics can reduce a species node degree relative to a randomly organized community, but positive interactions (e.g., facilitation) and shared habitat preferences may increase node degree.

1.1.2. *Clustering coefficient.* Consider a species v and its k_v neighbors—species directly connected to v —in a SAN. The CC of v measures the likelihood that these neighboring nodes or species are

also neighbors of each other (Figure 1). Specifically, for an undirected graph:

$$(1) \quad CC_V = \frac{E_V}{k_V(k_V - 1)/2},$$

where E_V is the number of edges among the k_V species (excluding v itself). Here, the value $k_V(k_V - 1)/2$ is the maximum possible number of edges among all k_V species connected to v . The CC ranges from zero (i.e., the neighbors of v are not neighbors of each other) to one (i.e., each neighbor of v is also a neighbor of every other neighbor of v). The value of CC_V measures the “cliquishness” of a graph, that is, the extent to which species form small groups that show preferential interactions within the group. This could occur, for example, if several species were restricted to a specific habitat type and were thus more likely to be associated. Species that have extreme habitat requirements or that exert strong negative effects on neighbors are more likely to have low values of CC_V . The CC of a graph is the average of the CCs of the individual nodes.

1.1.3 Characteristic path length. Path length refers to the shortest distance between two connected nodes, measured by the number of edges that separate them. The CPL of a graph is the shortest path length averaged over all node pairs that are connected by a path of edges (Figure 1). We used a directed graph to calculate CPL. In a directed graph, CPL is affected by the ratio of ingoing edges to outgoing edges connected to each node. If most edges in a graph are predominantly of one type, a larger number of edges on average span the distance between two nodes, increasing CPL (see Figure 1, e.g.). Thus CPL is a measure of the overall density and symmetry of network connections. The more heterospecific neighbors each species in a SAN has, on average, the shorter the CPL of the SAN. Thus factors that enhance the likelihood that species co-occur as neighbors, such as high relative abundance, shared habitat preferences, facilitation, and well-mixed spatial distributions, can reduce the path length between any two species (nodes) in a SAN. By contrast, the connectedness of a SAN will be lower, and the CPL will be greater, if competition and abiotic filtering exert a strong negative influence on species co-occurrence.

1.2 Network analysis of community structure in a tropical dry forest.

To test the usefulness of network analysis, we compare the observed spatial pattern of tree species in a community to a random spatial pattern. Specifically, we analyze the effect of tree size on the spatial association of tree species in a Costa Rican dry forest. The number of tree species that coexist in the understory of a forest is influenced by multiple biological and environmental processes (Janzen [1970], Connell [1978], Cornell and Lawton [1992]). For example, species interactions may prevent or facilitate the coexistence of some species, variation in habitat conditions may cause certain species to be spatially aggregated and others to be segregated, and dispersal limitation can induce spatial autocorrelation in the density of species and cause random spatial variation in community composition (Legendre [1993], Dale [1999]). The relative importance of these factors is likely to change with the size or age of individuals. For example, the effect of weak competitive interactions on species coexistence may not manifest until neighboring individuals have grown beside each other for some time. Another factor that influences coexistence is that the number of species of a given size decreases with increasing size. This phenomenon is known as the self-thinning law (Yoda et al. [1963], Weller [1987]) and is also referred to as an allometric scaling law (White [1981], Niklas [1994], West et al. [1999], Enquist and Niklas [2001]). Thus the spatial distribution of species, and the likelihood that two species coexist in the understory, is the result of some combination of random and deterministic processes. Here, we ask whether network analysis can detect the influence of tree size on the spatial association between understory species and their taller overstory neighbors. If deterministic processes that influence coexistence change in magnitude according to tree size, we should see differences between random and observed patterns of association.

To address the above question, we represent community structure as a network that we call a SAN. Specifically, this network is a graph whose nodes are species. An edge connects a species v_1 to species v_2 if the understory of at least one large tree of species v_1 harbors a smaller tree of species v_2 . We define a tree to be part of the understory of a target tree if its crown overlaps the crown of the target by at least 50%. Previous experiments show a gradual decreasing trend in the fraction of the species community represented in a SAN with increasing crown

overlap. For example, using four tree size classes we examined the fraction of the species community included in each of 10 separate annuli of crown overlap, ranging from 1–10% to 90–100% crown overlap (Fuller [2004]). The average community fraction represented in the 90–100% overlap annulus is 64.9%, and the richest annulus (1–10% overlap) averaged 75.9% of the community. As increasing the percent overlap yields a difference of only 11% in the fraction of the community included, we chose here to use a single overlap category of 50%. This approach greatly simplified our analysis.

The data represent a community of tree species located within the San Emilio Forest of northwest Costa Rica. The San Emilio Forest is a seasonally dry lowland tropical forest (10° 45' N, 85° 30' W) within sector Santa Rosa, Area de Conservacion, Guanacaste. For details on the physical and biological characteristics of the study plot see Enquist et al. [1999]. The data represent all trees with a basal stem diameter at breast height greater than 3 cm. We refer to diameter at breast height simply as the *stem diameter*. The stem diameter, species identity, and geographic coordinates of each tree were recorded by B.J. Enquist and C.A.F. Enquist between 1995 and 1996 within a continuous 14.2 ha rectangular plot within the forest (Enquist et al. [1999]). The plot is heterogeneous with respect to age, topography, and degree of deciduousness. In total, 13,639 individual trees (106 species) were used in our analysis. To quantify the effect of local neighborhood composition on community structure, we restrict the definition of a neighborhood to include only those trees that grow close enough to one another that their crowns overlap by at least 50%. The crown overlap of two trees is the area of the ground shaded by their overlapping foliage. We define *percent crown overlap* as the percent of the crown area of a tree covered by the larger crown area of a taller tree.

We estimated percent crown overlap based on the spatial coordinates of individual trees and the empirical allometric relationships among stem diameter, stem height, and crown diameter. Empirical and theoretical studies reveal that allometric scaling relationships exists between the height and radius of a tree's crown and its diameter (Baker [1950], Niklas [1994] and references therein). For trees in our study area, B.J. Enquist quantified these allometric relationships using tree measurements recorded in the field. Specifically, he quantified the relationships among stem diameter, crown area, and tree height for most

of the abundant species within the plot (38 species). Enquist regressed data at both the inter- and intraspecific level and found no significant differences in the allometric functions between taxa. The empirical allometric relationships are described by the following power functions:

$$(2) \quad \text{Crown Area} = 62.99D^{0.66},$$

$$(3) \quad \text{Crown Height} = 1.61D^{0.59},$$

where D is the stem diameter at breast height. Here we use interspecific allometries for both crown area and height. The above allometric relationships predict well the empirically measured values of crown height and area ($R^2 = 0.918$ and 0.922 , respectively).

To estimate the percent of crown overlap shared among individual trees we use the established relationship between stem and crown area and the known intertrunk distance of any two trees in the study plot. We assumed that tree crowns are roughly circular and symmetrical on average, and centered on the trunk. Although not strictly correct, and subject to individual variation, this assumption has precedence in other studies of forest structure (Porter [1989], O'Brien et al. [1995]) and allows us to estimate the area of ground shaded by a tree's crown as the area of a circle. By extension, one can estimate the percent crown overlap of two trees with the following equation, which describes the area of the lens formed by two intersecting circles (Weisstein, E.W. <http://mathworld.wolfram.com/Circle-CircleIntersection.html>):

$$(4) \quad A = r^2 \cos^{-1} \left(\frac{d^2 + r^2 - R^2}{2dr} \right) + R^2 \cos^{-1} \left(\frac{d^2 + R^2 - r^2}{2dR} \right) \\ - \frac{1}{2} \sqrt{(R + r - d)(d + r - R)(d + R - r)(d + r + R)}.$$

Here, R is the crown radius of the larger tree, r is the crown radius of the smaller tree, and d is the distance among the crown centers, given by the intertrunk distance. We calculated crown radius from the estimate of crown area described by equation (2).

1.3 Species-association networks. To determine whether network analysis can detect vertical variation in the neighborhood

structure of trees in the understory of larger trees, our construction of a SAN centers on *large target trees* in the community, which we define to be trees with a stem diameter of at least 40 cm. Multiple processes influence the coexistence of species in the understory, some of which may depend on the length of time individual trees occur next to one another. We therefore performed separate analyses using trees grouped into seven classes, according to their stem diameter: 5–10, 10–15, 15–20, 20–25, 25–30, 30–35, and 35–40 cm. Thus the similarity between the size of the understory trees and the target trees on which neighborhoods are centered increases with size class. The largest class includes some trees that are of equal stature as the target trees. However, the number of trees (and species) in a size class decreases rapidly with increasing size interval.

We constructed seven SANs, one for each of the listed seven size intervals of understory trees. In this construction, we first established a graph that consisted only of isolated nodes—nodes without edges—each of which corresponds to one of the 106 species on the study plot. For each of these species (nodes), we then sequentially carried out the following steps. First, we identified all individuals of a given species whose stem had a diameter of at least 40 cm. These are our *large target trees* for the species. Second, for each large tree, we determined all trees in the given stem diameter interval that are not of the same species and that have a percent crown overlap of 50% or greater with the large tree. We call all trees that meet these criteria the *neighborhood* of the large tree. For any species v that has a tree in this neighborhood of a large tree, we establish an edge in the graph from the current species to species v (Figure 2).

To construct the networks, we used the LEDA 4.3 package of C++ class libraries (Mehlhorn and Näher [2000]). To calculate the CPL, CC, and node degree of the networks, we used the algorithms described in Even [1979], Gibbons [1985], and Gould [1988].

1.4 Significance testing and randomization. Because our observed data represent a single tree community, we cannot directly calculate a measure of variance for each measure of structure that allows us to test the statistical significance of differences among classes. Instead, we use a Monte Carlo method (Manly [1997]) to compare the observed patterns against a null model. Published analyses of plant

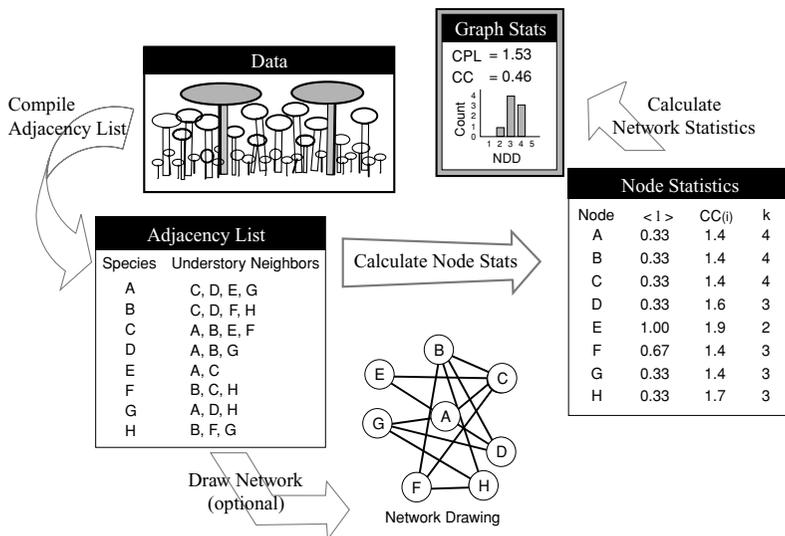


FIGURE 2. Constructing a species-association network (SAN). This figure uses a hypothetical example to show how we construct a SAN. First, data (top left box) on the position, size, and identity of individual trees are converted to adjacency lists (bottom left) for each species. A species “B” is placed on the adjacency list of species “A” if a shorter individual of B is found at least once in the understory of a large (≥ 40 cm stem diameter) individual of species A. An individual tree “B” is considered to be in the understory of another, taller tree “A” if the crown of B is overlapped by the crown of A by $\geq 50\%$. Once each species adjacency list is compiled, we use the lists to construct a graph (depicted here as a network drawing, bottom middle of figure) in which nodes represent species and edges connect each node to the species found on its adjacency list. We then calculate the node degree, k , clustering coefficient of each node, $CC(i)$, and path length, $\langle l \rangle$ of each species (rightmost box). See text for descriptions of these metrics. Finally, we calculate the node degree distribution (NDD histogram), clustering coefficient ($CC = \text{average of } CC(i)$), and characteristic path length ($CPL = \text{average of } \langle l \rangle$) for the graph as a whole.

community structure using other measures of spatial structure, such as Ripley’s K (Ripley [1976]), frequently compare the observed spatial pattern to a pattern representing complete spatial randomness (CSR). Under CSR, individuals are randomly and uniformly distributed in space (see examples in Fortin and Dale [2005]). Although widely used, CSR has been criticized as unrealistic (Legendre [1993], Dale [1999]).

Many processes can cause data values to be nonuniformly distributed in space. For example, physical limits on the dispersal distance of seeds and fruits can cause the density of individuals of a given species to be spatially autocorrelated. We recognize the importance of considering factors such as dispersal limitation when attempting to identify a specific process for a given spatial pattern. However, because our use of graph-theoretic metrics is a new approach to the study community structure, we consider it prudent to use CSR as a null model. Ecologists are familiar with CSR, and using it facilitates comparisons of our results to the many published studies that use CSR to analyze spatial patterns.

To generate distributions representing CSR, we simply rearranged the position of each tree according to a uniformly distributed random number. The new (randomized) position of each tree corresponds to the position previously occupied by a different tree. This technique preserves the observed pattern of where individual trees (but not species) occur on the landscape but removes any spatial structure, such as autocorrelation, in the distribution of each species. To prevent the occurrence of unlikely combinations of neighbors, such as replacing two neighboring small trees with two large trees, we randomized trees in the each size class separately from trees in the other size classes. We found no difference among the results for fully randomized data and those for restricted randomization. We therefore used the simpler full randomization approach.

To test the statistical significance of the observed patterns relative to CSR, we generate 1,000 iterations of each randomized SAN, from which we derive a distribution of values for each measure of network structure (e.g., the CC). For the NDD, we compare the observed distribution for each size class to the mean distribution (mean of 1,000 distributions) of the randomized SANs using the Kolmogorov–Smirnov goodness-of-fit test, which quantifies the difference between two distributions (Sokal and Rohlf [1995, pp. 708–715]. For the CC and CPL, we use the random distributions to determine the probability of occurrence for each of the observed values from each of the seven size classes. We consider observed values that are greater than the upper 95 percentile or less than the 5 percentile of the random distributions to be statistically significant. Finally, we compare the observed relationship between the CPL and tree size class to the corresponding relationship of the

random SANs using the Kolmogorov–Smirnov test. We use an identical comparison for the CC.

2. Results.

2.1 Node degree distribution. As described above, node degree is a measure of the species diversity of local neighborhoods. We compared the NDD among all seven empirically observed SANs and their randomized counterparts. Unsurprisingly, the two distributions were not absolutely identical (Figure 3). Specifically, the numbers of species with a given number of interactions differed between each size class and between the randomized SANs and their empirically observed counterparts. However, we observed no statistically significant differences between random and observed SANs in the NDD ($p > 0.05$ in Kolmogorov–Smirnov tests applied to all seven SANs). Thus we did not detect a substantial difference between random and observed networks in the number of understory species with which each target tree coexists.

2.2 Clustering coefficients. The CC of a node is proportional to the number of connections among its neighboring nodes. Here, neighboring nodes represent heterospecifics that occur in the understory of each species. Thus the CC of each species is proportional to the number of neighboring heterospecifics that also occur in each other's understory. In other words, the CC is a measure of preferential association among a subset of species. The CC of a SAN is computed as the average of the CCs of individual species. We find that the SAN CCs decrease with increasing size class (Figure 4a), but the nature of the relationship appears to differ between the random and observed SANs. The relationship between stem size diameter and CC is exponential. To illustrate this point we log transformed the values and fit second-order polynomial curves to the data using the least squares method (see Figure 4a). We then calculated the coefficient of determination for each fit as a measure of the variance in CC explained by stem size. We find that stem size explains less of the variation in the observed SAN ($R^2 = 0.8633$) relative to the random SAN ($R^2 = 0.9961$). All but one of the observed CC values fall within the 95% confidence intervals of the distributions representing the random

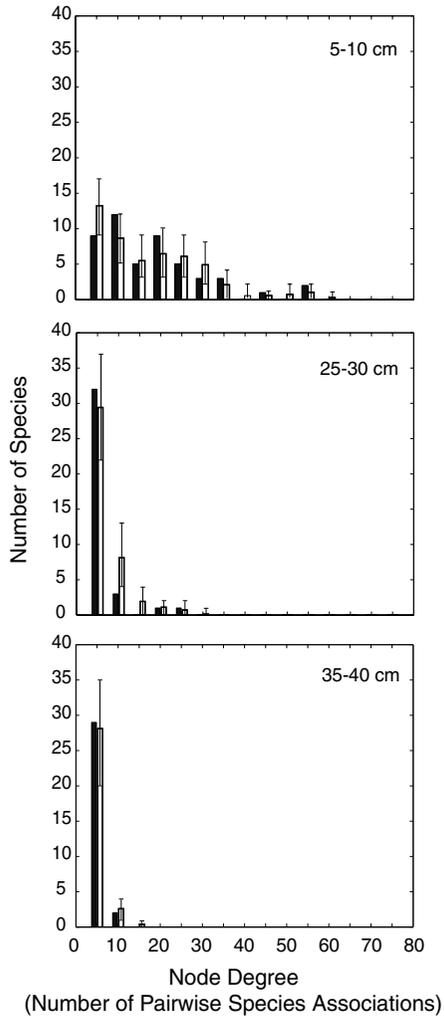


FIGURE 3. Effect of stem diameter on node degree distribution (NDD). This figure shows NDDs for three of seven SANs and illustrates the progression towards fewer pairwise species associations with increasing stem diameter of understory trees. Values in top right corner of boxes = stem diameter class. Shaded bars: empirically observed SANs; open bars: randomized SANs (mean of 1000 networks); error bars: 95% confidence limits of mean. The number of nodes of high degree decreases sharply with increasing stem diameter.

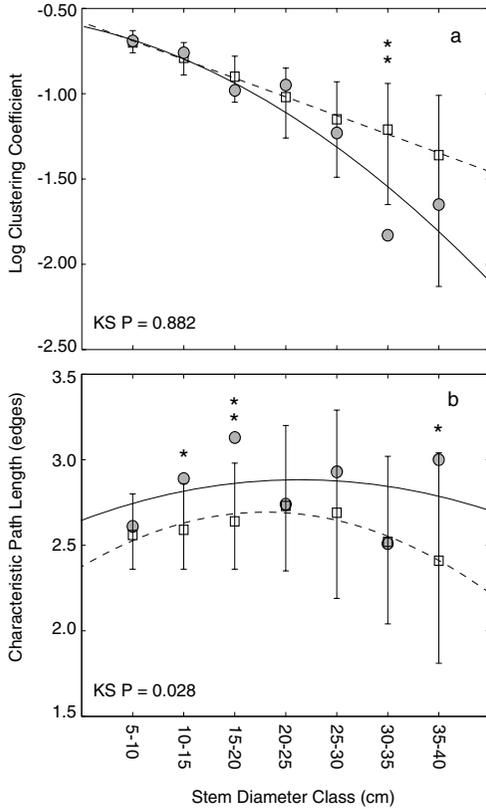


FIGURE 4. Effect of stem diameter on clustering coefficient (CC) and characteristic path length (CPL) of species networks. The horizontal axis indicates the stem diameter of understory trees used in the construction of the SANs we analyzed here. The vertical axes show the logarithmically transformed CCs (a) and CPLs (b). Shaded circles: values observed for empirically observed SANs; open squares: mean of 1,000 randomized SANs; error bars: 95% confidence limits of the mean; solid (dashed) lines: second-order polynomial fit for empirically observed (randomized) SANs. “KS p ” indicates the p -value of a Kolmogorov–Smirnov tests that assessed whether the distribution of CCs and CPL was identical for empirically observed and randomized SANs. One asterisk = significant at the 0.05 level; two asterisks = significant at the 0.01 level. The relationship between CPL and stem diameter is multimodal in the empirically observed networks but unimodal in randomized networks. The increasing error bars of CPLs and CCs with increasing stem diameter in the randomized networks is caused by the smaller number of large understory trees in the larger size classes.

SANs. The single exception is the observed SAN representing understory trees with 30–35-cm stem diameter that has a significantly lower CC ($p = 0.002$). The overall relationship of the observed CCs to tree size class (across all size classes) is not significantly different from the randomized SANs (Kolmogorov–Smirnov test, $p > 0.8$).

2.3 Characteristic path lengths. Exactly analogous to our analysis of CCs, we compared the CPL of the seven size classes. CPL is a measure of the density of network connections and reflects the equanimity of species spatial association. Higher path lengths indicate spatial segregation, but lower path lengths reveal a high level of spatial mixing. The mean values of the random distributions show a unimodal relationship with tree size class, with the mean random CPL being greatest in the middle tree size class (20–25 cm; Figure 4b). Thus, when tree positions are randomized, the density of network connections is a quadratic function of tree size class. By contrast, in the observed networks there is a bimodal relationship between tree size and CPL. In three of the seven stem diameter classes (10–15, 15–20, and 35–40 cm), understory trees show a statistically significant decrease in CPL in randomized networks. On average, the observed CPL of the SANs representing each size class tended to be higher than their random counterparts. Also, the observed overall relationship between the CPL and tree size class is significantly different from the randomized SANs (Kolmogorov–Smirnov test; $p < 0.028$).

3. Discussion. We examined the effect of tree age, size, and species identity on the ability of individual trees to coexist in the understory of large trees. We did this by constructing SANs of trees in seven different size classes and analyzing them using three measures of network structure (NDD, CC, and CPL; Figure 2). Our use of networks in which nodes are species and edges represent coexistence in local neighborhoods represents a completely new way of characterizing forest structure. Combining the results we conclude that SANs derived from four of the seven size classes (representing stem diameters 10–15, 15–20, 30–35, and 35–40 cm) can be distinguished from random SANs based on either the CC or the CPL. Note that our randomization method ensures that the *number* of trees and species found in each size class is the same for observed and random SANs. Thus any differences

detected among the SANs must be due to differences in the spatial distribution of species. Our inability to distinguish the pattern of species association of the remaining three size classes from random patterns suggests that one or more random processes dominate the distribution of trees in these classes. This does not rule out the action of deterministic processes, such as niche differentiation. But it does suggest that random processes dominate these levels of community organization. However, we cannot be certain whether we would have arrived at a different conclusion had we chosen a different network construction on the one hand or different measures of network structure on the other.

There are many ways to depict a tree community as a network, and each approach examines different aspects of community structure. For example, we could have constructed networks in which the nodes represent individual trees instead of species. This issue reveals both a strength and a challenge of graph-based approaches. Because the graph-based representation is very flexible, it has gained wide use in many disciplines. But the ability of network analysis to reveal patterns depends crucially on which measures of structure are chosen. Our approach is motivated in part by our question (how does the species composition of understory neighborhoods change with increasing tree size?) and in part by our knowledge of how environmental and ecological processes can influence the spatial distribution of plants. We chose three specific measures of network structure because they have been found useful in other disciplines. In the following paragraphs we interpret the results of each analysis based on the specific characteristics of each metric and our knowledge of the processes that govern forest structure.

The increasingly right-skewed NDDs of the observed and random SANs (Figure 3) indicate that the number of species in the understory declines with increasing tree size. This observation can be explained on the basis of well-known patterns of vertical structure in forests. Many studies have shown that tree density in closed-canopy forests decreases with increasing tree size (Yoda et al. [1963], Connell et al. [1984], Weller [1987], Hubbell et al. [2001]). This is a strong and universal pattern of forests and as noted above has been explained by the laws of tree thinning and allometry (White [1981], Niklas [1994], Enquist et al. [1998]). Here we find no statistical difference between the

observed and random NDDs. This indicates that, with regard to the distribution of the number of heterospecific neighbors of large trees, deterministic processes that limit tree density and species richness, such as allometric scaling laws, outweigh the potentially nonrandom effects of species identity, habitat selection, and dispersal limitation. This interpretation speaks to issues of adaptation, competition, and species diversity. For example, the node degree results suggest that the number of heterospecific neighbors a given tree has, and thus the number of competitors it encounters, changes predictably as the tree grows and regardless of where in the forest it resides. How does this predictability influence the evolution of resource allocation strategies? With regard to resource management, knowing that species diversity is spatially invariant could simplify management decisions. As used here, the node degree is proportional to the number of species in a given area. Might it therefore be useful in finding diversity hotspots within managed lands?

The CC analysis revealed statistically significant difference from random networks in only one size class. In the 30–35 cm size class (the second-largest class), the observed CC was significantly lower than expected for a randomly organized forest. Low values of CC may indicate that species tend to be spatially overdispersed or that species are less ecologically compatible as understory neighbors of each other. Previous studies have shown that large trees tend to be overdispersed or more regular in their spatial arrangement relative to smaller trees (Vacek and Leps [1996], Kenkel et al. [1997]). Thus, the CC of the 30–35-cm size class is consistent with overdispersion, but why then do we not see a similar pattern in the largest size class of 35–40-cm trees? Note that there is also an overall decreasing relationship between the CC and size class in the SANs. That this pattern is common to both observed and random of SAN suggests the influence of changes in tree density and species richness related to thinning or allometry. That the CC was unable to detect a statistically significant difference in all but the 30–35-cm size class raises the question of whether the growth history in this class is somehow different than that of other size classes? Perhaps this age cohort experienced a disturbance (disease, wind throw, fire, etc.) in the past that did not affect the other classes as intensely?

The CPL of a network is a measure of the density of the connections in a network. We interpret it as the degree to which the spatial distributions of the different species are well mixed. The more

heterospecifics found in the understory on average, the more connections there will be among species (nodes) and the lower the CPL. Spatial segregation of species reduces the diversity of local neighborhoods and should therefore increase the CPL. In four of the seven size classes, we find no difference in CPL between the observed and random SANs (Figure 4b). This indicates that in these classes, network analysis could not detect spatial segregation. By contrast, SANs representing three of the size classes have CPLs that are significantly higher than the corresponding random SANs. We conclude that in these size classes random processes and deterministic limitations on tree density and species richness are insufficient to overcome spatial structure imposed by other forces. Whether those forces are related to habitat selection, species interactions, or some combination is unclear. Interestingly, Kenkel et al. [1997] found that the spatial pattern of a stand of *Pinus banksiana* was random in the intermediate age class (30–40 years) but clumped in early stages and overdispersed at later stages. In other words, the nonrandomness in the spatial pattern was bimodal with respect to age class. In our networks that represent the observed species distribution, the intermediate size class is close to the mean of the random distributions in CPL whereas nonrandom values appear to either side of the mean. Thus the observed multispecies pattern, though complex, is in some ways similar to the pattern Kenkel found for a single population. This suggests that the processes that govern spatial variation in intraspecific density also influence spatial patterns in species diversity.

The experiments reported here decompose the overall structure of the forest into separate components based on trees size. Would we have obtained different results if we had performed our analysis without regard to tree size or with less restriction on crown overlap? Previous analyses reported elsewhere (Fuller [2004]) speak to this question. In separate experiments, we analyzed the structure of SANs constructed using all neighboring trees that had at least 10% crown overlap with a target tree. This approach generates neighborhoods that are more inclusive of shorter trees growing closer to the trunk of the target trees. When we applied this construction, we found that the SAN representing the observed forest plot had significantly higher CPL and significantly lower CC than that of the SANs representing randomized plots. These results suggest that the overall structure of the forest plot is far from random. However, by decomposing the overall variation into separate tree size

classes as we have done here, we are able to detect the influence of random processes that are obscured when the size classes are grouped in a single analysis.

Our results show that the choice of which measure of topological structure to use is crucial to detecting nonrandom patterns. Here, we employed three commonly used metrics. These metrics are frequently used to identify patterns and understand processes in other areas of biology (Fell and Wagner [2000], Jeong et al. [2000], Strogatz [2001], Wagner [2001], Williams et al. [2002]). However, there are many other measures of network structure we could have used, such as network diameter (average minimum distance among nodes), node connectivity (the smallest number of nodes whose removal results in a disconnected graph), or edge connectivity (the smallest number of edges whose removal results in a disconnected graph). These metrics examine network structure from different perspectives than the measures we chose.

As mentioned above, graph representations are highly flexible and can represent community structure at many levels of resolution. A related motivation for graph-based analyses is that any one graph can be characterized in many ways, and thus there are many ways in which it could reveal nonrandom patterns. There is growing recognition among ecologists of the need to examine their data from multiple perspectives. Many natural and anthropogenic processes influence the geographic variation in species patterns, and it is often difficult to identify the key processes responsible for a specific pattern. Hence, several ecologists and statisticians have recommended using multiple measures of spatial structure when comparing competing hypotheses (Burnham and Anderson [2002], Fortin and Dale [2005], Grimm and Railsback [2005]). Network analysis may be particularly useful in this regard, as it incorporates multiple measures of structure into a single framework. Again, the choice of characterization method is crucial here. Whether using other metrics would have uncovered stronger differences between the observed and random SANs than we found here is uncertain. Clearly, many kinds of analyses can be performed using network analysis on a single dataset, and this is one of its strengths. There are ample possibilities to develop both statistically and biologically sensible indicators of ecological structure. We hope our example will encourage others to consider using graph-based approaches for the analysis of ecological patterns.

Acknowledgments. This manuscript is based in part on Dr. Fuller's doctoral dissertation, completed in the Biology Department of the University of New Mexico (UNM). Our research was funded in part by financial support from the National Science Foundation through awards to UNM (DGE 9553623 and DEB 0083422) and the University of Tennessee (IIS-0427471). We are indebted to Dr. James H. Brown for support and guidance during all stages of this research. MF gratefully acknowledges Dr. Louis Gross for support at UT. AW would like to thank the Santa Fe Institute for its continued support. We are also sincerely grateful to Dr. Suzanne Lenhart and an anonymous reviewer for constructive editorial comments.

REFERENCES

- R. Albert, H. Jeong, and A.L. Barabasi [1999], *Internet: Diameter of the World-Wide Web*, *Nature* **401**, 130–131.
- F.S. Baker [1950], *The Principles of Silviculture*, McGraw-Hill Co., New York, NY.
- A.L. Barabasi, R. Albert, and H. Jeong [2000], *Scale-Free Characteristics of Random Networks: The Topology of the World-Wide Web*, *Physica* **281**, 69–77.
- F. Briand and J.E. Cohen [1984], *Community Food Webs have Scale Invariant Structure*, *Nature* **307**, 264–267.
- K.P. Burnham and D.R. Anderson [2002], *Model Selection and Multimodel Inference*, 2nd ed., Springer, New York, NY.
- M.D.F. Cantwell and R.T.T. Forman [1993], *Landscape Graphs: Ecological Modeling with Graph Theory to Detect Configurations Common to Diverse Landscapes*, *Landscape Ecol.* **8**, 239–255.
- J.E. Cohen and Z.J. Palka [1990], *A Stochastic Theory of Community Food Webs. V. Intervality and Triangulation in the Trophic-Niche Overlap Graph*, *Am. Nat.* **135**, 435–463.
- J.E. Cohen, F. Briand, and C.M. Newman [1986], *A Stochastic Theory of Community Food Webs III. Predicted and Observed Lengths of Food Chains*, *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **228**, 317–353.
- J.H. Connell [1978], *Diversity in Tropical Rain Forests and Coral Reefs*, *Science* **199**, 1302–1310.
- J.H. Connell, J.G. Tracey, and L.J. Webb [1984], *Compensatory Recruitment; Growth; and Mortality As Factors Maintaining Rain-Forest Tree Diversity*, *Ecol. Monogr.* **54**, 141–164.
- H.V. Cornell and J.H. Lawton [1992], *Species Interactions, Local and Regional Processes, and Limits to the Richness of Ecological Communities: A Theoretical Perspective*, *J. Anim. Ecol.* **61**, 1–12.
- M. Dale [1977], *Graph Theoretical Analysis of the Phytosociological Structure of Plant Communities: The Theoretical Basis*, *Vegetatio* **34**, 137–154.

- M. Dale [1985], *Graph Theoretical Methods for Comparing Phytosociological Structures*, *Vegetatio* **63**, 79–88.
- M.R.T.P. Dale [1999], *Spatial Pattern Analysis in Plant Ecology*, Cambridge University Press, Cambridge, U.K.
- M.R.T.P. Dale and R.D. Powell [1994], *Scales of Segregation and Aggregation of Plants of Different Kinds*, *Can. J. Bot.* **72**, 448–453.
- J.A.W. Dunne, R.J. Martinez, and N.D. [2002], *Food-Web Structure and Network Theory: The Role of Connectance and Size*, *Proc. Natl. Acad. Sci. USA* **99**, 12917–12922.
- B.J. Enquist and K.J. Niklas [2001], *Invariant Scaling Relations Across Tree-Dominated Communities*, *Nature* **410**, 655–660.
- B.J. Enquist, J.H. Brown, and G.B. West [1998], *Allometric Scaling of Plant Energetics and Population Density*, *Nature* **395**, 163–165.
- B.J. Enquist, G.B. West, E.L. Charnov, and J.H. Brown [1999], *Allometric Scaling of Production and Life-History Variation in Vascular Plants*, *Nature* **401**, 907–911.
- S. Even [1979], *Graph Algorithms*, Pitman, London.
- D. Fell and A. Wagner [2000], *The Small World of Metabolism*, *Nat. Biotechnol.* **18**, 1121–1122.
- M.J.D. Fortin and M.R.T. Dale [2005], *Spatial Analysis: A Guide for Ecologists*, Cambridge University Press, New York, NY.
- M.M. Fuller [2004], *Species Association Networks of Tropical Trees have a Non-Neutral Structure*, in *Chance, Determinism and Community Structure: An Assessment of Ecological Neutral Theory*, Chapter 4, Ph dissertation, Department of Biology, University of New Mexico, USA.
- D. Garlaschelli, G. Caldarelli, and L. Pietronero [2003], *Universal Scaling Relations in Food Webs*, *Nature* **423**, 165–168.
- A. Gibbons [1985], *Algorithmic Graph Theory*, Cambridge University Press, Cambridge, U.K.
- R. Gould [1988], *Graph Theory*, Benjamin/Cummings, Menlo Park, CA.
- V. Grimm and S.F. Railsback [2005], *Individual-based Modeling and Ecology*, Princeton University Press, Princeton, NJ.
- J. Gross and J. Yellen [1999], *Graph Theory and its Applications*, CRC Press, Boca Raton, FL.
- F. Harary [1969], *Graph Theory*, Addison-Wesley, Reading, Massachusetts.
- S.P. Hubbell, J.A. Ahumada, R. Condit, and R.B. Foster [2001], *Local Neighborhood Effects on Long-Term Survival of Individual Trees in a Neotropical Forest*, *Ecol. Res.* **16**, 859–875.
- D.H. Janzen [1970], *Herbivores and the Number of Tree Species in Tropical Forests*, *Am. Nat.* **104**, 501–528.
- H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi [2000], *The Large-Scale Organization of Metabolic Networks*, *Nature* **407**, 651–654.
- N.C. Kenkel, M.L. Hendrie, and I.E. Bella [1997], *A Long-Term Study of Pinus Banksiana Population Dynamics*, *J. Veg. Sci.* **8**, 241–254.

- P. Legendre [1993], *Spatial Autocorrelation: Trouble or New Paradigm?* Ecology **74**, 1659–1673.
- B. Manly [1997], *Randomization, Bootstrap, and Monte Carlo Methods in Biology*, 2nd ed., Chapman & Hall, London.
- K. Mehlhorn and S. Näher [2000], *LEDA: A Platform for Combinatorial and Geometric Computing*, Cambridge University Press, Cambridge, U.K.
- M.E.J. Newman [2003], *The Structure and Function of Complex Networks*, SIAM Rev. **45**, 167–256.
- K. Niklas [1994], *Plant Allometry*, Chicago University Press, Chicago.
- S.T. O’Brien, S.P. Hubbell, P. Spiro, R. Condit, and R.B. Foster [1995], *Diameter, Height, Crown, and Age Relationships in Eight Neotropical Tree Species*. Ecology **76**, 1926–1939.
- S.L. Pimm [1984], *The Complexity and Stability of Ecosystems*. Nature **307**, 321–326.
- J.R. Porter [1989], *Modules, Models and Meristems in Plant Architecture, in Plant Canopies: Their Growth, Form and Function*, (G. Russell, B. Marshall, and P.G. Jarvis, eds.), Cambridge University Press, Cambridge, U.K, pp. 143–159.
- B.D. Ripley [1976], *The Second Order Analysis of Stationary Point Processes*. J. Appl. Prob. **13**, 255–266.
- J. Scott [2000], *Social Network Analysis: A Handbook*, 2nd ed.. Sage Publications, London, U.K.
- R.R. Sokal and F.J. Rohlf [1995], *Biometry*, 3rd ed., Freehman and Company, New York.
- S.H. Strogatz [2001], *Exploring Complex Networks*, Nature **410**, 268–276.
- D.K. Urban and T. Keitt [2001], *Landscape Connectivity: A Graph-Theoretic Perspective*, Ecology **82**, 1205–1218.
- S. Vacek and J. Leps [1996], *Spatial Dynamics of Forest Decline: The Role of Neighbouring Trees*, J. Veg. Sci. **7**, 789–798.
- A. Wagner [2001], *How to Reconstruct a Genetic Network from n Single-Gene Perturbations in Fewer than n^2 Easy Steps*, Bioinformatics **17**, 1183–1197.
- S. Wasserman and K. Faust [1994], *Social Network Analysis*, Cambridge University Press, Cambridge, U.K.
- D.J. Watts and S.H. Strogatz [1998], *Collective Dynamics of Small-World Networks*, Nature **393**, 440–442.
- D.E. Weller [1987], *A Reevaluation of the $-3/2$ Power Rule of Plant Self-Thinning*, Ecol. Monogr. **57**, 23–43.
- G. West, J.H. Brown, and B. Enquist [1999], *A General Model for the Structure of Plant Vascular Systems*, Nature **400**, 664–667.
- J. White [1981], *The Allometric Interpretation of the Self-Thinning Rule*, J. Theor. Biol. **89**, 475–500.
- R.J.M. Williams and N.D. Martinez [2000], *Simple Rules Yield Complex Food Webs*, Nature **404**, 180–183.

R.J. Williams, E.L. Berlow, J.A. Dunne, A.L. Barabási, and N.D. Martinez [2002], *Two Degrees of Separation in Complex Food Webs*, Proc. Natl. Acad. Sci. USA **99**, 12913–12916.

K. Yoda, T. Kira, H. Ogawa and K. Hozumi [1963], *Self-Thinning in Overcrowded Pure Stands under Cultivated and Natural Conditions (Intraspecific Competition Among Higher Plants XI)*, J. Inst. Polyt., Osaka City Univ., Ser. D **14**, 107–129.