Quantifying co-occurrence patterns in space and time across aquatic systems with network analysis

Christopher J. Patrick[1]*, Kyle Cavanaugh[2], Talina Konotchick[3], and Hannes Peter[4]
[1]Smithsonian Environmental Research Center, Edgewater MD, USA – patrickc@si.edu
[2]Department of Geography, University of California, Los Angeles, CA, USA – kcavanaugh@geog.ucla.edu
[3]J. Craig Venter Institute, La Jolla, CA, USA – tkonotch@jcvi.org
[4]Institute of Ecology, Lake and Glacier Ecology Research Group, University of Innsbruck, Innsbruck, Austria – hannes.peter@uibk.ac.at

Abstract

Network analytic techniques have been used to analyze the connections between organisms, which is a central theme in community ecology. We analyze species co-occurrence patterns from four aquatic ecosystems (the coastal ocean, kelp forest, streams, and lakes) using decadal-scale datasets. Both spatial and temporal co-occurrence networks are analyzed using a set of commonly used network metrics and compared to scale-free and random model networks. Additionally, we present a conceptual model that links increases in spatial connectivity (and species dispersal) to increases in the size and connectedness of spatial co-occurrence networks and decreases in the size and connectedness of temporal co-occurrence networks within those metacommunities. We then discuss the results from each ecosystem in relation to this conceptual model. For temporal co-occurrence networks, size metrics (i.e., diameter and average path length) were high in the stream case study, and both kelp forests and streams showed high connectivity. For spatial co-occurrence networks, the coastal ocean had the high connectivity and the lakes and streams had relatively high size metrics. Stability of spatial networks through time was variable between the four ecosystems, with kelp forests showing the most year-to-year variability across all metrics. Species co-occurrence patterns from long-term ecological datasets can yield valuable insights into spatial and temporal dynamics of ecosystems.

Section 1. Introduction

Understanding the processes that structure communities is a central theme of community ecology. Recently, network analytic techniques have been developed that may be used to analyze the connections (i.e., edges) between organisms (i.e., nodes) (Sole and Montoya 2001; Montoya et al. 2006; Ings et al. 2009; Barberan et al. 2012). Such links can include predator-prey interactions, mutualistic interactions such as flower-pollinator relationships, facilitative interactions such as co-metabolism, and spatial/temporal co-occurrence. Networks developed by linking species that co-occur in space or time (co-occurrence networks) are useful tools for understanding the complexity and sensitivity of these interactions and can be used to infer the processes that structure communities (Shade et al. 2010; Steele et al. 2011). Network topology describes the structure of a network (i.e., the pattern of nodes and edges) and can be quantified with statistics (Blonder et al. 2012). Most ecological networks have many links and are densely clustered (Montoya and Sole 2002). Moreover, most species are linked by only a few intermediate species, a phenomenon known as small-world behavior, which has been described for other networks such as the Internet or the power grid (Albert and Barabasi 2002; Albert et al. 2000). However, the underlying causes of these patterns remain obscure (Montoya et al. 2006).

Identifying interactions in natural communities in many cases requires observation and/or experimental manipulation, which is often limited to a few species pairs (Olesen et al. 2011). One way to overcome this limitation and gain a greater understanding of the entire community of connections is the analyses of co-occurrence patterns in space and time, which

can help to identify potential links between many species (Ruan et al. 2006; Fuhrman and Steele 2008). Co-occurrence patterns might be shaped by species interactions, by common preferred environmental conditions, or by dispersal pressure. Spatial co-occurrence explicitly considers spatial overlap in species distributions across sites. For example, if a researcher has observations of ecological communities at multiple sites, a spatial co-occurrence network can be constructed whereby each site is represented by a node in the network, and nodes with the same species are linked to one another. Understanding spatial co-occurrence patterns is one of the goals of metacommunity analysis and spatial co-occurrence networks can be a useful tool for advancing our understanding of metacommunity dynamics (Leibold and Mikkelson 2002). Alternatively, co-occurrence networks can be used to consider temporal co-occurrence within a single habitat and community. Here, the links between nodes represent significant correlations in species abundance through time. This is akin to examining temporal synchrony between taxa living in the same environment (Rusak et al. 2008).

Here, we provide a conceptual framework for how co-occurrence patterns might be used to investigate drivers of community composition, such as dispersal and environmental variability, and use this to propose a set of working hypotheses. First, the number of temporally or spatially co-occurring species and the number of links between these species might provide information about the strength of underlying mechanisms (Fig. 1). Assuming that a low dispersal rate within a metacommunity promotes species sorting (Leibold et al. 2004), species interactions might be able to translate into more connected temporal co-occurrence networks on the local scale within individual habitats. On the other hand, massive dispersal, such as under a mass effect scenario (Leibold et al. 2004), likely dominates over species interactions, and therefore temporal co-occurrence at the local scale might be more variable through time, resulting in fewer temporal connections.

Dispersal and colonization affect regional species pools, and hence, high dispersal rates might lead to large, well-connected spatial co-occurrence networks (Fig. 1). Environmental variability might form another gradient shaping co-occurrence patterns. Whereas asynchronous fluctuating environmental conditions may disrupt spatial synchrony and therefore co-occurrence, stable conditions or synchronous environmental fluctuations may allow enough time for the establishment of species interactions. Hence, along a gradient of synchrony in environmental variability, less synchronous conditions may support rather small networks of spatially connected species, whereas stable conditions may allow for larger and denser co-occurrence networks. Higher levels of dispersal should also maintain the stability of spatial networks through time, even in the face of environmental change, since organisms will rapidly be able to respond by colonizing new areas quickly (Fig. 1).

Based on the concept outlined above, we hypothesize that as spatial connectivity between locations increases, the
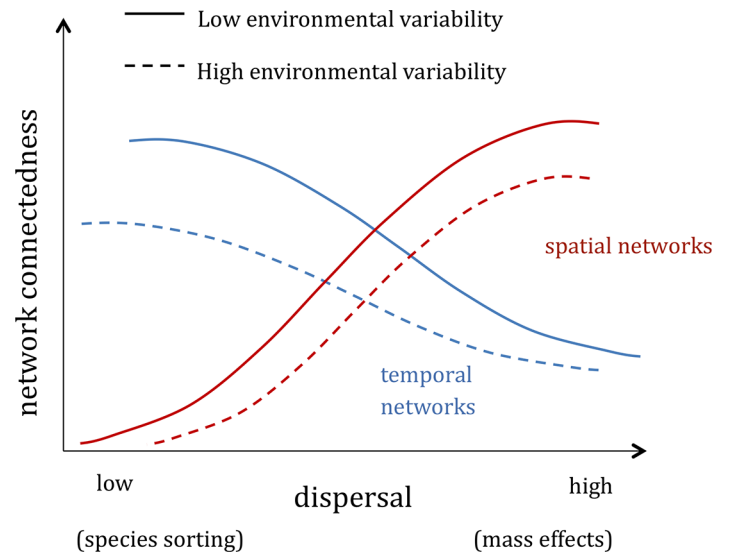

Fig. 1. Conceptual framework of how temporal and spatial networks may inform about meta-community properties. Low dispersal rates and stable environmental conditions may allow for species sorting and large correlational networks over time. Such networks feature great numbers of species and/or many connections among species. Higher dispersal rates may lead to mass effects and sustain synchronicity in metapopulation abundance and therefore spatial correlations. Increasing environmental variability might interrupt the formation of species interactions or synchronicity and hence lead to smaller, less connected co-occurrence networks.

size and connectedness of spatial co-occurrence networks increases whereas the size and connectedness of temporal co-occurrence networks for individual communities within those metacommunities decreases (Fig. 1). We test the predictions of the outlined framework using four decadal-scale datasets of aquatic communities in space and time as separate case studies: (1) phytoplankton species in lakes across Sweden, (2) invertebrates living in streams across a region of the eastern United States, (3) mixed communities of algae, invertebrates, and fish living in kelp forests off the coast of southern California, USA, and (4) ichthyoplankton found in the coastal ocean of the Southern California Bight.

Section 2. Materials and methods

Data sources

Data were collected from the Swedish University of Agricultural Sciences, the Santa Barbara Coastal Long Term Ecological Research Project (SBC LTER), the National Park Service, and the California Cooperative Ocean Fisheries Investigation (CalCOFI) (Table 1). Each dataset contained species count data for multiple locations that had been sampled annually. Sampling was not consistent for all sites in the data sets, so before our analyses, we removed sites that had not been sampled in at least 75% of the years. All data used in the analysis were absolute abundances of individuals, not proportions, so our analysis avoids the issues that come with using

Table 1. Description of the data sets for each ecosystem.

System	Nr taxa	Nr sites	Nr years	Average site Shannon diversity	Average site α richness	Average cross site dissimilarity in a year (Bray Curtis)	Average species pool dissimilarity across years (Jaccards Index)	Community type	Source
Boreal lakes	183	67	17	1.785	31.916	0.477	0.233	Phytoplankton	Swedish University of Agricultural Sciences
Small streams	121	36	27	1.987	23.47	0.447	0.384813	Invertebrates	USGS
Kelp forests	97	11	12	1.893	27.531	0.762	0.271	Algae, invertebrates, fish	Santa Barbara Coastal Long Term Ecological Research Project
Southern CA Bight, USA	131	63	28	0.624	5.279	0.9	0.567	Fish eggs and larvae	California Cooperative Oceanic Fisheries Investigation

compositional data (Friedman and Alm 2012). Moreover, correlational co-occurrence networks have been demonstrated to be sensitive to sampling size, however to plateau at about 25 samples (Berry and Widder 2014).

CalCOFI Icthyoplankton time series

The decline of the Pacific sardine fishery in the 1940s led to the creation of the CalCOFI ichthyoplankton time series (for a review, see Ohman and Venrick 2003). The ichthyoplankton surveys are used for the detection and appraisal of fishery resources and the study of population dynamics (including age specific data) and systematics of fishes (Smith and Richardson 1977). Larval abundance has been shown to be a reliable index to estimate biomass of later life stages of pelagic fish, including the economically important sardine, anchovy, and Pacific mackerel (Smith and Richardson 1977). In this analysis, we include both eggs and larvae from all species found in surface tows from the spring (Feb-May) cruises of 1984-2011 from the 75 station pattern (i.e., lines 76.7 to 93.3; the Southern California Bight; oceaninformatics.ucsd.edu). Eggs and larvae can have different ecological interactions so we treat them as separate nodes in our analysis. Ichthyoplankton data originate from the NOAA National Marine Fisheries Service, Southwest Fisheries Science Center, collected by NOAA CalCOFI survey teams and were identified by the SWFSC Ichthyoplankton Laboratory (Kramer et al. 1972; Smith and Richardson 1977).

Kelp forest time series

The giant kelp, *Macrocystis pyrifera*, in southern California grows on rocky reefs in shallow waters (5-30 m depth). The patchy distribution of rocky substrate along the coast of California causes kelp forests to occur in discrete stands that range from tens of meters to kilometers in alongshore length. Neighboring stands are typically separated by distances on the order of hundreds of meters (Cavanaugh et al. 2014). Kelp forests provide both food and habitat for a diverse array of algae, invertebrates, and fish (Dayton 1985; Graham et al. 2007). Many of these species are relatively sedentary as adults, with home ranges on the order of 0.1–10 km (Kinlan and Gaines 2003). However, the larval stages of these species are dispersed by ocean currents, and they can travel distances that are orders of magnitude greater than their home ranges (100s km; Kinlan and Gaines 2003). Previous network analysis of kelp forest communities has found that these systems exhibit small-world network properties, whereby well-connected hub populations increase the connectivity and robustness of the kelp forest metapopulation (Watson et al. 2011).

Kelp forest community surveys were conducted at 11 sites in the Santa Barbara Channel each summer between late July and early August from 2001 to 2012 through the SBC LTER project (http://sbc.lternet.edu). At each site, divers measured the abundance of giant kelp, understory macroalgae, invertebrates, and reef fish along a permanent 40 m transect. Giant kelp fronds were counted in the 2 m × 40 m area around each

transect. The abundance of invertebrates and understory macroalgae was assessed in six 1 m^2 quadrats, four 2 m × 20 m band transects, or 80 uniformly spaced points along the transect, depending on the species. Fish abundance was measured in a 2-m wide swath that was centered along the transect.

Lake time series

We analyzed summer phytoplankton data (July or August) from 67 small, oligotrophic, mostly humic lakes in Sweden from 1992 through 2008 and covering a latitudinal gradient between 56°N to 68°N; this dataset was made available through the Swedish National Lake Inventory (http://www.slu.se/vatten-miljo, Willén 2007). Samples for the inventory were taken from the middle of the lakes, preserved with an acidic Lugol's solution, and determined using the Utermöhl technique (EN15204, 2006) by a single certified laboratory.

Stream invertebrate time series

Data were extracted from the United States Geological Survey (USGS) Chester County 1970–1997 data set in the Aquatic Community Classification database maintained by The Nature Conservancy (TNC). This data set includes aquatic invertebrate data from greater than 1000 samples at more than 50 sites in eastern Pennsylvania. The county encompasses 1973 km^2 of land and has an elevation gradient ranging from 21 to 326 m asl (CCPC 2009). Invertebrate samples were collected with a modified Hess sampler from 10 separate 30.5 × 30.5-cm plots within the same riffle. The 10 samples were pooled before sorting and identification. Invertebrate samples were sorted in the laboratory, and all individuals were identified to genus except for Chironomidae and Planaridae, which were identified to family.

Network construction

There are numerous ways to explore temporal and spatial co-occurrence patterns (e.g., checkerboard scores, mutual information, local similarity analysis), and to our knowledge, a thorough review of the limitations of the individual metrics is still lacking. We chose ranked correlations based on absolute abundance data to compare between rather different types of ecosystems to avoid spurious correlations, which may inflate networks and reduce specificity (Friedman and Alm 2012). A recent analysis of modeled metacommunities has shown that co-occurrence networks based on Spearman and Pearson correlation coefficients outperform other metrics such as similarity based metrics (Bray-Curtis, Sörensen) in terms of sensitivity and specificity (Berry and Widder 2014).

We calculated all possible Spearman rank correlations between taxa for the temporal and spatial datasets using the packages 'Hmisc' and 'qvalue' in R v. 2.13.0 (R Development Core Team 2014). Significance was adjusted for multiple testing using the BH procedure. An edge, or link between two species, in the network was considered if the absolute correlation coefficient was larger 0.3 and significant at the 95%

significance level (Junker and Schreiber 2008). Similar cutoffs were previously used to construct co-occurrence networks of soil microbial assemblages (Barberan et al. 2012), and recent work on microbial metapopulations based on generalized Lotka-Volterra models confirms the high accuracy and sensitivity of co-occurrence networks constructed from rank-correlation matrices (Berry and Widder 2014). Networks were built using unweight edges (i.e., interactions were considered irrespective of the sign and strength of the correlation), visualized, and analyzed using the R packages 'network' and 'igraph' as well as the freeware software Cytoscape (Saito and Smoot 2012) and gephi (Bastian et al. 2009). For each data set, spatial co-occurrence networks were created for each year for which there was data (n = 12–28 years in each dataset) and temporal networks for each spatial location (n = 11–67 sites in each system; Fig. 2).

We calculated the following topology metrics for each network: clustering coefficient, average node connectivity, diameter of the network, and average path length between nodes (species) of each network. The clustering coefficient measures the degree to which nodes of the network tend to cluster together and is a measure of the connectedness of the network and is indicative of the degree of species linkage in the network (Newman 2003). Node connectivity is a measure of the number of nodes that need to be removed to disconnect the remaining nodes from each other and is a measure of network robustness (Diestel 2005). In an ecological context, this node connectivity can also be inferred to be a measure of stability of interactions. Average path length is the average number of steps along the shortest paths for all possible pairs of network nodes, and diameter is the greatest distance between any pair of nodes. In our analysis, both diameter and average path length are considered measures of the size of the network. Larger networks are less connected, meaning that the likelihood of a strong connection between any two randomly selected species is low.

Statistical analyses

The network attributes (average path length, diameter, and cluster coefficient) for the temporal and spatial co-occurrence networks were compared across systems using one-way ANOVA. All of those attributes met the assumptions of normality and equal variance. Average node connectivity was compared among groups using a nonparametric Kruskall-Wallis rank sum test. The coefficient of variation for spatial co-occurrence networks was calculated for each system and compared across systems to determine if temporal stability in network strength matched our conceptual model.

Random and scale free networks

We compared network properties of the observed networks to scale free and random reference networks of the same size. Scale-free and random network versions of each network were calculated using Graph Crunch 2 (Kuchaiev et al. 2011).
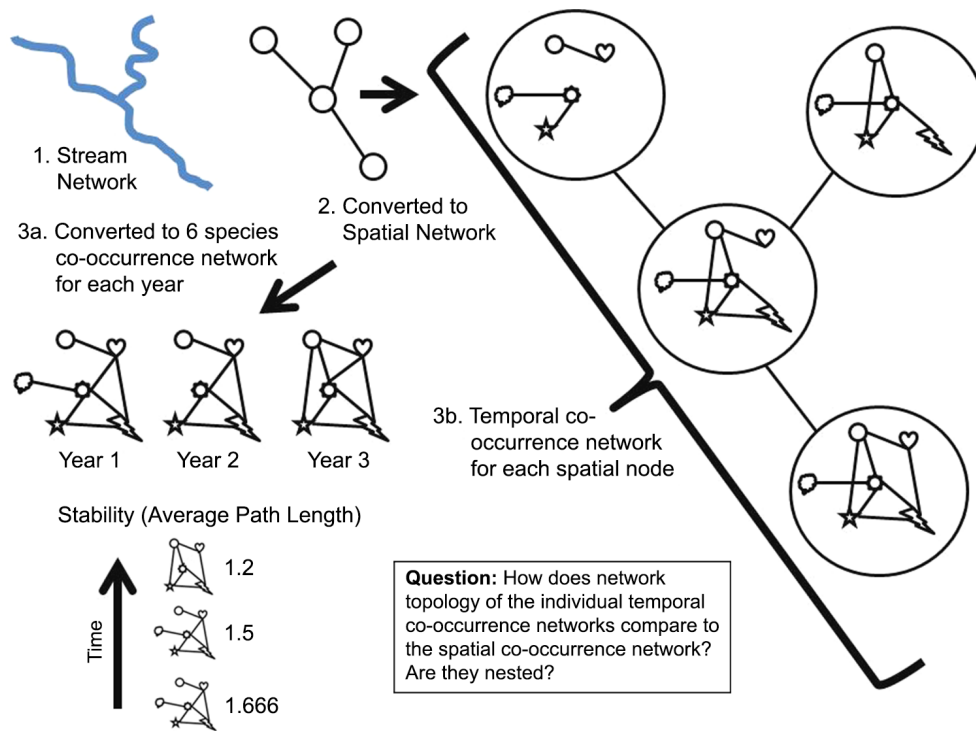
Fig. 2. Conceptual example of the network analysis. Shows the two contrasting methods that we used in the analysis. First a stream network (1), is divided into 4 spatial nodes (2), each with 3 years of data. From those data we either build a spatial co-occurrence network for each year (3a) and contrast those networks through time, or we build a temporal co-occurrence network for each spatial location (3b) and contrast them to each other. Building off this basic analysis (3a and 3b), we will be able to compare these results across systems to address general questions regarding the effect of aquatic community type on network topology.

Scale-free networks form by preferential attachment of new nodes to existing and well-connecting nodes. Networks such as the internet, power grids, and some biological networks are scale free (Albert and Barabasi 2002; Albert et al. 2000). On the other hand, disturbance may disrupt the network and lead to random co-occurrence patterns. To obtain a normalized measure of network stability, we calculated the distance, *D*, between the observed network properties and the scale-free and random models as

$$D = \frac{(sf - o)}{(o - r)}$$ (1)

where *sf* is the scale free vector, *r* is the random value, and *o* is the observed value. On this scale, a value of 1 for *D* means the network is perfectly balanced between random and scale free; values that are > 1 are more scale free; and values < 1 are more random.

Section 3. Results

Temporal co-occurrence networks

For the temporal co-occurrence networks, several of the systems differed significantly from one another in terms of average path length between nodes (species), network diameter, and clustering coefficient (Table 2, Fig. 3). Average node connectivity did not differ among systems ($\chi^2 = 4.22$, df = 3, $P = 0.24$). Streams had the highest average path length followed by kelp forests and the open ocean (which did not differ from one another; $P = 0.94$), and then lakes. An identical pattern was observed for network diameter. Lakes and oceans had the two lowest clustering coefficients and did not differ significantly from one another ($P = 0.53$). Kelp and stream ecosystems had higher clustering coefficients and did not differ significantly from one another ($P = 0.36$).

Spatial co-occurrence networks

For the spatial co-occurrence networks, the systems again differed significantly from one another in terms of average path length, network diameter, and clustering coefficient (Table 3, Fig. 4). Oceans and kelp forest had significantly lower average path lengths and smaller network diameters than lakes and streams. Oceans had significantly higher clustering coefficients than streams, but these two systems did not differ from lakes and kelp forest communities.

Temporal variation in spatial network properties differed between the four ecosystem types (Fig. 5, Table 4). Lakes phytoplankton communities consistently had the lowest variation in network density (i.e., average path length and clustering

Table 2. ANOVA table comparing the stability of temporal co-occurrence network parameters in different aquatic ecosystems.

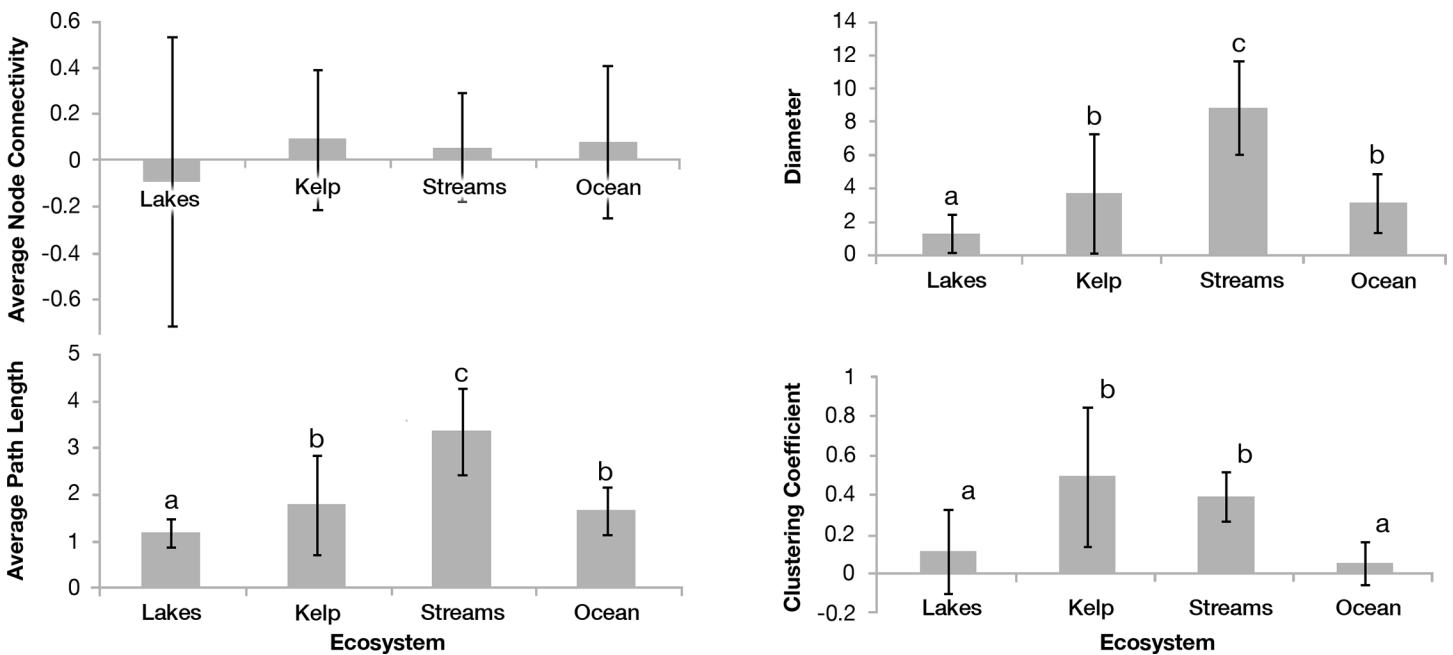| Source	Df	Sum sq	Mean sq	F	P
Average path length	3	105.35	35.12	86.67	<0.0001
Residuals	156	63.21	0.41		
Diameter	3	1351.00	450.30	114.20	<0.0001
Residuals	173	682.10	3.9		
Clustering Coefficient	3	3.29	1.096	42.54	<0.0001
Residuals	115	2.96	0.026		

Fig. 3. Comparison of the network characteristics of temporal co-occurrence networks for communities found in four different aquatic ecosystems. Ecosystem type is displayed on the x-axis, response variable is displayed on the y-axis. Ecosystems were found to differ significantly in terms of average path length between any two nodes in the network, diameter of the network, and clustering coefficient for the network (Table 2). Significant differences ($\alpha = 0.05$) between ecosystems for significant ANOVA models were determined by Tukey's HSD test and displayed as letters over the bars; different letters denote statistically different groups.

coefficient) through time. Streams had the second lowest variation for average path length and network diameter, but the second highest variation for clustering coefficient. Open ocean communities had the second lowest variation for clustering coefficient, and the third lowest for average path length and for network diameter. Kelp forest communities displayed the most variable spatial networks through time in every category.

Random and scale free networks:

The ratios of clustering coefficient and path length showed different patterns within each system type (Fig. 6). Average path length in streams tended to be in between scale-free and random networks in most years but matched random networks in 1976 and 1982. Clustering coefficient for stream networks was more variable throughout the time series. Average path length in lakes was very stable and was in the

middle between scale free and random for the entire time series. Clustering coefficient in lakes was more variable than average path length, however, save for dips toward random in 1992 and 1994, clustering coefficient in lakes was closer to scale free than random. Kelp forests were much closer to random than the preceding networks. Average path length ratio was less than 1 in all but 3 years, and clustering coefficient was less than one in all but 2 years. Open ocean networks exhibited oscillations between random and scale free that were not present in the other systems. Average path length ratio tended to be greater than 1, but moved between 1.5 and 4.5, indicating scale free tendencies. Conversely, the clustering coefficient ratio tended to move between zero and one indicative of random networks. For cluster coefficient in the open ocean, there was a period from 1999 to 2005 where the oscillatory dynamics ceased (Fig. 6 D1: solid line).

Table 3. ANOVA table comparing the stability of spatial co-occurrence network parameters in different aquatic ecosystems.

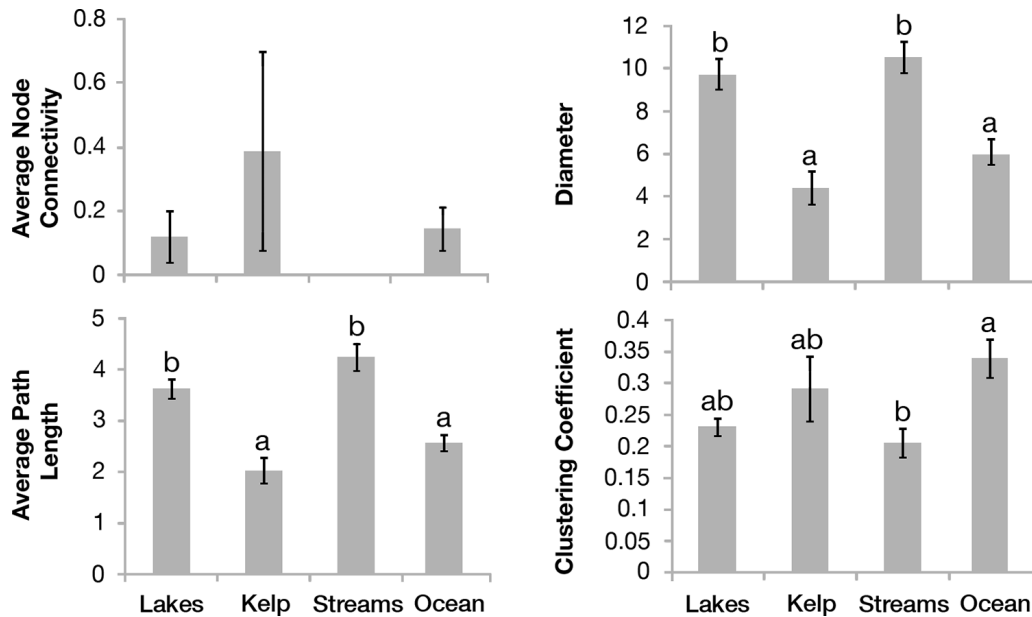| Source	Df	Sum sq	Mean Sq	F	P
Average path length	3	61.32	20.44	18.60	<0.0001
Residuals	81	89.02	1.10		
Diameter	3	506.10	168.70	17.03	<0.0001
Residuals	81	802.30	9.91		
Clustering Coefficient	3	0.27	0.091	4.763	0.0042
Residuals	79	1.52	0.019		

Fig. 4. Comparing the stability of spatial co-occurrence networks for communities found in four different aquatic ecosystems. Ecosystem type is displayed on the x-axis; response variable is displayed on the y-axis. Ecosystems were found to differ significantly in terms of average path length between any two nodes in the network, diameter of the network, and clustering coefficient for the network (Table 3). Node connectivity for streams was effectively zero. Significant differences ($\alpha = 0.05$) between ecosystems for significant ANOVA models were determined by Tukey's HSD test and displayed at letters over the bars; different letters denote statistically different groups.

Section 4. Discussion

Our conceptual model predicted that in ecosystems that are highly connected via dispersal we would observe highly connected spatial co-occurrence networks and weakly connected temporal co-occurrence networks. The results of our analyses of streams and oceans matched this conceptual model: the relatively open ocean system was characterized by highly connected spatial co-occurrence networks and weakly connected temporal co-occurrence networks, whereas the dispersal-limited stream system showed the opposite pattern. However, the kelp forest and lake networks deviated somewhat from our expectations. Kelp forest communities had well-connected spatial co-occurrence networks, as would be expected from an open system, but also had surprisingly highly connected temporal co-occurrence networks. These deviations from our conceptual model suggest that processes other than spatial connectivity are influencing the topology of the co-occurrence

networks. We suspect that variation in disturbance regimes and the life history of the resident organisms may be responsible for some of the deviations from our predictions. For example, frequent asynchronous disturbance could act to reduce the connectivity of spatial occurrence networks by causing increased variability in community structure across sites. On the other hand, frequent synchronous disturbance could increase both the connectivity of temporal occurrence networks and the variability in spatial co-occurrence networks through time. Below we discuss the detailed temporal and spatial co-occurrence patterns of each aquatic ecosystem in turn and place those results in the context of our conceptual model.

Streams

Stream metacommunities are characterized by a hierarchical dendritic network of habitats moving from stable larger streams to successively smaller and less environmentally stable
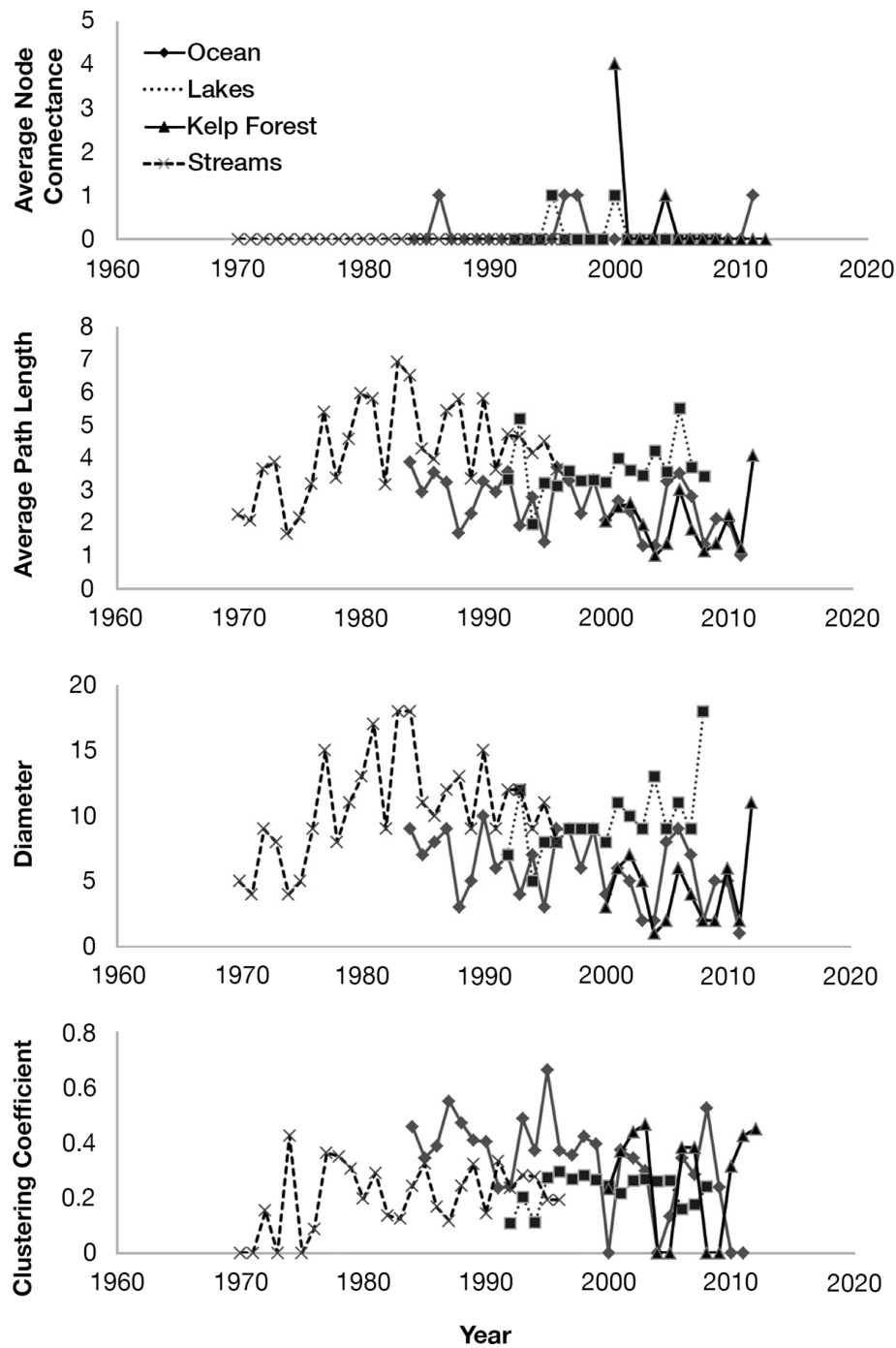
Fig. 5. Changes in spatial co-occurrence networks through time for ichthyoplankton (gray diamonds), lakes (dotted squares), kelp forest (black triangles), and streams (dashed black).

Table 4. Coefficient of variation for different network stability metrics through time for each system.

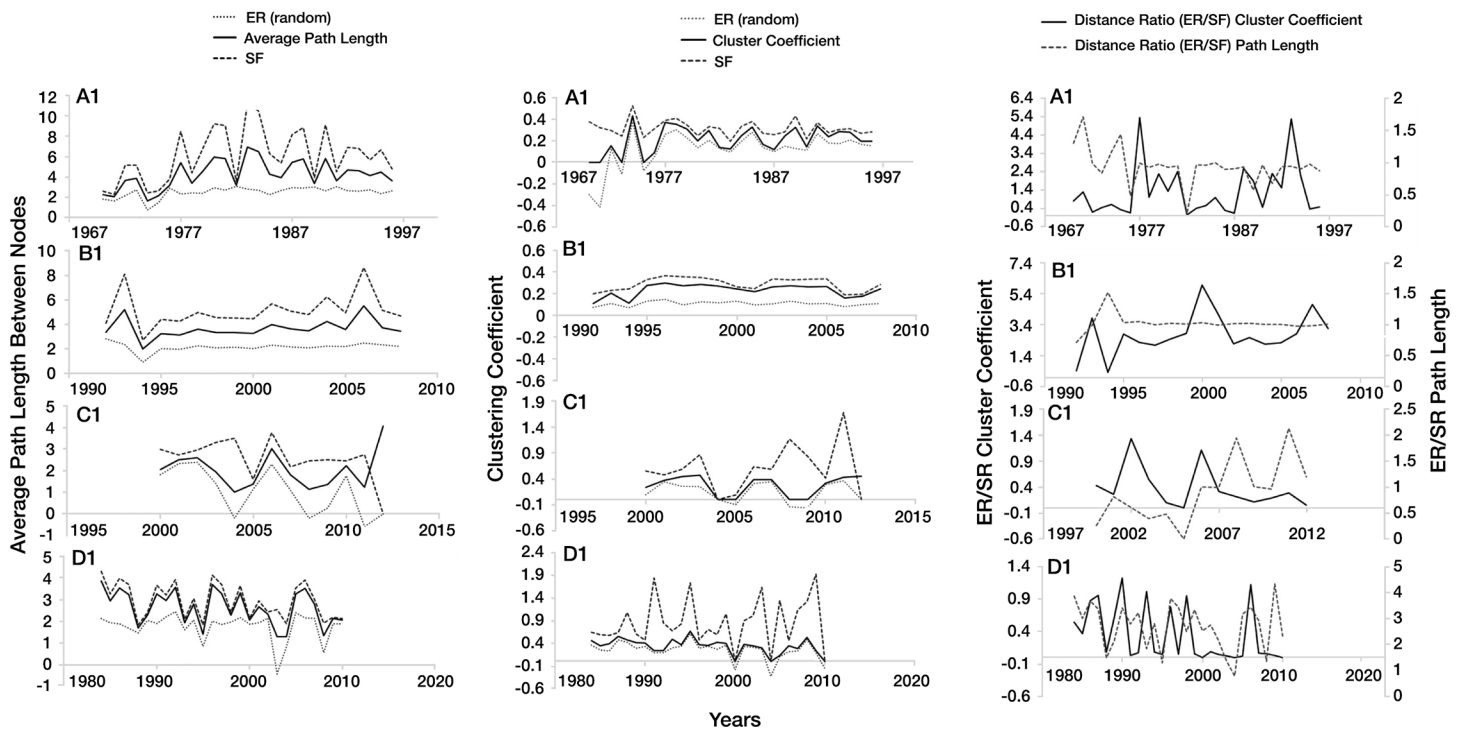| Measure	Ocean	Lakes	Kelp	Streams
Average node connectivity	2.49	2.82	2.91	0
Average path length	0.33	0.22	0.43	0.33
Diameter	0.44	0.29	0.64	0.37
Clustering coefficient	0.48	0.26	0.64	0.59

Fig. 6. Observed networks compared to random (ER) and scale free (SF) networks. A = Streams, B = Lakes, C = Kelp Forest, and D = Open Ocean. 1 = average path length, 2 = clustering coefficient, and 3 = ratio of random/scale free.

second- and first-order streams. The smaller the stream, the more likely it will be affected by disturbance events like hydrologic flashiness and period droughts (Poff and Ward 1989). Both of these events are capable of severely affecting the composition of a local community. There is evidence for dispersal constraints affecting the composition of macroinvertebrate communities in streams, particularly further up in the stream network (Patrick and Swan 2011; Swan and Brown 2010). There is also evidence that the stream communities used in this analysis experienced environmental filtering (Patrick and Swan 2011).

The data set used for these analyses was composed of small streams. Seventy-five percent of the streams were third order or smaller, and of the remaining 25%, one was a fifth order and the rest were fourth order. Given the size of the streams, it is likely that frequent disturbance and re-colonization was an important part of the ecology of these communities. Variability in these disturbance and recolonization patterns may have led to the observed low temporal synchrony in these communities. These particular communities are undergoing a great deal of turnover between 1970 and 1997 offering support for this interpretation of the results (Patrick and Swan 2011).

The relatively large size and connectedness of temporal co-occurrence networks in streams matched our conceptual model (Fig. 1) and could be the result of dispersal limitation allowing communities in the metacommunity to reach their equilibrium behavior. However, frequent disturbance altering

the composition of the communities may have prevented the communities from slowly developing stronger spatial co-occurrence associations, maintaining the relative strength of the networks as they are. This explanation fits with the observation that while the streams have poorly connected spatial co-occurrence networks, network strength is remarkably consistent through time, surpassed only by the relatively disturbance-free lake phytoplankton communities.

Ocean

The Southern California Bight, the area in which the ichthyoplankton were sampled, can be considered a fairly open ecosystem in terms of dispersal and connectivity compared with the other ecosystems in this study. We found that the network connectedness of this community was low for temporal networks and high for spatial networks relative to the other ecosystems, again matching our conceptual model (Fig. 1). Our results also indicated low to medium size metrics (compared with the other ecosystems) for both the temporal and spatial dimensions. Perhaps the low temporal connectedness is related to known and temporally variable oceanographic conditions.

Ichthyoplankton abundance has been linked to oceanographic conditions such as temperature fluctuations associated with climate variability including El Niño-Southern Oscillation (ENSO) and Pacific Decadal Oscillation (PDO), events that vary on interdecadal to decadal time scales (Smith

and Eppley 1982; Moser et al. 2000; Hsieh et al. 2005). Additionally, variability in realized larval connectivity and larval abundance through time has been observed on seasonal to interannual time scales (Walker et al. 1987; Watson et al. 2010). The environmental variability in this system might also explain the smaller network size (relative to the other systems) metrics seen in both time and space. Additionally, spatial patterns of ichthyoplankton counts may be related to physical features such as current, mesoscale eddies and fronts, that themselves are variable in space and time. A recent study showed similar simulated potential and realized larval connectivity for two fish species in the Southern California Bight, indicating the importance of circulation processes affecting spatial distribution (Watson et al. 2010). It should be noted that ocean plankton patchiness in time and space can make sampling of this environment difficult (Haury et al. 1978). For stability of spatial networks through time, the coastal ocean was ranked in the middle compared with the other ecosystems. Through time, average path length was shown to be scale free, indicating strong robustness and the presence of key 'hub' species; on the other hand, connectivity and distance ratio tended toward random, perhaps a result of the stochastic nature of the oceanic environment.

Kelp

The kelp and ocean system are very similar in terms of network characteristics (see Figs. 3, 4, 5). Both systems also track each other in the timing of the stability of the spatial co-occurrence networks (Fig. 5). The kelp forests sampled in the dataset are nested within the Southern California Bight so both systems experience similar oceanographic and climate influences. The high connectedness of kelp spatial co-occurrence networks matched our expectations, as kelp forests are a relatively open system and most kelp-associated species have relatively high larval dispersal capabilities. Large amounts of dispersal would increase the connectivity of kelp forests, giving many species the opportunity to occur in most or all of the sites when conditions are suitable for colonization. The kelp forests sites were all located in a relatively small region (within ~ 70 km), and so there was not much spatial variability in environmental conditions such as sea surface temperature, nutrient levels, and wave disturbance. The high connectedness of the temporal co-occurrence networks may have been due to the fact that many species responded similarly to synchronous environmental variability, disturbance, or variability in larval transport due to changes in circulation patterns. Previous work has shown that giant kelp forest food-web structure and diversity is dependent on the abundance of giant kelp itself (Foster and Schiel 1985; Graham 2004; Byrnes et al. 2011). It follows that changes in the abundance of giant kelp could impact many kelp-associated species in a similar manner. Whereas network stability of spatial and temporal co-occurrence networks was high, the long-term stability of kelp forest spatial co-occurrence networks was low. This matches

with previous observations of high variability in the food-web structure of kelp forest communities (Byrnes et al. 2011). Giant kelp abundance is highly variable in time (Cavanaugh et al. 2011), and this variability could be influencing the population dynamics of kelp-associated species.

Lakes

Compared with ocean and stream ecosystems, lakes represent a patchy environment embedded in an otherwise mostly hostile environment for phytoplankton communities. Nevertheless, lake phytoplankton communities form large spatial co-occurrence networks. Phytoplankton species may disperse via water flow, on animals, and via the atmosphere, and individual dispersal capabilities depend on the ability to withstand transport conditions. Many phytoplankton species are known to form spores, which better endure dry and high-radiation conditions (Kristiansen 1996). One might assume that dispersal among lakes might be limited and related to spatial distance, however, regionally abundant taxa have been shown to also be widespread (Östman et al. 2010). In the same study, the authors report that phytoplankton dispersal might be less important in structuring local communities when environmental variation was high. The dataset analyzed here covered a latitudinal gradient between 56°N and 68°N, and latitudinal diversity gradients, which are known for other microbes (e.g., Fuhrman and Steele 2008; Lozupone and Knight 2007), might explain some of the spatial co-occurrence patterns (i.e., some species gradually disappearing toward the North). However, phytoplankton community composition and function along this gradient have been shown to exhibit significant shifts depending on growing season length rather than gradual changes (Wehyenmeyer et al. 2012). Such a natural borderline, which is known for plants as the limes norrlandicus, might affect dispersal in a distance-independent way. Also, environmental variability within these lake ecosystems might covary with latitude, with higher latitude lakes being exposed to long winters and short growing seasons. The temporal co-occurrence networks of lake phytoplankton were comparably small and poorly connected but featured a high long-term stability. This might indicate that temporal turnover of lake phytoplankton, and the influence of disturbances in patchy environments are relatively low. In conclusion, factors such as patchiness, dispersal limitation, environmental gradients, and regime shifts complicate the prediction of the influence of dispersal and fluctuating environmental conditions for pelagic freshwater ecosystems. However, co-occurrence networks might be a useful tool to further explore metacommunity dynamics in such systems.

Factors to consider when performing comparative network analyses

Our results highlight the difficulty in reliably distinguishing the specific mechanisms underlying the patterns that network analysis identifies. The differences in the topology of

the networks can be attributed to a large number of potential factors, which include species sorting, disturbance, sampling efforts, generation times of the organisms, and the role of dominant species. In addition, it can be difficult to characterize the connectivity of systems based on the spatial arrangement of habitat and species' dispersal capabilities because comparable quantitative estimates for dispersal mode and ability do not exist for most of the described taxa on earth.

Our conceptual model does not account for the strength of species sorting (local niche-based processes), a major factor that has become a recognized part of the metacommunity concept (Logue et al. 2011). As species sorting decreases in strength, we might predict a decreased impact of dispersal and connectivity on the stability of the constructed co-occurrence networks. Frequency of disturbance and evolutionary adaption to disturbance is another factor that is unaccounted for in the conceptual model. Streams and kelp forests, in particular, are much more disturbance prone than the ocean and lake systems (Byrnes et al. 2011, Resh et al. 1988); however, patchiness might counteract this vulnerability if only a subset of systems is affected by a perturbation.

Sampling effort may have also impacted the results that we present here. We assume that the data sets have sampled the majority species in each ecosystem and the consistent sampling of these ecosystems through the decades makes this a reasonable assumption. However, the total number of nodes sets the theoretical maximum diameter of a given network. Thus, increased sampling efforts (e.g., in extent, effort and/or frequency; additional trophic levels e.g., microbes; time of day or seasonality) might cause the network topology to change (Butts 2009).

Furthermore, the temporal frequency of sampling, while the same for the data analyzed here, was not always in synchrony with the generation times of all the organisms. Annual sampling is very appropriate for the invertebrates in the stream data sets, which were univoltine, but may have been less appropriate for organisms like phytoplankton and zooplankton, which have short generation times. It is difficult to determine what portion of the variation among case studies is attributable to differences in generation time. Further study is necessary to determine how differences in sampling regime (Blonder et al. 2012) and generation time affect the topology of spatial and temporal co-occurrence networks. This could be accomplished through analyses of replicate ecosystems surveying the same type of organisms. This would enable a deeper look into mechanisms affecting co-occurrence patterns.

Another concern is the role of dominant species in affecting network connections. Recently, Friedman and Alm (2012) reported that in communities with a high degree of dominance, such as in bacterial communities, changes in the relative abundance of dominant species might lead to the detection of spurious correlations, and hence, false co-occurrence networks. Compared with bacterial communities, the communities addressed in this study feature relatively little dominance, and

we used ranked correlations of absolute abundance data to infer co-occurrence patterns. Hence, we expect that spurious correlations due to changes in relative abundance should not affect our analyses. Moreover, the analysis of modeled metacommunities confirmed that co-occurrence networks constructed from correlation matrices outperform other matrices, and hence, are useful tools to investigate network properties (Berry and Widder 2014). Despite these challenges, there is much to be learned from these valuable long-term datasets.

Section 6. Final conclusions

Connectivity and dispersal are an important but poorly understood part of aquatic ecosystem dynamics (Shurin et al. 2009; De Bie et al. 2012; Perkin and Gido 2012). We used ecological networks to describe the associations between species found in a range of aquatic ecosystems. By simultaneously considering both spatial and temporal co-occurrence network characteristics, we were able to distill the major patterns in the spatial and temporal co-occurrence across aquatic systems and compare different habitats. Furthermore, we were able to demonstrate the types of inference that can be drawn from this analysis and believe that this could be used to study a variety of ecosystems where more information about species interactions and dynamics of the system is desired. Network analysis could be especially important in the context of global climate change as another way to measure changes in the properties of biological systems as they experience changes in environmental conditions (Roemmich and McGowan 1995).

This analysis also highlights the importance of collecting long-term ecological time-series data. Long-term ecological datasets are an extremely valuable tool for providing baseline metrics with which to compare future changes, studying and predicting future trends in ecosystems, evaluating ecological processes that occur on decadal scales, and distinguishing anthropogenic activities from underlying natural change (Magurran et al. 2010). We would not have been able to complete the analyses in this chapter without such datasets. Complete understanding of the spatial and temporal dynamics of ecological systems is rarely accomplished; however, it is clear from the growing body of literature on metacommunities that these complex dynamics are key components of ecosystems (Leibold et al. 2004; Logue et al. 2011). Ultimately, better understanding of metacommunity dynamics will help resource managers increase the success of conservation and restoration projects that seek to manage entire systems (Bie et al. 2012; Moser et al. 2001).

Statement of Authorship

The lead author, Christopher Patrick, wrote the code for all of the analyses, wrote text for the introduction/discussion, synthesized the results, and edited the manuscript. Hannes Peter and Chris Patrick designed the conceptual framework of the network analysis, and Hannes Peter developed the specific network construction and measurement approach. Each

author performed analyses on their respective dataset (Kyle Cavanaugh – kelp forest, Hannes Peter – lake phytoplankton, Chris Patrick – stream invertebrates, Talina Konotchick – open ocean mixed plankton), and wrote text describing and interpreting results for their respective study systems. All authors contributed to editing the final manuscript and provided text for the introduction, methods, results, and discussion.

References

Albert, R., H. Jeong, and A. L. Barabasi. 2000. Error and attack tolerance of complex networks. Nature 406:378-382 [doi:10.1038/35019019].

———, and A. L. Barabasi. 2002. Statistical mechanics of complex networks. Rev. Modern Physics 74:47-97 [doi:10.1103/RevModPhys.74.47].

Barberan, A., S. T. Bates, E. O. Casamayor, and N. Fierer. 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J 6:343-351 [doi:10.1038/ismej.2011.119].

Bastian, M., S. Heymann, and M. Jacomy. 2009. Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

Berry, D., and S. Widder. 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Front. Microbiol. 5 [doi:10.3389/fmicb.2014.00219].

Blonder, B., T. W. Wey, A. Dornhaus, R. James, and A. Sih. 2012. Temporal dynamics and network analysis. Methods Ecol. Evol. 3:958-972 [doi:10.1111/j.2041-210X.2012.00236.x].

Butts, C. T. 2009. Revisiting the foundation of network analysis. Science 325:414-416 [doi:10.1126/science.1171022].

Byrnes, J. E., and others. 2011. Climate-driven increases in storm frequency simplify kelp forest food webs. Global Change Biol. 17:2513-2524 [doi:10.1111/j.1365-2486.2011.02409.x].

Cavanaugh, K. C., D. A. Siegel, D. C. Reed, and P. E. Dennison. 2011. Environmental controls of giant-kelp biomass in the Santa Barbara Channel, California. Mar. Ecol. Prog. Ser. 429:1-17 [doi:10.3354/meps09141].

———, D. A. Siegel, P. T. Raimondi, and F. Alberto. 2014. Patch definition in metapopulation analysis: a graph theory approach to solve the mega-patch problem. Ecology 95:316-328 [doi:10.1890/13-0221.1].

Dayton, P. K. 1985. Ecology of kelp communities. Annual Rev. Ecol. System. 16:215-245 [doi:10.1146/annurev.es.16.110185.001243].

De Bie, T. D., and others. 2012. Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. Ecol. Lett. 15:740-747 [doi:10.1111/j.1461-0248.2012.01794.x].

Diestel, R. 2005. Graph theory. 2005. Springer.

Foster, M. S., and D. R. Shiel. 1985. The ecology of giant kelp forests in California: a community profile. U.S. Fish and Wildlife Services Biological Report 85(7.2).

Friedman, J., and E. J. Alm. 2012. Inferring correlational

networks from genomic survey data. PLoS Comp. Biol. 8(9):e1002687 [doi:10.1371/journal.pcbi.1002687].

Fuhrman, J. A., and J. A. Steele. 2008. Community structure of marine bacterioplankton: patterns, networks, and relationships to function. Aquat. Microb. Ecol. 53:69-81 [doi:10.3354/ame01222].

Graham, M. H. 2004. Effects of local deforestation on the diversity and structure of Southern California giant kelp forest food webs. Ecosystems 7:341-357 [doi:10.1007/s10021-003-0245-6].

———, J. A. Vasquez, and A. H. Buschmann. 2007. Global ecology of the giant kelp Macrocystis: from ecotypes to ecosystems. Oceanogr. Mar. Biol. 45:39-88.

Haury, L., J. A. McGowan, and P. H. Wiebe. 1978. Patterns and processes in the time-space scale of plankton distributions, pp. 277-327. *In* J. H. Steele (Ed.), Spatial patterns in plankton communities. Plenum Press [doi:10.1007/978-1-4899-2195-6_12].

Hsieh, C., and others. 2005. A comparison of long-term trends and variability in populations of larvae of exploited and unexploited fishes in the Southern California region: a community approach. Progr. Oceanogr. 67:160-185 [doi:10.1016/j.pocean.2005.05.002].

Ings, T. C., and others. 2009. Ecological networks—beyond food webs. J. Animal Ecol. 78:253-269 [doi:10.1111/j.1365-2656.2008.01460.x].

Junker, B. H., and F. Schreiber. 2008. Correlation networks. *In* Analysis of biological networks. Wiley-Interscience [doi:10.1002/9780470253489].

Kinlan, B., and S. Gaines. 2003. Propagule dispersal in marine and terrestrial environments: A community perspective. Ecology 84:2007-2020 [doi:10.1890/01-0622].

Kramer, D., and others. 1972. Collecting and processing data on fish eggs and larvae in the California Current region. NOAA Technical report NMFS CIRC-370.

Kristiansen, J. 1996. Dispersal of freshwater algae — a review. Hydrobiologia 336:151-157 [doi:10.1007/BF00010829].

Kuchaiev, O., A. Stevanović, W. Hayes, and N. Pržulj. 2011. GraphCrunch 2: Software tool for network modeling, alignment and clustering. BMC Bioinformatics 12:24 [doi:10.1186/1471-2105-12-24].

Leibold, M. A., and G. M. Mikkelson. 2002. Coherence, species turnover, and boundary clumping: elements of metacommunity structure. Oikos 97(2):237-250 [doi:10.1034/j.1600-0706.2002.970210.x].

———, and others. 2004. The metacommunity concept: a framework for multi-scale community ecology. Ecol. Lett. 7:601-613 [doi:10.1111/j.1461-0248.2004.00608.x].

Logue, J. B., N. Mouquet, N. H. Peter, H. Hillebrand, and T. M. W. Group. 2011. Empirical approaches to metacommunities: a review and comparison with theory. Trends Ecol. Evol. 26:482-491 [doi:10.1016/j.tree.2011.04.009].

Lozupone, C.A. & R. Knight. 2007. Global patterns in bacterial diversity. PNAS. 104(27):11436-11440 [doi:10.1073/

pnas.0611525104].

Magurran, A. E., and others. 2010. Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. Trends Ecol. Evol. 25(10):574-582 [doi:10.1016/j.tree.2010.06.016].

Montoya, J. M., and R. V. Sole. 2002. Small world patterns in food webs. J. Theor. Biol. 214:405-412 [doi:10.1006/jtbi.2001.2460].

———, S. L. Pimm, and R.V. Sole. 2006. Ecological networks and their fragility. Nature 442:259-264 [doi:10.1038/nature04927].

Moser, H. G., and others. 2000. Abundance and distribution of rockfish (Sebastes) larvae in the Southern California Bight in relation to environmental conditions and fishery exploitation. CalCOFI Reports 41:132-147.

———, and others. 2001. The CalCOFI ichthyoplankton time series: potential contributions to the management of rocky-shore fishes. CalCOFI Reports 42:112-128.

Newman, M. E. J. 2003. The structure and function of complex networks. SIAM Rev 45:167-167-256.

Ohman, M. D., and E. L. Venrick. 2003. CalCOFI in a changing ocean. Oceanography 16:76-85 [doi:10.5670/oceanog.2003.34].

Olesen, J. M., J. Bascopmte, Y. Dupont, H. Elberling, C. Rasmussen, & P. Jordano. 2011. Missing and forbidden links in mutualistic networks. Proceedings of the Royal Society of London Series B. 278: 725-732 [doi:10.1098/rspb.2010.1371]

Östman, Ö., and others. 2010. Regional invariance among microbial communities. Ecol. Lett. 13:118-127 [doi:10.1111/j.1461-0248.2009.01413.x].

Patrick, C. J., and C. M. Swan. 2011. Reconstructing the assembly of a stream-insect metacommunity. J. N. Am. Benthol. Soc. 30(1):259-272 [doi:10.1899/09-169.1].

Perkin, J. S., and K. B. Gido. 2012. Fragmentation alters stream fish community structure in dendritic ecological networks. Ecol. Appl. 22:2176-2187 [doi:10.1890/12-0318.1].

Poff, N. L., and J. V. Ward. 1989. Implications of streamflow variability and predictability for lotic community structure: a regional analysis of streamflow patterns. Can. J. Fish. Aquat. Sci. 46:1805-1818 [doi:10.1139/f89-228].

R Development Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Resh, V. H., and others. 1988. The role of disturbance in stream ecology. J. N. Am. Benthol. Soc. 7:433-455 [doi:10.2307/1467300].

Roemmich, D., and J. McGowan. 1995. Climatic warming and the decline of zooplankton in the California Current. Science 267(5202):1324-1326 [doi:10.1126/science.267.5202.1324].

Ruan, Q. S., and others. 2006. Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. Bioinformatics 22:2532-2538 [doi:10.1093/bioinformatics/btl417].

Rusak, J. A., N. D. Yan, and K. M. Somers. 2008. Regional climatic drivers of synchronous zooplankton dynamics in north-temperate lakes. Can. J. Fish. Aquat. Sci. 65:878-889 [doi:10.1139/f08-043].

Saito, R., and M. E. Smoot. 2012. A travel guide to Cytoscape plugins. Nat. Methods 9(11):1069-1076 [doi:10.1038/nmeth.2212].

Shade, A., C. Y. Chiu, and K. D. McMahon. 2010. Differential bacterial dynamics promote emergent community robustness to lake mixing: an epilimnion to hypolimnion transplant experiment. Environ. Microbiol. 12:455-466 [doi:10.1111/j.1462-2920.2009.02087.x].

Shurin, J. B., K. Cottenie, and H. Hillebrand. 2009. Spatial autocorrelation and dispersal limitation in freshwater organisms. Oecologia 159:151-159 [doi:10.1007/s00442-008-1174-z].

Smith, P. E., and R.W. Eppley. 1982. Primary production and the anchovy population in the Southern California Bight: comparison of time series. Limnol. Oceanogr. 27:1-17 [doi:10.4319/lo.1982.27.1.0001].

———, and S. L. Richardson. 1977. Standard techniques for pelagic fish egg and larva surveys. FAO Fisheries Technical Paper. 175.

Sole, R. V., and J. M. Montoya. 2001. Complexity and fragility in ecological networks. Proc. R. Soc. Biol. Sci. B 268:2039-2045 [doi:10.1098/rspb.2001.1767].

Steele, J. A., and others. 2011. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. ISME J. 5:1414-1425 [doi:10.1038/ismej.2011.24].

Swan, C. M., and B. L. Brown. 2011. Advancing theory of community assembly in spatially structured environments: local vs. regional processes in river networks. J. N. Am. Benthol. Soc. 30(1): 232-234 [doi:10.1899/10-150.1].

Walker, H. J., W. Watson, and A. M. Barnett. 1987. Seasonal occurrence of larval fishes in the nearshore Southern California Bight off San Onofre, California. Estuar. Coast. Shelf Sci. 25(1):91-109 [doi:10.1016/0272-7714(87)90028-X].

Watson, J. R., and others. 2010. Realized and potential larval connectivity in the Southern California Bight. Mar. Ecol. Prog. Ser. 401:31-48 [doi:10.3354/meps08376].

———, and others. 2011. Identifying critical regions in small-world marine metapopulations. Proc. Natl. Acad. Sci. U.S.A. 108:E907–E913 [doi:10.1073/pnas.1111461108].

Wehyenmeyer, G. A., H. Peter, and E. Willén. 2012. Shifts in phytoplankton species richness and biomass along a latitudinal gradient—consequences for relationships between biodiversity and ecosystem functioning. Freshw. Biol. 58:612-623 [doi:10.1111/j.1365-2427.2012.02779.x].

Willén, E. 2007. Phytoplankton in lakes. Environmental Quality Criteria. Department of Environmental Assessment.