

ORIGINAL RESEARCH

The genetic network of greater sage-grouse: Range-wide identification of keystone hubs of connectivity

Todd B. Cross^{1,2} | Michael K. Schwartz¹ | David E. Naugle² | Brad C. Fedy³ |
 Jeffrey R. Row³ | Sara J. Oyler-McCance⁴

¹USDA Forest Service, National Genomics Center for Wildlife and Fish Conservation, Rocky Mountain Research Station, Missoula, Montana

²College of Forestry and Conservation, University of Montana, Missoula, Montana

³School of Environment, Resources and Sustainability, University of Waterloo, Waterloo, ON, Canada

⁴U.S. Geological Survey Fort Collins Science Center, Fort Collins, Colorado

Correspondence

Todd B. Cross, USDA Forest Service, National Genomics Center for Wildlife and Fish Conservation, Rocky Mountain Research Station, Missoula, MT.
 Email: tbcross@fs.fed.us

Funding information

This study was supported by grants from the Montana and Dakotas Bureau of Land Management (07-IA-11221643-343, 10-IA-11221635-027, and 14-IA-11221635-059), the Great Northern Landscape Conservation Cooperative (12-IA-11221635-132), and the Natural Resources Conservation Service—Sage-grouse Initiative (13-IA-11221635-054). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. The views in this article are those of the authors and of the U.S. Geological Survey; however, these views do not necessarily reflect those of other employers.

Abstract

Genetic networks can characterize complex genetic relationships among groups of individuals, which can be used to rank nodes most important to the overall connectivity of the system. Ranking allows scarce resources to be guided toward nodes integral to connectivity. The greater sage-grouse (*Centrocercus urophasianus*) is a species of conservation concern that breeds on spatially discrete leks that must remain connected by genetic exchange for population persistence. We genotyped 5,950 individuals from 1,200 greater sage-grouse leks distributed across the entire species' geographic range. We found a small-world network composed of 458 nodes connected by 14,481 edges. This network was composed of hubs—that is, nodes facilitating gene flow across the network—and spokes—that is, nodes where connectivity is served by hubs. It is within these hubs that the greatest genetic diversity was housed. Using indices of network centrality, we identified hub nodes of greatest conservation importance. We also identified keystone nodes with elevated centrality despite low local population size. Hub and keystone nodes were found across the entire species' contiguous range, although nodes with elevated importance to network-wide connectivity were found more central: especially in northeastern, central, and southwestern Wyoming and eastern Idaho. Nodes among which genes are most readily exchanged were mostly located in Montana and northern Wyoming, as well as Utah and eastern Nevada. The loss of hub or keystone nodes could lead to the disintegration of the network into smaller, isolated subnetworks. Protecting both hub nodes and keystone nodes will conserve genetic diversity and should maintain network connections to ensure a resilient and viable population over time. Our analysis shows that network models can be used to model gene flow, offering insights into its pattern and process, with application to prioritizing landscapes for conservation.

KEYWORDS

Centrocercus urophasianus, graph theory, multiscale conservation prioritization

1 | INTRODUCTION

Understanding population structure and quantifying genetic connectivity are important for guiding ongoing conservation and restoration efforts (Crooks & Sanjayan, 2006). Traditionally, population structure is first analyzed and subpopulations delineated, then genetic connectivity among subpopulations is quantified. However, this process need not be completed in two stages. Genetic network models can be used to simultaneously gain an understanding of population structure and to quantify genetic connectivity among populations in natural systems (Bunn, Urban, & Keitt, 2000; Dyer, 2007; Dyer & Nason, 2004).

Genetic networks are constructed of components called nodes and edges, where nodes may represent populations and edges represent genetic connectivity among nodes (Sallaberry, Zaidi, & Melançon, 2013). Each node can be weighted by the genetic diversity within the nodes and each edge by the genetic covariance among local populations (Bunn et al., 2000). The overall structure of the network provides a means by which to rank the importance of how each component contributes to maintaining network connectivity (Jacoby & Freeman, 2016). One can think of network structure in terms of the commercial airline model. In such models, nodes and edges are known, as are nodes of high and low connectivity (henceforth, hub nodes and spoke nodes, respectively). In the airline industry, hub nodes are strategically selected to maximize efficiency of air traffic, while spoke nodes are selected based on limited need for services. For wildlife populations, where populations serve as hub nodes and where populations serve as spoke nodes are unknown.

Qualifying genetic network structure and identifying nodes that act as hubs can be very informative to conservation and management of wildlife species (Garroway, Bowman, Carr, & Wilson, 2008; Lookingbill, Gardner, Ferrari, & Keller, 2010). Knowledge of which nodes are connected to one another and which nodes rank highly in network centrality can facilitate prioritization for management (Jacoby & Freeman, 2016). Prior network modeling of wildlife populations have shown that which nodes function as hubs and which function as spokes is not intuitive (Bunn et al., 2000; Garroway, Bowman, & Wilson, 2011; Garroway et al., 2008; Koen, Bowman, & Wilson, 2015). One might expect a node's proximity to the center of the species' range would influence that node's importance to connectivity, where centrally located nodes have greater genetic exchange than peripheral nodes. However, it has been shown that populations at the periphery of a species' range can act as critical hub nodes, connecting populations across the network, and that populations located toward the center of the range do not necessarily function as hub nodes (Bunn et al., 2000).

Emergent properties of genetic networks can be used to identify hub nodes and spoke nodes and the sensitivity of the entire network to the loss of connectivity (Dunne, Williams, & Martinez, 2002). There are three common network structures: (1) single-scale ("regular"), (2) broad-scale ("random"), and (3) small-world—a subset of which are known as scale-free (Amaral, Scala, Barthélémy, & Stanley, 2000; Bray, 2003). Regular networks are highly structured

such that proximal nodes tend to be linked to each other, while distant nodes tend not to be linked: a structure comparable to the isolation by distance pattern commonly discovered in the population genetics literature (e.g., a stepping-stone model; Wright, 1943). In a regular genetic network, genetic connectivity is between neighboring nodes and nodes separated by a greater number of edges will be more isolated from one another. In regular networks, hub nodes are nonexistent as all nodes are equally connected. Random networks are unstructured such that proximity of nodes is irrelevant to whether nodes are connected or not and to the strength of connections: a structure most similar to the theoretical island model first proposed by Wright (1931) and analogous to the population genetic concept of panmixia. In a random genetic network, genetic connectivity is unencumbered across the entire network because the number of steps between any two nodes is relatively small such that close and distant nodes have equal chances of being linked. In random networks, there are no hub nodes, but there exist thoroughfares through the network that foster quick transit among any set of nodes. In contrast, small-world networks are composed of few highly connected nodes (hub nodes) and a greater number of more isolated nodes (spoke nodes), much like the hub-and-spoke model characteristic of the familiar commercial airline model. Most nodes can be reached from every other node by a small number of steps, often routed through central hub nodes, which foster connectivity among the spoke nodes. Redundancy is an important characteristic of small-world networks. In small-world genetic networks, genetic connectivity is greatest among nearest neighbor nodes, but genetic connectivity can exist between any two nodes by a small number of steps through hub nodes which are nodes at which genetic connectivity is concentrated such that these nodes serve to connect other distal nodes (also known as, spoke nodes). An extreme form of small-world networks is scale-free networks. In scale-free networks, there is less redundancy in internode connections and greater centrality for the hub nodes.

Within any network's structure, individual node importance to network connectivity can be quantified by centrality indices. There are several centrality indices, each of which quantifies the importance of a node to network connectivity in a different way (Table 1). Some centrality indices rank nodes based on local connectivity and some based on network-wide connectivity. Therefore, these function-valued centrality indices can be easily transformed into node-specific rankings, and these rankings can be used to prioritize conservation (Jacoby & Freeman, 2016).

The greater sage-grouse (*Centrocercus urophasianus*; hereafter, sage-grouse; Figure 1) is a lekking gallinaceous bird of conservation concern, an indicator species for sagebrush (*Artemisia* spp.) communities (Rowland, Wisdom, Suring, & Meinke, 2006), and an indicator species for landscape-scale connectivity across the western United States and southern Canada (Aldridge et al., 2008). Sage-grouse once occupied over 1.2 million km² (Edminster, 1954; Schroeder et al., 2004). The species now occupies less than 0.67 million km² across 11 western states and two Canadian provinces (Patterson, 1952; Schroeder et al., 2004)—56% of its range compared to pre-Western

TABLE 1 Network parameters used to quantify connectivity, the unit for which each is calculated, and the definition of the parameter, and relation of the parameter as pertains to the greater sage-grouse population network. All but characteristic path length and weight are centrality indices

Network parameter	Network unit	Definition source	Ecological interpretation
Characteristic path length	Entire network	The mean of all pairwise network distances connecting nodes ^b	Mean number of steps for genetic exchange among all nodes along all possible paths
Betweenness centrality	Node	The number of shortest paths upon which a particular node lies ^b	The importance of node to maintaining network-wide genetic exchange along the most direct routes
Closeness centrality	Node	The mean shortest path between node and all other nodes (connected network)	Mean number of steps for the most direct path of genetic exchange between any two nodes
Clustering coefficient	Node	The probability that two nodes connected to a given node are also connected (ranges from 0–1) ^b	An index of genetic connectivity among nodes that are both connected to another node
Degree centrality	Node	The number of edges connected to a node ^b	The number of other nodes with which a given node exchanges genes
Eigenvector centrality	Node	The direct and indirect connectivity: per node and immediate neighbors ^a	An index of how well connected a given nodes' connections are as follows: that is, how much genetic exchange occurs at a node's immediate connections
Strength	Node	The sum of all edge weights ^b	An index of the magnitude of genetic exchange with all nodes connected to a given node
Weight	Edge	The magnitude of covariance between connected nodes	The magnitude of genetic exchange between any two connected nodes

^aGarroway et al. (2008).

^bNewman (2006).



FIGURE 1 A male greater sage-grouse (*Centrocercus urophasianus*) displays on a lek in the early morning. Photograph credit: Rick McEwan

Settlement (Schroeder et al., 2004). An additional 29% of the remaining species' range is likely at risk of extirpation (Aldridge et al., 2008). Increased geographic isolation and declines in sage-grouse populations range-wide coincide with fragmentation and loss of sagebrush due to changes in land use (Copeland, Doherty, Naugle, Pocewicz, & Kiesecker, 2009; Knick et al., 2003; Schrag, Konrad, Miller, Walker, & Forrest, 2011).

Sagebrush provides essential cover, is a staple of the species' diet, is where the bird nests and rears its broods, and congregates in the spring to display and breed on leks (Beever & Aldridge, 2011;

Hagen, Connelly, & Schroeder, 2007; Patterson, 1952; Remington & Braun, 1985; Wallestad & Eng, 1975). On leks, males battle with one another to claim the center and energetically display to potential mates. Lek attendance by males is significantly correlated with female lek attendance (Bradbury, Vehrencamp, & Gibson, 1989).

Despite long seasonal migratory movements (up to 240 km; Smith, 2012) and large home ranges (4–195 km²; Connelly, Hagen, & Schroeder, 2011; Connelly, Rinkes, & Braun, 2011), fidelity to leks and stability in lek location are well documented (Cross, Naugle, Carlson, & Schwartz, 2017; Dalke, Pyrah, Stanton, Crawford, & Schlatterer, 1963; Dunn & Braun, 1985; Emmons & Braun, 1984; Patterson, 1952; Schroeder & Robb, 2003; Wallestad & Schladweiler, 1974). However, sage-grouse may shift or abandon leks because of persistent disturbance or alteration of sagebrush cover (Holloran, Kaiser, & Hubert, 2010; Walker, Naugle, & Doherty, 2007).

Sage-grouse are known to disperse during the breeding season and are capable of long-distance breeding dispersal movements (Cross et al., 2017). While the fashion of long-distance dispersal movements is unknown, most migratory movements are made in stepping-stone fashion (Tack, 2009), and short-distance abrupt singular movements are common when suitable habitat is lacking (Dunn & Braun, 1986).

The lek mating system of sage-grouse is well suited to network analyses because leks are fairly fixed spatial locations. Given the species' patterns of dispersal, we would expect that network structure should be composed of clustered, hub node-like nodes characteristic of a small-world network (Garroway et al., 2008).

TABLE 2 Sample summary listed by U.S. state or Canadian province for all samples used to construct range-wide greater sage-grouse genetic network. Total individuals sampled per state/province, leks sampled per state/province, and total number of nodes per state/province

State/Province	Individuals	Leks sampled	Nodes
CA	53	14	6
CAN (SASK)	6	2	1
CO	679	106	38
ID	988	281	80
MT	1881	358	130
ND	7	2	1
NV	430	116	45
OR	296	52	31
SD	75	15	6
UT	607	114	44
WY	902	120	76

New connectivity insights would come at a critical time for sage-grouse. In 2010, sage-grouse were added to the federal Endangered Species Act (ESA) candidate list following several petitions for protection (U.S. Fish and Wildlife Service 2010). In September 2015, a U.S. Fish and Wildlife Service determination found current efforts by state and federal agencies and other partners adequate to obviate the need for listing. However, significant conservation challenges remain, and the species' status will again be reviewed in 2020 (U.S. Fish and Wildlife Service 2015). Understanding genetic connectivity is a critical step toward comprehending the relationship between the distribution and abundance of extant populations across fragmented landscapes. Perhaps more importantly, network analysis will greatly benefit planning in identifying conservation targets with the greatest benefit to maintaining genetic connectivity.

In this study, we had two primary objectives. First, we sought to determine the network structure of connectivity among leks, weighting edges by genetic divergence based on genetic covariance among leks. Second, we sought to identify which leks were important to maintaining overall population connectivity and persistence using network centrality indices. Within this second objective, we also sought to identify keystone nodes, that is, nodes that act as more important to maintaining gene flow than their size or location within the species range alone would indicate.

2 | METHODS

2.1 | Study area and sampling

We used 16,420 spatially referenced sage-grouse feather and blood samples collected from 2,139 leks (mean of 7.68 samples per lek) across the entire contiguous range of the species in the United States of America and Canada from 2005 to 2015. The spatial distribution of our sampling was optimized as described in Hanks et al.

(2016), following a smaller pilot sample. Feather samples were collected from leks using noninvasive methods (Bush, Vinsky, Aldridge, & Paszkowski, 2005; Segelbacher, 2002) after having been dropped by sage-grouse during breeding activity, while blood samples were collected from sage-grouse on leks as part of radiotelemetry field research. The only location throughout the entire distribution of the species that we did not use was Washington State because samples from this location were collected during a different period (from 1992 to 1999) than the rest of the samples.

2.2 | DNA extraction

Genetic analysis was conducted at two molecular biological laboratories: the National Genomics Center for Wildlife and Fish Conservation at the USFS Rocky Mountain Research Station and the Molecular Ecology Lab at the USGS Fort Collins Science Center. Protocols were established at the inception of the study to ensure consistency among laboratory genotyping and are described below. Feather DNA was extracted from the quill (calamus) using QIAGEN DNeasy Blood and Tissue Kit and the user-developed protocol for purification of total DNA from nails, hair, or feathers. The protocol was modified by incubating samples for a minimum of 8 h after addition of Proteinase K to maximize tissue lysis and by eluting DNA with 100 μ l of Buffer AE to increase the final DNA concentration in the eluate. Blood samples were extracted using QIAGEN DNeasy Blood and Tissue Kit and the protocol for nucleated blood. At USGS, parts of the extraction process were automated using a QIAcube (Qiagen).

2.3 | Microsatellite DNA amplification and genotyping

We based our analysis upon a panel of neutral, polymorphic microsatellite loci both to identify individuals from noninvasively collected samples (of unknown individual origin) and to quantify relatedness (i.e., functional movement resulting in gene flow). We amplified 15 variable microsatellite loci (BG6, SGMS064, SGMS066, SGMS068, MSP11, MSP18, SG28, SG29, SG36, SG39, SGCA5, SGCA11, SGCTAT1, TUT3, and TUT4) and one sex-diagnostic locus [CHD gene, using the primers 1237L and 1272H (Kahn, St. John, & Quinn, 1998)]. Primer design, PCR conditions, and electrophoresis used at USFS and USGS are detailed in Cross, Naugle, Carlson, and Schwartz (2016); Cross et al. (2017) and Row et al. (2015).

To ensure correct genotyping from low-quality and low-quantity feather DNA samples, each sample was PCR-amplified twice across the 15 variable microsatellite loci to screen for allele dropout, stutter artifacts, and false alleles (DeWoody, Nason, & Hipkins, 2006). To minimize genotyping error, at least two independent observers scored each sample. If any locus failed to amplify in either replicate or if there was a discrepancy between locus genotypes as scored by the two observers, PCR amplification and genotyping were repeated twice more. If a genotype was confirmed by this repeat analysis, then it was retained. If a genotype failed again, the sample was

assigned a missing score at the failed locus. To ensure consistency among laboratories, both laboratories genotyped the same 70 individuals (approximately 1% of the retained genotypes). Each laboratory's genotypes for these individuals were compared, and necessary shifts to synchronize allele calls were made for all samples.

To screen samples for quality control, we removed from analysis any individual for which amplification failed at one-third of the loci (i.e., five loci). After removal of poor-quality samples, genotypes were screened to ensure consistency between allele length and length of the microsatellite repeat motif using MICROCHECKER v2.2.3 (Van Oosterhout, Hutchinson, Wills, & Shipley, 2004). To identify and remove multiple captures of the same individual and to screen for and correct genotyping error, we used DROPOUT 2.3 (McKelvey & Schwartz, 2005 as implemented in Schwartz et al. 2006), MICROCHECKER v2.2.3 (Van Oosterhout et al., 2004), and package ALLELEMATCH 2.5 (Galpern, Manseau, Hettinga, Smith, & Wilson, 2012) in program R 3.3.0 (R Core Team 2016). Finally, we quantified the power of our microsatellite locus panel to discern individuals using probability identity (P_{ID} ; Evett & Weir, 1998) which calculates the probability that two individuals drawn at random from the population have the same genotype across all loci.

2.4 | Network construction

A minimum of four or more individuals per node is required to calculate within-node genetic variation (Dyer, 2014b). Therefore, before constructing the network, we performed a hierarchical clustering analysis of lek locations. First, we calculated a distance-based tree using the geographic coordinates for the leks from which the 6,242 individual samples were collected (using the HCLUST function in base R). Second, we clustered all lek locations within 15 km of one another (cut distance implemented using the CUTREE function in base R). We selected 15 km as the cut distance, as this is the best estimate of median breeding dispersal distance among leks for sage-grouse (Cross et al., 2017). Third, we removed any resultant clusters of leks composed of fewer than four individuals.

Following clustering, we constructed a weighted population network among the resulting clusters, which we henceforth refer to as nodes. For all clustered samples and for all nodes, we calculated the mean and standard deviation for number of alleles per locus, effective number of alleles, expected and observed heterozygosity, and F_{IS} . We estimated genetic covariance among nodes, where microsatellite genetic covariance represents the weight of each network edge connecting nodes. We used the packages GSTUDIO (Dyer, 2014a) and POPGRAPH (Dyer, 2014b) in program R to estimate the conditional genetic covariance network following the methods of Dyer and Nason (2004) using default parameters ($\alpha = 0.05$ and tolerance = 1×10^{-4} ; Garroway et al., 2008). Following pruning using the recommended settings, the resultant minimal incidence matrix contained the smallest set of edges that sufficiently capture the among-node genetic covariance structure (Dyer & Nason, 2004). We also calculated the minimum spanning tree, which is the subset of network edges that connect all nodes together with the maximum

genetic covariance among nodes (edge weight), without any cycles. To test for structure within the minimum spanning tree, we tested for correlation between weighted (factoring genetic covariance) distances among nodes in the minimum spanning tree and geographic distance (great circle distance) among all nodes calculated using the RDIST.EARTH function in the FIELDS package (Nychka, Furrer, Paige, & Sain, 2015) in R.

2.5 | Network structure determination

To determine the network structure, we compared the degree distribution, clustering coefficient, and characteristic path length of the sage-grouse genetic network to that of 1000 Erdos-Renyi model random networks with the same number of nodes, edges, and edge weight distribution as the range-wide sage-grouse genetic network. The characteristic path length is defined as the average shortest path length between all pairs of nodes in the network, and it provides an understanding of how long it takes alleles to traverse the network. We generated the random networks using package IGRAPH (Csardi & Nepusz, 2006) in program R and tested for significant differences between the degree distribution, clustering coefficient, and characteristic path length of the true sage-grouse network and the random networks using permutation tests, following the methods of Garroway et al. (2008). We used the results of these comparisons to determine whether network structure was purely a function of the number of nodes and edges or whether network structure was a result of nonrandom processes. For example, if we found a characteristic path length that did not deviate significantly from that of the random networks coupled with a significantly higher clustering coefficient than that of the random networks, then we could conclude that the network had small-world or scale-free characteristics (Watts & Strogatz, 1998). Furthermore, if we found a degree distribution that did not follow the power law (which would indicate scale-free network), was not binomial (which would indicate a random network) or fixed (which would indicate a regular network), but instead that was fat-tailed, we would conclude that the likely network structure was that of the hub node-and-spoke small-world network.

We quantified pairwise conditional genetic distance among all nodes. Conditional genetic distance is the length of the shortest path connecting each pair of nodes conditioned on network structure (Dyer, Nason, & Garrick, 2010) or the relative strength of the genetic covariance between nodes along the connecting edges (Koen, Bowman, Garroway, & Wilson, 2013). When compared to geographic distance among nodes, conditional genetic distance can provide insight into network structure. For example, if conditional genetic distance is correlated with geographic distance, one can conclude that the process of isolation by distance shaped a genetic network (Dyer et al., 2010).

We calculated six centrality indices (Table 1) and used these metrics to quantify connectivity and relative isolation of each node across the network. To calculate standard error (SE) of the mean and median as well as their respective 95% confidence intervals (CI), we calculated 1000 resampled networks of 75% of the nodes.

Betweenness centrality quantifies the importance of a node in terms of the bottleneck to gene flow it creates, eigenvector centrality quantifies how connected a node is, and strength quantifies how strong the connection is between a node and all its neighboring nodes (Garroway et al., 2008). Eigenvector centrality is an index of both how well a node is connected and how well a node's immediate connections are connected—in essence, measuring both direct and indirect connectivity. These properties make eigenvector a better index than betweenness if one is interested in quantifying the strength of connections. This is because eigenvector centrality increases not only just with increased immediate connectivity of the node of interest, but also the immediate connectivity of the nodes to which the node of interest is connected. When quantifying connectivity, we used the centrality index of node strength rather than node degree, as Koen et al. (2015) found that strength more adequately depicts migration and gene flow than degree centrality. To examine relationships between network centrality indices, we tested for pairwise correlation between all indices using Spearman's rank correlation. We also tested for correlation between each network centrality index and mean peak male count per node (calculation described below). All network centrality indices were calculated using the IGRAPH package in R, and all correlations were calculated in base R.

To identify hub nodes of genetic exchange, we screened all nodes within the top 1% of each network centrality index (Table 1), and within the top 50% of all node centrality indices combined in order to identify those nodes that were most important to local network (regional) and network-wide (range-wide) connectivity. These nodes represent the top hub nodes of genetic exchange that maintain connectivity at all scales. To identify spoke nodes, we identified the nodes with the lowest ranking for each centrality index.

2.6 | Keystone nodes

We hypothesized that nodes composed of the most highly attended leks and the most geographically central nodes in the species' range would rank highest for centrality (i.e., node abundance and range centrality would be positively correlated with node centrality). To calculate abundance, we used the per-lek high male counts recorded between 2005 and 2015 (WAFWA, 2015) and calculated the mean peak male count per node over these years using all leks constituting each node (male lek attendance is significantly correlated with female lek attendance; Bradbury et al., 1989). Using mean peak male count, we tested for correlation with network centrality indices.

We defined range centrality as the great circle distance from the center of the geographic center of the sampling distribution.

We calculated range centrality as the distance of each node from the centroid of a minimum convex polygon enveloping all nodes, such that increased magnitude of distance was equivalent to inverse range centrality. We calculated the minimum convex polygon using the GCONVEXHULL function and calculated the centroid of the minimum convex polygon using the GCENTROID function in the RGEOS package (Bivand & Rundel, 2017) in R. Finally, we calculated the distance of every node from the centroid using the RDIST.EARTH function in the FIELDS package (Nychka et al., 2015) in R. Using range centrality, we tested for correlation with node centrality indices.

Finally, we sought to identify nodes with greater importance to genetic connectivity than the magnitude of lek attendance within the node or node location within the species range alone might indicate. We call these nodes keystone nodes. We identified keystone nodes as those that were low in attendance or peripheral to the range, but that still ranked high in centrality. To identify keystone nodes, we plotted both the mean peak male count within a node and the range centrality of each node against each network centrality index. We then called the outliers of these plots, keystone nodes.

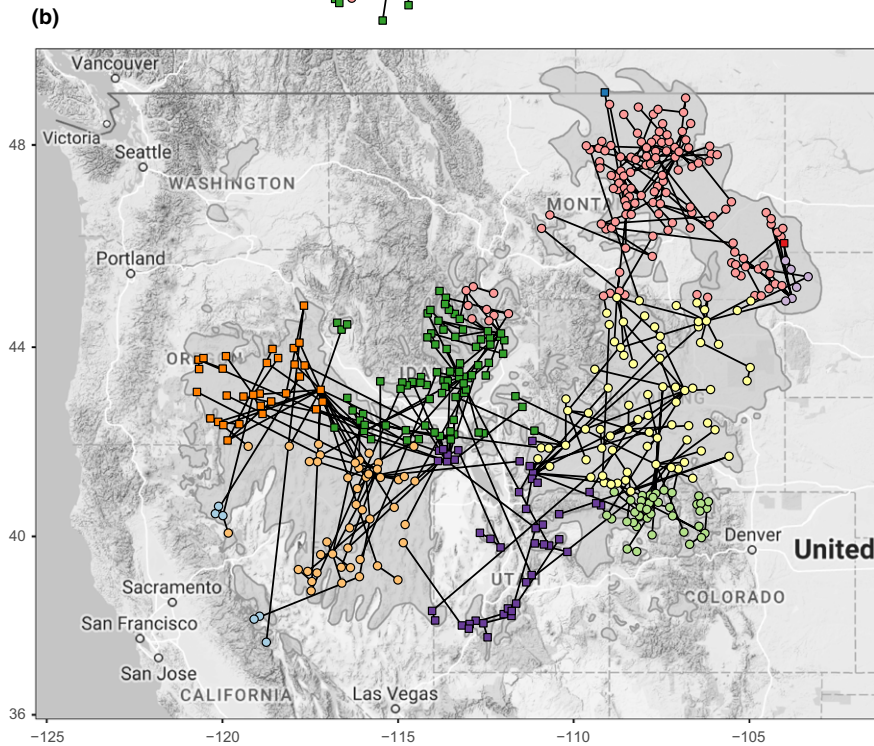
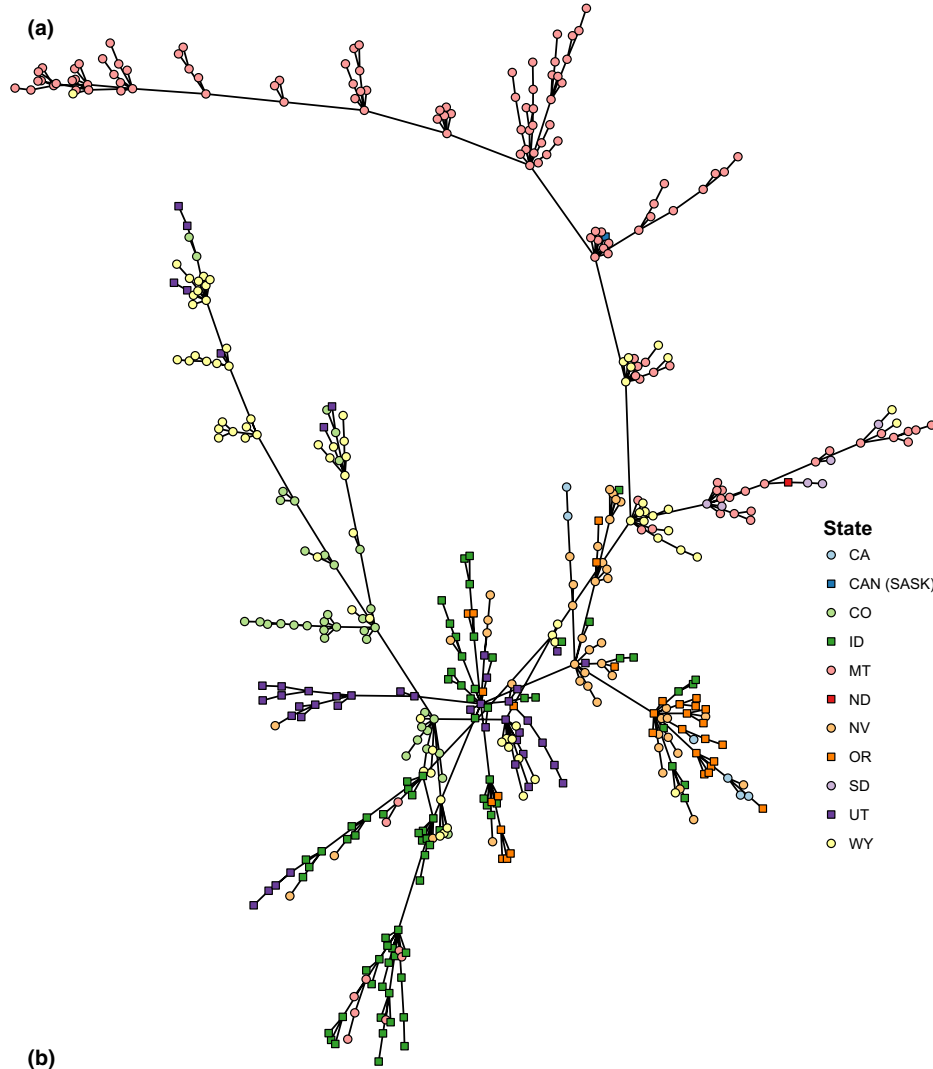
3 | RESULTS

3.1 | Genotyping and network construction

After genotyping the same 70 individuals at each laboratory, we compared allele calls. For two loci, the laboratories had a two base pair difference across all allele calls (BG6 and SGCA5), for one locus, there was a four base pair difference (SGCA11), and for another locus, a seven base pair difference (SGCTAT1) as well as two alleles called off-step on a dimer repeat motif changed to comply with the tetramer repeat motif. Each laboratory's genotypes for these individuals were shifted to synchronize allele calls for all samples for these loci. With the additional ALLELEMATCH analysis, we discovered no additional multiple matches (identical genotypes resulting from the same sample being genotyped by both laboratories). P_{ID} for the complete microsatellite panel was 2.20×10^{-22} , providing evidence that our microsatellite panel was adequate for distinguishing individuals. Our genotyping efforts resulted in 6,729 individual genotypes from 1,388 leks (median = 3 individuals per lek, IQR = 4 individuals per lek, range = 1–62 individuals per lek) following duplicate removal and quality control.

Hierarchical clustering and removal of nodes with fewer than four individuals yielded 5,924 samples from 1,180 leks clustered into 458 nodes from 2006 to 2015 (median = 10 individuals per node, IQR = 7.00–15.75, range = 4–90 individuals per node, Table 2).

FIGURE 2 The greater sage-grouse range-wide genetic network minimum spanning tree. The minimum spanning tree is pruned such that only the most highly weighted edges (i.e., the connections representative of the greatest genetic covariance) are shown between all nodes ($n = 458$). Distance among nodes in the minimum spanning tree was highly correlated with geographic distance between nodes ($r_s = 0.61$, $p < 2.2 \times 10^{-16}$). (a) Fruchterman-Reingold plot (layout with minimal edge overlap). (b) Geographic map of the range-wide greater sage-grouse genetic network nodes connected by edges retained within the minimum spanning tree. Node color indicates geographic location by state. Edges are shown as black lines. The species' range is shown as light gray polygons



We were able to use 1,057 mean peak male counts for 399 nodes. For all clustered samples and across all nodes, average number of alleles was 17.0 ± 7.48 (mean \pm SD) and 6.43 ± 1.47 , for the effective number of alleles was 7.30 ± 2.88 and 4.41 ± 0.75 , for expected heterozygosity was 0.84 ± 0.06 and 0.74 ± 0.05 , for observed heterozygosity was 0.78 ± 0.05 , and for F_{IS} was 0.07 ± 0.02 and -0.06 ± 0.06 . We constructed a network composed of 458 nodes connected by 14,481 edges. The minimum spanning tree (Figure 2a) was correlated with geographic distance ($r_s = 0.61$, $p < 2.2 \times 10^{-16}$). Montana nodes compose a large group within the minimum spanning tree and are joined to the rest of the network through nodes in Wyoming and Idaho (Fig 2). Nodes in Colorado also group together and are linked by nodes in Wyoming and Utah (Figure 2). There was strong evidence for a positive correlation between conditional genetic distance and geographic distance (Table 4).

3.2 | Network structure determination

The sage-grouse genetic network deviated from random network structure in both mean clustering coefficient and mean characteristic path length. Both indices were significantly greater for the sage-grouse genetic network than for the 1000 random networks with the same number of nodes, edges, and edge weight distribution as the sage-grouse network; none of the random networks had a greater mean clustering

coefficient or characteristic path length than the sage-grouse network (clustering coefficient: $p < .001$, characteristic path length: $p < .001$). Both the mean clustering coefficient ($0.19 \pm 3.35 \times 10^{-3}$ [SE]) and the mean characteristic path length ($1.88 \pm 7.04 \times 10^{-3}$ [SE], [1.88, 1.91]) were short, and the node degree distribution was fat-tailed (Figure 4d).

3.3 | Node properties

In order to describe node location, we used USGS hydrologic cataloging units, also known as watersheds (<https://water.usgs.gov/GIS/metadata/usgswrd/XML/huc250k.xml>). We discovered hub nodes within the C.J. Strike Reservoir watershed in Idaho and the Big Horn Lake and Upper Green-Slate watersheds in Wyoming. These nodes rank highly across multiple centrality indices indicating the importance of these regions to maintaining genetic connectivity across the network. Collectively, these two basins contained nodes with the maximum centrality indices (Figure 3). Many other regions contained hub nodes as determined with one or more centrality indices. Notably, the Big Horn Lake watershed in Wyoming, the Bullwhacker-Dog and Middle Musselshell watersheds in Montana, the Fremont watershed in Utah, and the Middle Snake-Succor watershed in Idaho and Oregon contain nodes that rank high for centrality indices indicative of their functioning as hub nodes that maintain local connectivity. The Crazy Woman watershed, the Upper Green-Slate watershed,

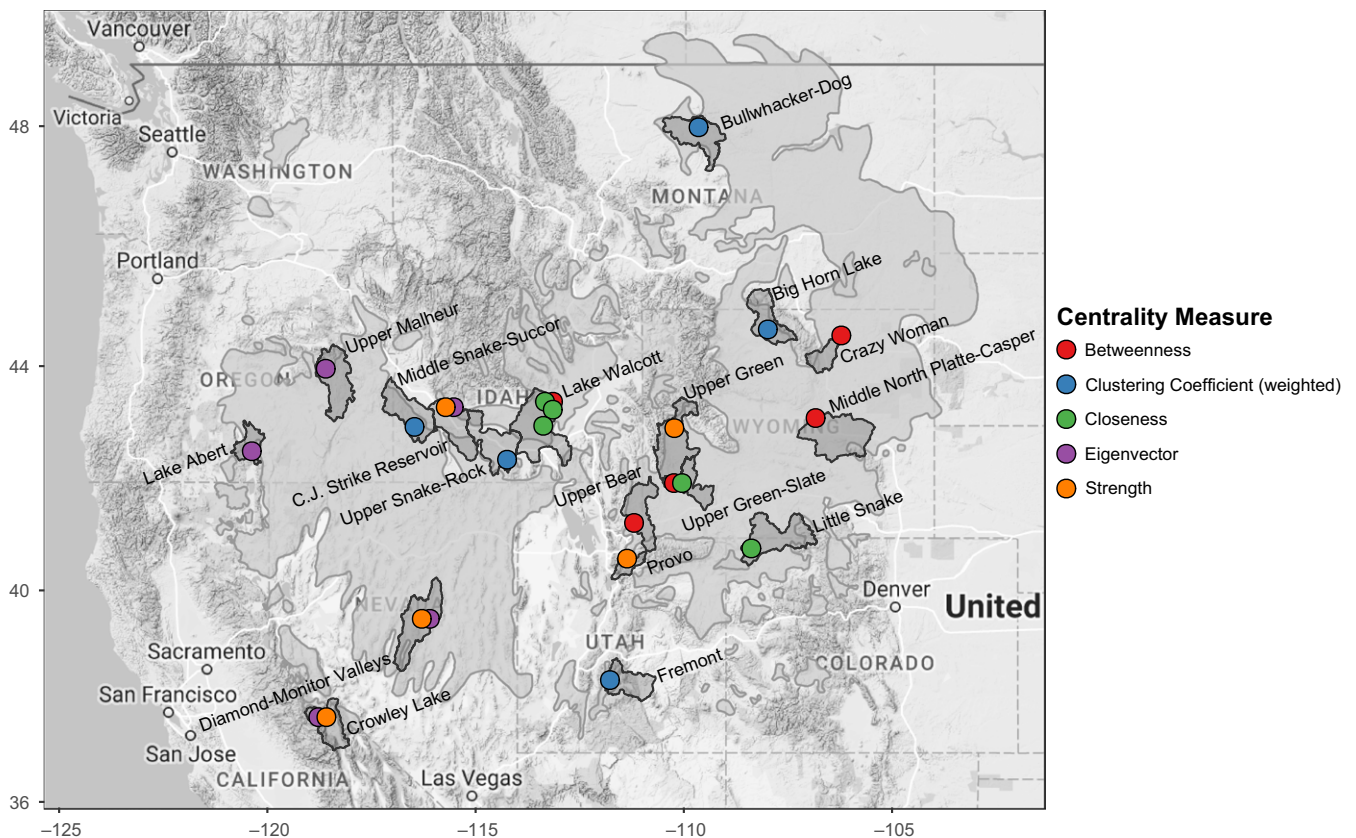


FIGURE 3 The top 1% of nodes for each of the six centrality indices ($n = 20$ per index). Nodes in the top 1% of more than one index are offset to the left or right to reveal both. Node color indicates centrality measure. Shaded polygons depict the watershed within which these top-ranking nodes are located. The species' range is shown as light gray polygons

and the Middle North Platte-Casper watershed in Wyoming, the Lake Walcott watershed in Idaho, and the Upper Bear watershed in Utah all contain nodes that rank high for centrality indices indicative of their functioning as hub nodes that maintain network-wide connectivity (Figure 3).

Only five nodes ranked in the top 50th percentile of all network centrality indices, indicating their importance to genetic connectivity both locally and network-wide as well as the rarity of this combination of local and network-wide importance. The range of each centrality index for these hub nodes was large (Table 3). These nodes were located within the Idaho Falls, Lake Walcott, and Salmon Falls watersheds in Idaho, and the Middle North Platte-Casper and Sweetwater watersheds in Wyoming (Figure 3); locations central to both the sampling extent and the geographic range of the species.

Nodes with high betweenness centrality act as bridges between different parts of the network, so their loss can have network-wide impacts on genetic connectivity (Garroway et al., 2008). We identified several hub nodes whose betweenness ranking was high, located toward the center of the species' range. Three of the top 1% of nodes ranked by betweenness centrality were located in Wyoming—one in the Crazy Woman watershed, one in the Middle North Platte-Casper, and one in the Upper Green-Slate watershed. Of the remaining two nodes, one was located in the Lake Walcott watershed in Idaho, and one was located in the Upper Bear watershed in

Utah (Figure 3). Of these, the node with the greatest betweenness (1491) was located just south of Grand Teton National Park in the Upper Green-Slate watershed. This same node also indexed high for strength (700.89). Most nodes in the network were important to network-wide connectivity (right-skewed distribution: Table 3, Figure 4). However, seventeen nodes had a betweenness of zero, indicating relatively low importance to fostering genetic connectivity across the network. These spoke nodes had an average strength of 429.2 (± 218.3 [SD]), indicating strong connections to other nodes despite low importance to network-wide genetic connectivity. Many of these spoke nodes with the lowest betweenness indices were located toward the periphery of the species' range.

To identify nodes that covaried the greatest with all other nodes in the network, we ranked nodes by closeness centrality. Closeness is an index of the average shortest path between a node and all other nodes in the network. Hence, a smaller closeness index indicates shorter paths on average, and therefore, greater connectivity. There were a small number of very closely covarying nodes in the network (left-skewed distribution: Table 3, Figure 4). The top-ranked closeness nodes were central to the species' range, away from the periphery. Two of the nodes in the top 1% of closeness centrality were located in the Lake Walcott watershed in Idaho. The remaining nodes were located in the Little Snake watershed in Colorado and the Sweetwater and Upper Green-Slate watersheds in Wyoming (Figure 3). The node with the minimum closeness (4.91×10^{-5}) was

TABLE 3 Network centrality indices (betweenness, closeness, clustering coefficient, degree, and eigenvector) and network connectivity (strength and weight) for the range-wide greater sage-grouse genetic network (a) and networks (b) calculated from 1000 networks constructed from a resample of 75% ($n = 343$ nodes) of the originally sampled 458 nodes (sampled without replacement). Listed are the network centrality index, the component for which each index was calculated, minimum, mean, median, standard error (SE) of the mean and median, and 95% confidence intervals (CI) of the mean and median

(a)					
Centrality index	Component	Min	Mean \pm SD	Median (IQR)	Max
Betweenness centrality	Node	0.00	203.60 \pm 249.63	90.50 (29.25–270.50)	1491.00
Closeness centrality	Node	4.91×10^{-5}	$1.34 \times 10^{-4} \pm 1.51 \times 10^{-5}$	1.37×10^{-4} (1.28×10^{-4} – 1.45×10^{-4})	1.59×10^{-4}
Clustering coefficient	Node	0.12	0.19 \pm 0.022	0.18 (0.17–0.20)	0.33
Eigenvector centrality	Node	0.07	0.55 \pm 0.18	0.57 (0.43–0.69)	1.00
Strength	Node	88.19	619.10 \pm 181.51	634.70 (488.60–752.50)	1085.00
Weight	Edge	3.02	9.82 \pm 2.23	9.68 (8.42–11.00)	35.61
(b)					
Centrality index	Component	Resampled networks			
		Mean \pm SE [95% CI]		Median \pm SE [95% CI]	
Betweenness centrality	Node	155.77 \pm 1.28 [153.28, 158.26]		87.00 \pm 4.87 [78.00, 98.00]	
Closeness centrality	Node	$1.75 \times 10^{-4} \pm 1.83 \times 10^{-6}$ [1.71×10^{-4} , 1.78×10^{-4}]		$1.78 \times 10^{-4} \pm 1.87 \times 10^{-6}$ [1.75×10^{-4} , 1.82×10^{-4}]	
Clustering coefficient	Node	0.18 \pm 0.0032 [0.18, 0.19]		0.18 \pm 0.0032 [0.17, 0.19]	
Eigenvector centrality	Node	0.55 \pm 0.025 [0.49, 0.59]		0.56 \pm 0.027 [0.50, 0.60]	
Strength	Node	449.46 \pm 8.82 [431.19, 466.40]		455.71 \pm 10.22 [435.06, 475.58]	
Weight	Edge	9.86 \pm 0.066 [9.74, 9.99]		9.74 \pm 0.069 [9.60, 9.87]	

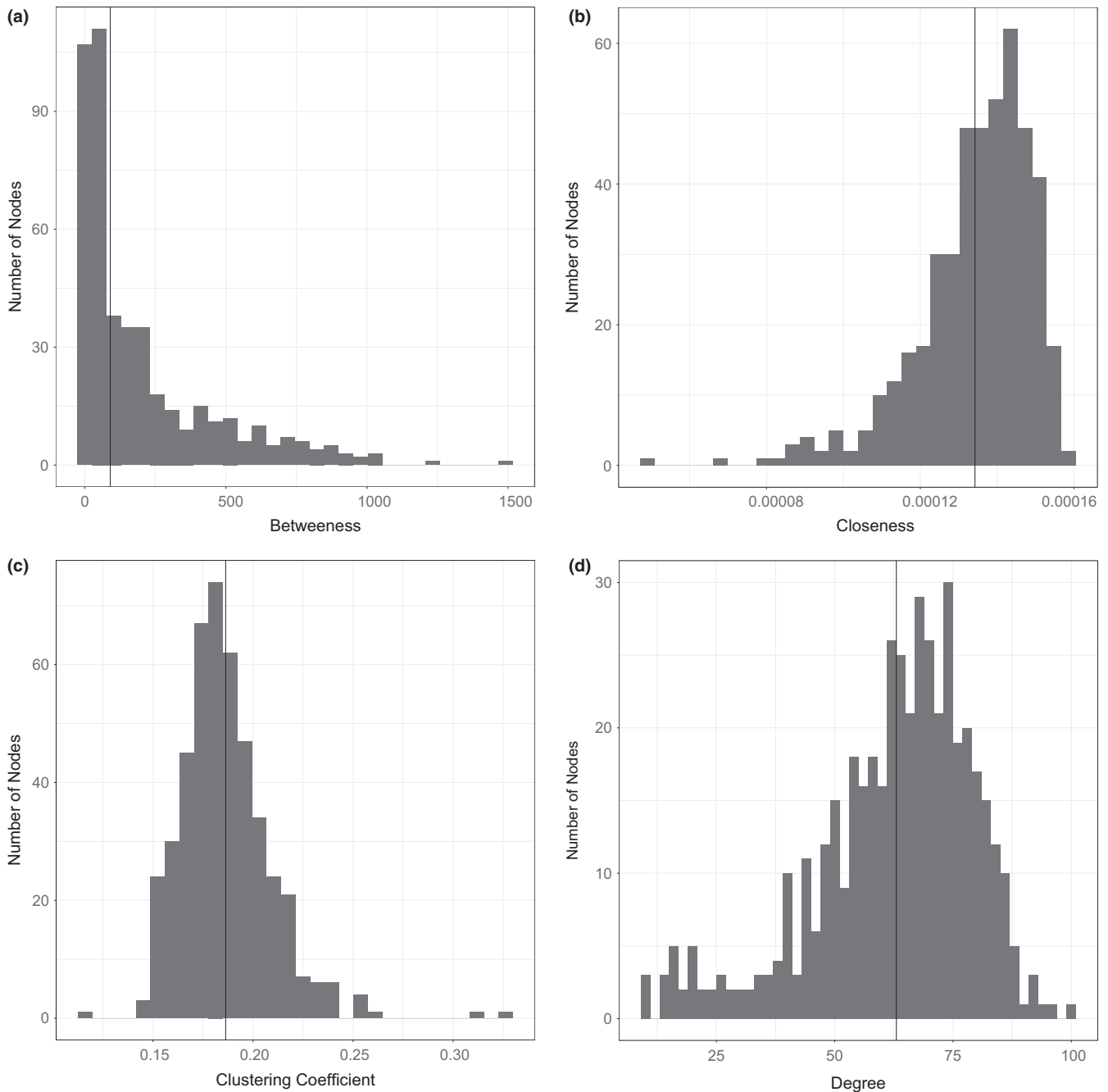


FIGURE 4 Centrality index distributions for all nodes (a–f; $n = 458$) and edges (g; $n = 14,433$) in the greater sage-grouse genetic network. The solid vertical black line shows the median for each index

located in the Spring-Step toe Valleys watershed in Nevada and had low betweenness (0). The node with the greatest closeness in the network (1.59×10^{-4}) was located just south of Grand Teton National Park in the Upper Green-Slate watershed in Wyoming and had a relatively high betweenness (1521) and strength (700.89).

To identify nodes that anchor tightly knit groups connected by a high number of edges, we examined node rankings by clustering coefficient. Increased clustering coefficient is indicative of small-world characteristics. Network-wide, there was a low chance that any two nodes connected to a given node were also connected to

one another (right-skewed distribution: Table 3, Figure 4). The nodes in the top 1% of clustering coefficient were found across the species' range and were mostly toward its periphery (Figure 3). The southernmost hub node in the top 1% was in the Fremont watershed in Utah. Other hub nodes in the top 1% of centrality were located in the Middle Snake-Succor watershed in Idaho, the Big Horn Lake watershed in Wyoming, and the Bullwhacker-Dog watershed and Middle Musselshell watershed in Montana. The hub node with the greatest clustering coefficient (0.33) was found in the Middle Snake-Succor watershed. This node also had low betweenness (0), strength

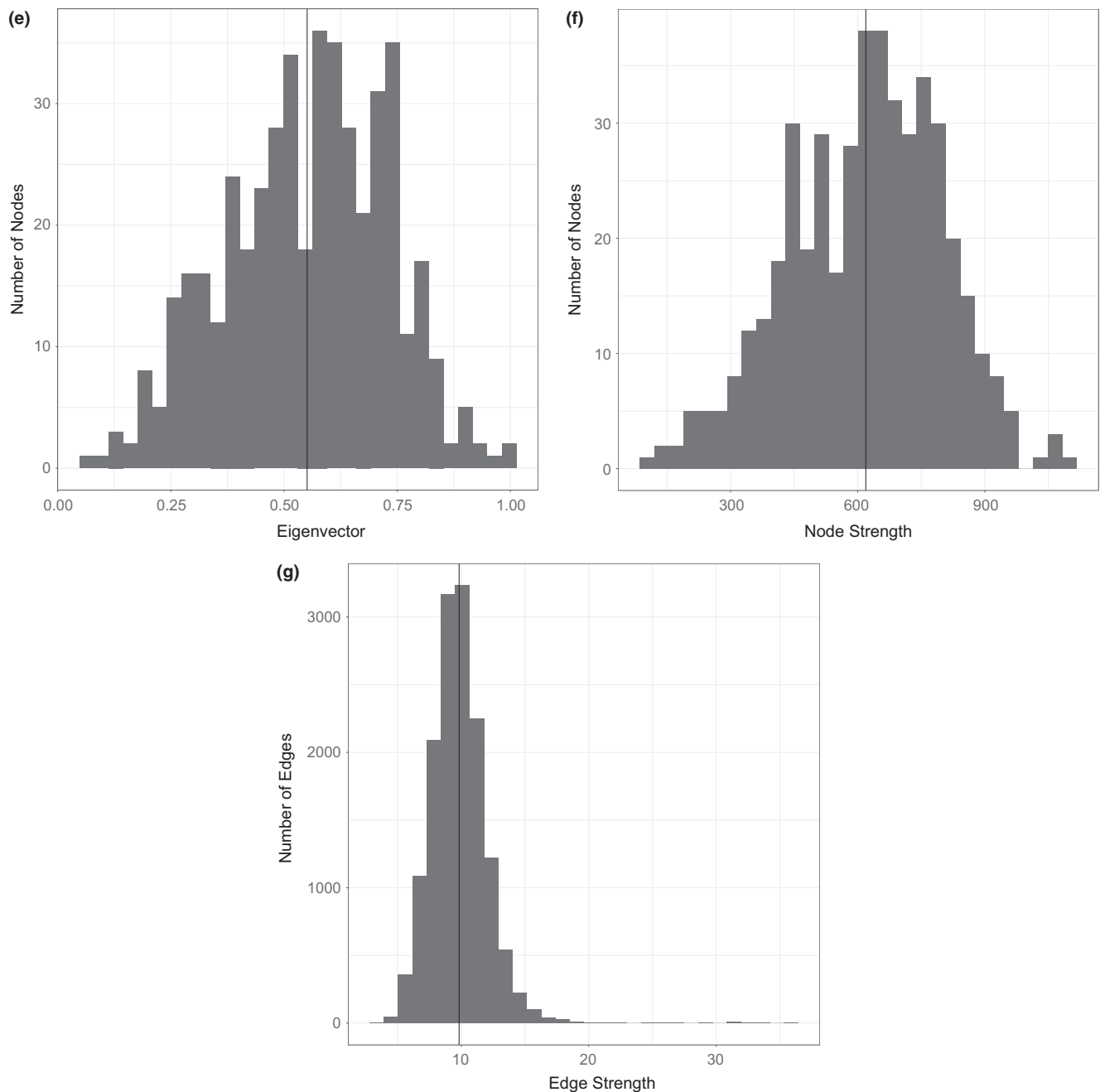


FIGURE 4 (continued)

(223.65), and eigenvector centrality (0.21). The spoke node with the lowest clustering coefficient (0.12) was found in the Shoshone watershed in northcentral Wyoming and had low betweenness (7), strength (327.34), and eigenvector centrality (0.25).

To identify the most highly networked nodes, we examined node rankings by eigenvector centrality. Eigenvector centrality increases for nodes that are highly connected to other highly connected nodes. All but one of the top one percent of nodes ranked by eigenvector centrality were located in the Great Basin, indicating increased genetic connectivity therein. These hub nodes are located in the Crowley Lake watershed in California, the Lake Abert watershed in Oregon, the Diamond-Monitor Valleys watershed in Nevada, and

the Upper Malheur watershed in Oregon (Figure 3). The exception was a single node outside the Great Basin in the C.J. Strike Reservoir watershed in Idaho. This hub node had the greatest eigenvector centrality (1.00) and had very high strength (1064.07), but very low betweenness centrality (3). The spoke node with the lowest eigenvector centrality (0.067) was located in the Escalante Desert watershed in Utah and was a terminal node on the minimum spanning tree. This spoke node was also low for strength (88.2) and betweenness (69): low centrality both locally and network-wide. Eigenvector centrality was normally distributed (Table 3, Figure 4).

To determine which nodes covaried closely with many other nodes, we calculated the strength of each node. The top 1% of

TABLE 4 Correlation between network centrality indices, range centrality, and lek attendance per node for the range-wide greater sage-grouse genetic network. Spearman's rank correlation (r_s) is shown below the diagonal with significance (p) above

	A	He	Betweenness	Closeness	Clustering coefficient	Eigenvector	Strength	Samples in cluster	Overall mean peak male count	Range centrality (inverse)
A	–	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000
He	0.872	–	0.000	0.000	0.000	0.000	0.000	0.000	0.028	0.000
Betweenness	0.706	0.565	–	0.000	0.749	0.000	0.000	0.000	0.028	0.237
Closeness	0.670	0.595	0.914	–	0.380	0.008	0.007	0.000	0.046	0.005
Clustering coefficient	0.341	0.278	–0.015	–0.041	–	0.000	0.000	0.000	0.314	0.429
Eigenvector	–0.515	–0.307	–0.280	–0.124	–0.499	–	0.000	0.000	0.003	0.000
Strength	–0.551	–0.359	–0.281	–0.125	–0.567	0.983	–	0.000	0.006	0.002
Samples in cluster	0.860	0.602	0.665	0.547	0.340	–0.577	–0.613	–	0.021	0.310
Overall mean peak male count	–0.157	–0.110	–0.110	–0.100	–0.051	0.149	0.137	–0.116	–	0.449
Range centrality (inverse)	–0.234	–0.388	–0.055	–0.131	0.037	–0.206	–0.147	–0.048	0.038	–

nodes ranked by strength was located toward the range periphery in both the central and southwestern part of the species' range, indicating increased genetic covariance among a greater number of nodes (Figure 3). Of the hub nodes with the greatest strength, one hub node was in the Crowley Lake watershed in California, one in the Diamond-Monitor Valleys watershed in Nevada, one in the C.J. Reservoir Strike watershed in Idaho, one in the Provo watershed in Utah, and one in the Upper Green watershed in Wyoming. All five of the top 1% of nodes by strength were terminal nodes on the minimum spanning tree. The nodes with the greatest and least strength (1085.0 and 88.2), located in the C.J. Strike Reservoir watershed in Idaho and the Escalante Desert watershed in Utah, were also similarly ranked for eigenvector centrality (two indices for which there was very strong evidence for a positive correlation (Table 4)). Node strength was normally distributed (Table 3, Figure 4).

We found evidence for a strongly positive significant correlation between betweenness and closeness, and eigenvector and strength (Table 4). All other significant correlations among centrality indices were weak and negative. We found evidence for a strongly positive significant correlation between number of alleles per node and betweenness and closeness, although all other significant correlations were moderately negative or weak and positive. The evidence for a correlation between mean peak male count and centrality indices was weak when significant. Finally, we found evidence for a strongly positive significant correlation between the number of samples in a node and betweenness, but only moderate or weak relationships when testing for correlation with other centrality indices.

3.4 | Edge properties

Edge weight is an index of the magnitude of genetic covariance between nodes and can be used to identify nodes most closely linked. Overall, genetic connectivity among nodes has led to increased network connectivity, with the occurrence of some highly connected nodes evidenced by a skewed right distribution of edge weight (Table 3; Figure 4). The top 0.1% of edges with the greatest genetic covariance emanated from a node in the Spring-Steptoe Valleys watershed in Nevada. This node also has the lowest closeness centrality and low eigenvector centrality (0.33) and very low betweenness (0). The edge of least weight connected two nodes within the Fremont watershed in Utah (in the southcentral UT group of nodes in Figure 3). This node was of moderate importance to network-wide connectivity (betweenness: 115), but had low connectivity to other nearby nodes (eigenvector centrality: 0.11).

3.5 | Keystone nodes

It was common that the hub nodes—those with the highest centrality rankings—were also those with lower mean peak male count (Figure 5). There was strong evidence for a weak positive correlation between mean peak male count and eigenvector centrality and mean peak male count and strength and strong evidence for weak

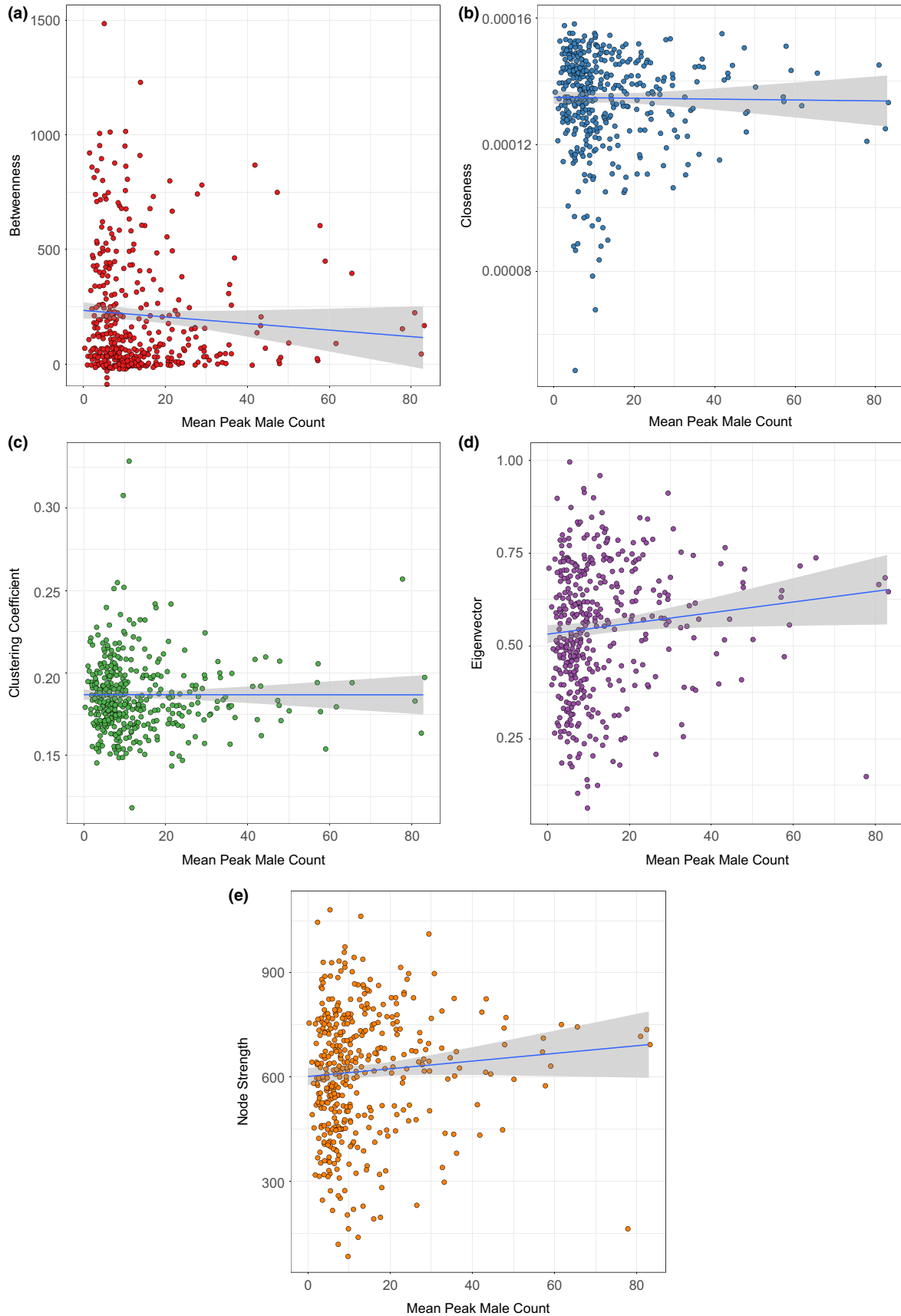


FIGURE 5 Relationships between centrality index (y-axis) and mean peak male count per node (x-axis). Red circles envelope keystone nodes. The fitted linear model and confidence interval are shown (blue line with shaded confidence interval)

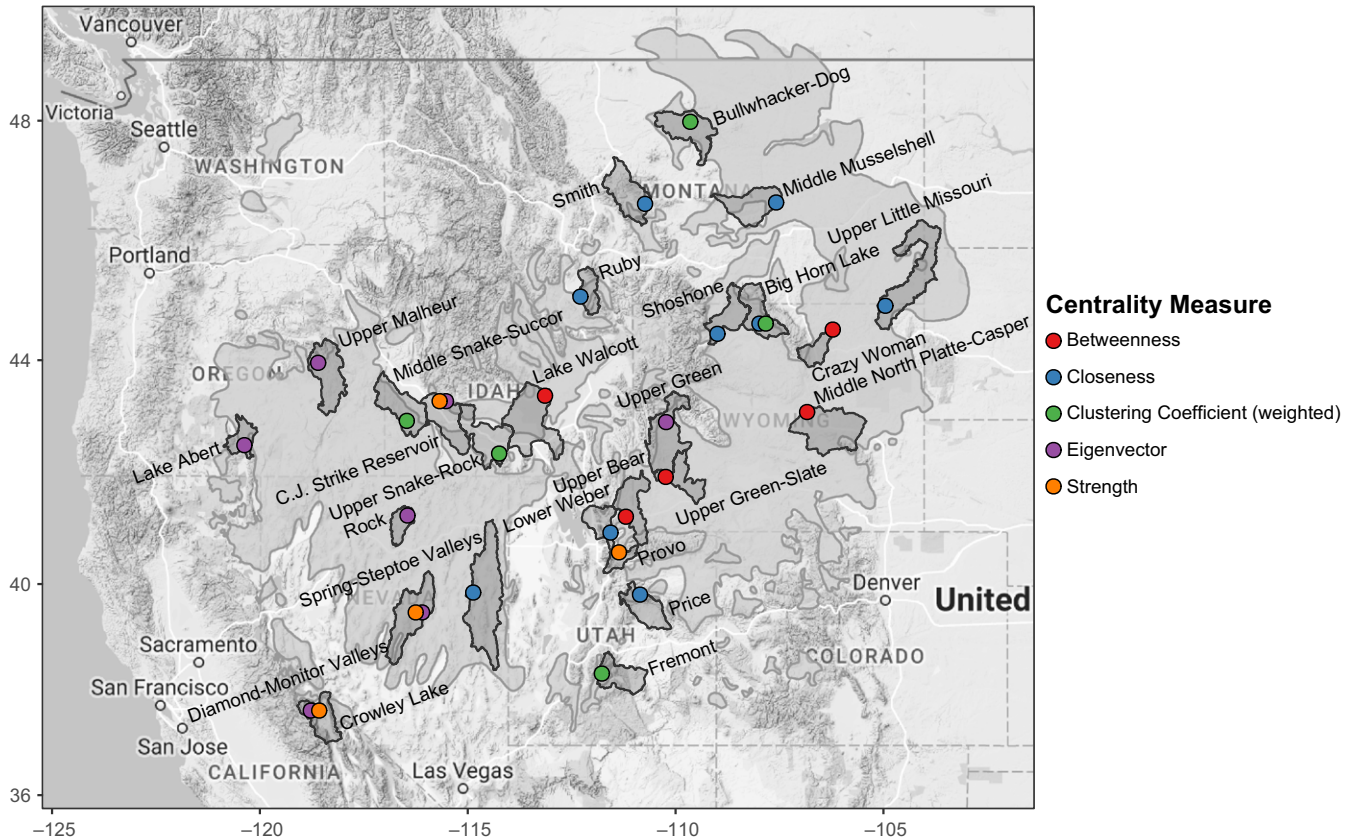


FIGURE 6 Keystone nodes ($n = 26$): nodes with greater importance to genetic connectivity than the magnitude of lek attendance within the node or node location within the species range alone might indicate. These nodes were low in mean peak high male count relative to their network centrality rankings. Points representing keystone nodes for more than one centrality index are offset to the left or right, such that these offset touching points represent the same node. Node color indicates centrality measure. Shaded polygons depict the watershed within which these top-ranking nodes are located. The species' range is shown as light gray polygons

negative correlation between mean peak male count and betweenness and between mean peak male count and closeness (Table 4 and Figure 5).

Across all centrality indices, we discovered 26 nodes that ranked high for network centrality despite having lower mean peak male count than nodes of similar ranking (Figure 6). These 26 keystone nodes were located across the entire species' range. Four of these keystone nodes ranked highly for more than one centrality index, with high rankings coupled for eigenvector centrality and strength and for closeness and clustering coefficient. In all cases, these nodes were keystone for betweenness and closeness or for eigenvector and strength.

4 | DISCUSSION

4.1 | Emergent network properties

The greatest utility of our network analysis is its ability to be used to prioritize and target conservation efforts to the nodes most important to maintaining network connectivity at any desired scale. Our network approach allows nodes to be ranked across multiple

centrality indices, indicative of different scales and patterns of connectivity, each with unique importance to conservation.

We discovered that the sage-grouse range-wide genetic network is best characterized as hub-and-spoke topology most resembling the structure of a small-world network and not that of a random or regular network. Both the mean clustering coefficient ($0.19 \pm 3.35 \times 10^{-3}$ [SE]) and the mean characteristic path length ($1.88 \pm 7.04 \times 10^{-3}$ [SE], [1.88, 1.91]) were shorter than has been reported for other species (e.g., 0.254 and 2.26 in Garroway et al., 2008). The fat-tailed distribution of node degree (Figure 4d) confirmed small-world network structure by ruling out scale-free structure, for which the degree distribution follows a power law.

Many hub nodes of connectivity within the network are located across the species' range (Figure 3), with most spoke nodes located along the periphery of the range. This hub-and-spoke topology is evident in the minimum spanning tree, with important hub nodes of genetic connectivity occurring in nearly every state across the contiguous range (Figure 2). Loss of one of these highly connected hub nodes within several major basins could severely affect overall network connectivity.

We documented strong connectivity across the entire network, evidenced by high-ranking nodes and edges across the species' range. This means that some of the nodes may be able to recover should they be extirpated but the habitat remain intact or be restored (e.g., following a local extinction caused by West Nile virus—e.g., Naugle et al., 2004; or return after restoration following natural resource extraction—e.g., Naugle et al., 2011). The ability to recover is exhibited in the network's traversability (i.e., the apparent low resistance to gene flow). However, to the best of our knowledge, node recovery has not been previously investigated in wildlife networks. The minimum spanning tree can serve as a powerful guide in making management decisions related to the relative importance of individual nodes to overall landscape connectivity (Urban & Keitt, 2001), as it is possible to model which nodes or which parts of the range will most likely be affected by the loss of any given node.

Our results suggest that distance plays an important role in structuring genetic connectivity (also known as isolation by distance, Wright, 1943). The vast majority of edges in the minimum spanning tree—those connections that represent the greatest covariance—connect geographically proximal nodes (Figure 2a). Similarly, there was a correlation between conditional genetic distance and geographic distance. These results support prior findings of isolation by distance across the species' range (Bush et al., 2011; Cross et al., 2016; Davis, Reese, Gardner, & Bird, 2015; Fedy, Row, & Oyler-McCance, 2017; Oyler-McCance, Taylor, & Quinn, 2005; Schulwitz, Bedrosian, & Johnson, 2014). However, nodes with greater centrality, important to both local and network-wide genetic exchange, are located across the species' range.

Both Cross et al. (2016) and Oyler-McCance et al. (2005) found that sage-grouse subpopulations in southwestern Montana were diverged from populations in the rest of the state. We confirm this prior finding, showing that sage-grouse from these same subpopulations are more closely related to conspecifics in Idaho than to subpopulations in Montana, as is evident in edge connectivity within the minimum spanning tree (Figure 2b). Cross et al. (2016) also found that the population in Northern Montana was diverged from the subpopulation in Southeastern Montana and the Dakotas and from the southcentral Montana subpopulation (the SE-W subpopulation in Cross et al., 2016). We confirm these findings here, showing nodes with very high clustering coefficient (indicative of highly interconnected network subunits) within the same regions (Figure 3). We expect that the other top-ranked nodes for clustering coefficient in the Middle Snake-Succor watershed in Idaho and the Fremont watershed in Utah might also be embedded at the core of their respective subpopulations. Schulwitz et al. (2014) found that the subpopulations in southern and southeastern Montana and the Dakotas were both highly connected to leks in northern Wyoming. We also found the same pattern of connectivity, evident in the hub-and-spoke topology of the minimum spanning tree. In our case, a hub node in Wyoming/southcentral Montana is located in the Big Horn Lake watershed of northcentral Wyoming, and a hub node for Wyoming/southeastern Montana

subpopulations is located within the Crazy Woman watershed of northeastern Wyoming (Figures 2b and 3). Davis et al. (2015) found that the small northern California population known to have experienced population declines had retained genetic diversity. We confirm this understanding by finding that the nodes in this area show elevated local connectivity (covariance) within the area. We also found that genetic connectivity into the northern California nodes comes from nodes to the north in Oregon (Figure 2b). Oyler-McCance, Casazza, Fike, and Coates (2014) discovered a northern and a southern subpopulation within the Bi-State population in southern California and southwestern Nevada. We found the same break evidenced by a lack of edges connecting these two units in the minimum spanning tree (Figure 2b). This lack of interconnectivity among nodes in the northern and southern groups is especially surprising, given that both groups exhibit greater covariance with far more geographically distant nodes. Fedy et al. (2017) documented genetic differentiation between birds in the Bighorn and Powder River Basins of Wyoming as well as differentiation between the northern and southern parts of the state, differences reflected in our analysis as evidenced by edge connectivity within the minimum spanning tree.

4.2 | Hubs of genetic exchange

We identified nodes with high importance to large-scale, network-wide genetic connectivity (i.e., nodes with high betweenness), and nodes within the top 50% of all centrality indices important to both network-wide and local connectivity. These top-ranked hub nodes are located across the entire range of the species. The locations of these hub nodes important to network-wide connectivity are in areas that should foster range-wide genetic connectivity due to their location in the topography of the western landscape.

Connectivity of these hubs is apparent in the minimum spanning tree (Figure 2b) where connectivity across the range appears crescent shaped, with one point of the crescent in northern Montana/Saskatchewan and the other in Oregon. The Upper Snake Basin of Idaho (Lake Walcott watershed) forms a thumb terminating in southwestern Montana to the northeast. Hub connectivity opens up over the Columbia Plateau of Idaho (Upper Snake-Rock, C.J. Strike Reservoir, and Middle Snake-Succor watersheds). Connectivity extends south into the Great Basin composing most of eastern California, all of Nevada and western Utah. Here, the Southwest River Basin in Idaho (Middle Snake-Succor and C.J. Strike Reservoir watersheds) connects to the Malheur Basin (Upper Malheur watershed) to the west and to the South Lahontan River Basin (Crowley Lake watershed) by way of the Central Nevada Desert (Diamond-Monitor Valleys watershed) to the southwest. The Green River Basin of Wyoming (Upper Green and Upper Green-Slate watersheds) sits just west of a low section of the North American Continental Divide connecting the Upper Snake Basin and Great Basin to the rest of Wyoming and farther up into the northeastern part of the species range. The Green River Basin also sits just north of the Yampa and White River Basins in Colorado (Little Snake watershed), and the

Bear River Basin in Utah (Upper Bear watershed), both connected by low valleys to the south. Similarly, the Powder and Tongue River Basins of Wyoming (Crazy Woman watershed) connect to the north with both of the Dakotas and eastern Montana. The Bighorn River Basin, which ranks lower for other connectivity indices, connects to the southeastern-west subpopulation in the Yellowstone River Basin of Montana (Cross et al., 2016)—nodes in the Big Horn Lake watershed anchor both basins. We suspect that the topology of the genetic network is largely shaped by the topography of the landscape, a hypothesis previously posited for sage-grouse (Cross et al., 2016; Row et al., 2015; Schulwitz et al., 2014), and which has been found to influence genetic structure in other species (e.g., Roffler et al., 2014).

We identified 26 keystone nodes across the range of sage-grouse that stand out with increased importance to genetic connectivity despite having lower mean peak male count (Figure 6). These keystone nodes do not follow the presupposition of increased centrality with increased mean peak male count (i.e., a proxy for population size for any given node) and include the highest ranking nodes for each centrality index, regardless of the population size (Figure 5). We believe that these keystone nodes and other hub nodes (Figure 3) are top candidates for targeted conservation efforts, as their protection will help secure range-wide genetic connectivity. The keystone nodes are also distributed across the entire species' range, from the core to the periphery (Figure 6). Therefore, neither range centrality nor local population size alone should be trusted proxies for prioritizing targeted conservation actions for sage-grouse.

4.3 | Limitations of the study and future directions

Prior research has modeled range-wide sage-grouse connectivity using a network approach. Knick and Hanser (2011) weighted nodes using lek attendance and limited edge connections using hypothesized dispersal thresholds. However, these imposed dispersal thresholds may have affected the resultant network structure. For example, Knick and Hanser (2011) used an exponential decay function to determine the probability of connectivity of leks. Imposing dispersal thresholds likely oversimplified the contribution that each priority area for conservation made to network connectivity by assuming dispersal limitations is equal among all nodes regardless of the internal population dynamics within nodes and environmental conditions within and among nodes. Crist, Knick, and Hanser (2017) used network approaches to generate several models of hypothesized connectivity among sage-grouse priority areas for conservation, which are areas that protect larger leks (i.e., those with more males visible during breeding) and surrounding area. They characterized the centrality of each priority area for conservation and concluded that several subnetworks exist across the species' range. However, in their analysis, patch size, shape, and boundary length all had an effect on the pattern of connectivity and centrality. Our analysis provides insight into genetic connectivity using centrality indices based solely on the species' biology: the genetic covariance resulting from cumulative dispersal and breeding, a quantitative metric.

We have confidence in the cut distance we used to cluster leks into nodes, as it is empirically based on dispersal distances documented over a vast area, across multiple years, involving both sexes (Cross et al., 2017). Our clustering approach increased genetic variance within nodes, but also increased covariance among nodes (Dyer, 2015). Choice of cut distance depends on the desired scale of analysis for conservation and management application. We could have performed this analysis using individual leks. Doing so would have resulted in finer resolution for our results. However, it also would have resulted in fewer individuals per node, which would have limited our characterization of within-node genetic variation. Furthermore, we would have had to cut many leks from our analysis due to the minimum node composition requirement of four individuals. By clustering leks into nodes, we were restricted to making statements about the connectivity of larger landscapes that extend beyond the size of an individual lek and which were potentially representative of leks unsampled within the same landscapes. Furthermore, our clustering approach reflects the biology of the species, as prior research has shown that both female and male sage-grouse attend multiple leks within a breeding season (Cross et al., 2017; Dunn & Braun, 1985; Semple, Wayne, & Gibson, 2001).

We found evidence for correlation between some network centrality indices and samples per node and mean peak male count. However, when significant, these relationships were only moderate or weak in all but one case: that of betweenness and samples within a node (Table 4). Therefore, we do not believe that sample size drove the centrality of a node. Larger populations acting as hub nodes might be expected, as these highly populated hub nodes would be expected to house greater genetic diversity to be the sources of dispersers. However, as discussed above, the highest ranking nodes for each centrality index were never those with the greatest mean peak male count (Figure 5).

Future work should examine the effect of the spatial distribution of individuals composing nodes on the resultant network model structure. For example, constructing a genetic network where priority areas for conservation serve as nodes may help prioritize conservation based on existing management boundaries at a larger landscape scale. It is worth noting that if priority areas for conservation are treated as nodes, larger priority areas for conservation may score higher for centrality indices due to the within-node proportion of the genetic covariance, which will increase centrality.

4.4 | Applications and future directions

We believe that the greatest utility of our network analysis will be its use in prioritizing and targeting conservation efforts to the nodes most important to maintaining network connectivity. This network approach allows for the ranking of nodes by multiple centrality indices, indicative of different scales and different patterns of connectivity. These indices can be used to locate the top-ranking nodes—and more importantly, the leks which compose those nodes—which can then be prioritized in accordance with management goals (Bottrill

et al., 2008). If goals are to conserve hubs of genetic exchange that connect the greatest number of nodes range-wide, then ranking nodes based on betweenness is most relevant. If goals are to conserve hubs of genetic exchange that connect immediate connections, then ranking nodes based on eigenvector centrality is most relevant. If goals are to conserve nodes that have the greatest genetic exchange with their immediate connections, then ranking nodes based on strength is most relevant. If goals are to conserve hubs of local connectivity, then ranking nodes based on closeness or clustering coefficient is most relevant.

Conservation actions may be targeted first toward the top-ranking nodes, or managers may first choose to combine network centrality with economic cost before deciding where to act. We can imagine many additional ways in which network centrality may be combined with additional metrics to target conservation resources. Our hope is that the empirically based sage-grouse genetic network we constructed will prove a useful tool to conservation planners.

ACKNOWLEDGMENTS

We thank Rodney Dyer, Fred Allendorf, Jeff Good, Jon Graham, and Katie Zarn for their comments on this manuscript. This research would not have been possible without Steve Knick's leadership and the many state, federal, and nongovernmental organization (NGO) biologists and technicians who have spent lifetimes working tirelessly in the field and behind desks for the conservation of the Greater sage-grouse. We specifically acknowledge biologists and technicians with the Bureau of Land Management, California Department of Fish & Game, Colorado Division of Wildlife, Idaho Fish & Game, Montana Fish, Wildlife & Parks, Montana Audubon, Nevada Division of Wildlife, North Dakota Game and Fish, Natural Resources Conservation Service: Sage-Grouse Initiative, Oregon Fish & Wildlife, South Dakota Game, Fish & Parks, Utah Wildlife Resources, U.S. Forest Service (USFS), the Western Association of Fish and Wildlife Agencies (WAFWA), Washington Department of Fish and Wildlife, and Wyoming Game & Fish Department. Our molecular genetics analyses and manuscript preparation would not have been possible without Cory Engkjer, Kevin McKelvey, and Kristy Pilgrim, with laboratory work by Kara Bates, Nasreen Broomand, Taylor Dowell, Jennifer Fike, Scott Hampton, Randi Lesagonicz, Inga Ortloff, Sara Schwarz, Kate Welch, the USFS Rocky Mountain Research Station National Genomics Center for Wildlife and Fish Conservation, and the USGS FORT Molecular Ecology Laboratory.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

T.B.C. conceived the idea, analyzed the data, conducted the research, and wrote the paper. T.B.C. and M.K.S. developed and designed the

methods. T.B.C. and S.J.O. generated the data. T.B.C., D.E.N., S.J.O., J.R.R. B.C.F, S.T.K, and M.K.S. substantially edited the paper.

DATA ACCESSIBILITY

Sample microsatellite genotypes and node membership are available on USGS ScienceBase: <https://doi.org/10.5066/f73n22pn>.

REFERENCES

- Aldridge, C. L., Nielsen, S. E., Beyer, H. L., Boyce, M. S., Connelly, J. W., Knick, S. T., & Schroeder, M. A. (2008). Range-wide patterns of greater sage-grouse persistence. *Diversity and Distributions*, *14*, 983–994. <https://doi.org/10.1111/j.1472-4642.2008.00502.x>
- Amaral, L. A. N., Scala, A., Barthélemy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences*, *97*, 11149–11152. <https://doi.org/10.1073/pnas.200327197>
- Beever, E. A., & Aldridge, C. L. (2011). Influences of free-roaming equids on sagebrush ecosystems, with a focus on greater sage-grouse. In S. T. Knick & J. W. Connelly (Eds.), *Greater Sage-Grouse: Ecology and conservation of a landscape species and its habitats. Studies in avian biology* (pp. 273–290), 38th edn. Berkeley, CA: University of California Press.
- Bivand, R. S., & Rundel, C. (2017). RGEOS: Interface to Geometry Engine – Open Source (GEOS). R package version 0.3-26. <https://CRAN.R-project.org/package=rgeos>.
- Bottrill, M. C., Joseph, L. N., Carwardine, J., Bode, M., Cook, C., Game, E. T., ... Pressey, R. L. (2008). Is conservation triage just smart decision making? *Trends in Ecology & Evolution*, *23*, 649–654. <https://doi.org/10.1016/j.tree.2008.07.007>
- Bradbury, J., Vehrencamp, S., & Gibson, R. (1989). Dispersion of displaying male sage-grouse. *Behavioral Ecology and Sociobiology*, *24*, 1–14. <https://doi.org/10.1007/BF00300112>
- Bray, D. (2003). Molecular networks: The top-down view. *Science*, *301*, 1864–1865. <https://doi.org/10.1126/science.1089118>
- Bunn, A. G., Urban, D. L., & Keitt, T. H. (2000). Landscape connectivity: A conservation application of graph theory. *Journal of Environmental Management*, *59*, 265–278. <https://doi.org/10.1006/jema.2000.0373>
- Bush, K. L., Dyte, C. K., Moynahan, B. J., Aldridge, C. L., Sauls, H. S., Battazzo, A. M., ... Carlson, J. (2011). Population structure and genetic diversity of greater sage-grouse (*Centrocercus urophasianus*) in fragmented landscapes at the northern edge of their range. *Conservation Genetics*, *12*, 527–542. <https://doi.org/10.1007/s10592-010-0159-8>
- Bush, K. L., Vinsky, M. D., Aldridge, C. L., & Paszkowski, C. A. (2005). A comparison of sample types varying in invasiveness for use in DNA sex determination in an endangered population of greater sage-grouse (*Centrocercus urophasianus*). *Conservation Genetics*, *6*, 867–870. <https://doi.org/10.1007/s10592-005-9040-6>
- Connelly, J. W., Hagen, C. A., & Schroeder, M. A. (2011). Characteristics and dynamics of greater sage-grouse populations. In S. T. Knick & J. W. Connelly (Eds.), *Greater Sage-Grouse: Ecology and conservation of a landscape species and its habitats. Studies in avian biology* (pp. 53–67), 38th edn. Berkeley, CA: University of California Press.
- Connelly, J. W., Rinkes, E., & Braun, C. (2011). Characteristics of greater sage-grouse habitats: A landscape species at micro and macro scales. In S. T. Knick & J. W. Connelly (Eds.), *Greater Sage-Grouse: Ecology and conservation of a landscape species and its habitats. Studies in avian biology* (pp. 69–83), 38th edn. Berkeley, CA: University of California Press.
- Copeland, H. E., Doherty, K. E., Naugle, D. E., Pocewicz, A., & Kiesecker, J. M. (2009). Mapping oil and gas development potential in the US

- Intermountain West and estimating impacts to species. *PLoS ONE*, 4, e7400. <https://doi.org/10.1371/journal.pone.0007400>
- Crist, M. R., Knick, S. T., & Hanser, S. E. (2017). Range-wide connectivity of priority areas for Greater Sage-Grouse: Implications for long-term conservation from graph theory. *Condor: Ornithological Applications*, 119, 44–57. <https://doi.org/10.1650/CONDOR-16-60.1>
- Crooks, K. R., & Sanjayan, M. (2006). Connectivity conservation: Maintaining connections for nature. In K. R. Crooks & M. Sanjayan (Eds.), *Connectivity conservation. Conservation biology series* (vol. 14, pp. 383–405). Cambridge, MA: Cambridge University Press. <https://doi.org/10.1017/CBO9780511754821>
- Cross, T. B., Naugle, D. E., Carlson, J. C., & Schwartz, M. K. (2016). Hierarchical population structure in greater sage-grouse provides insight into management boundary delineation. *Conservation Genetics*, 17, 1417–1433. <https://doi.org/10.1007/s10592-016-0872-z>
- Cross, T. B., Naugle, D. E., Carlson, J. C., & Schwartz, M. K. (2017). Genetic recapture identifies long-distance breeding dispersal in Greater Sage-Grouse (*Centrocercus urophasianus*). *Condor: Ornithological Applications*, 119, 155–166. <https://doi.org/10.1650/CONDOR-16-178.1>
- Csardi, G., & Nepusz, T. (2006). The IGRAPH software package for complex network research. *InterJournal, Complex Systems*, 1695, 38.
- Dalke, P. D., Pyrah, D. B., Stanton, D. C., Crawford, J. E., & Schlatterer, E. F. (1963). Ecology, productivity, and management of sage-grouse in Idaho. *Journal of Wildlife Management*, 27, 811–841.
- Davis, D. M., Reese, K. P., Gardner, S. C., & Bird, K. L. (2015). Genetic structure of Greater Sage-Grouse (*Centrocercus urophasianus*) in a declining, peripheral population. *Condor: Ornithological Applications*, 117, 530–544. <https://doi.org/10.1650/CONDOR-15-34.1>
- DeWoody, J., Nason, J. D., & Hipkins, V. D. (2006). Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes*, 6, 951–957. <https://doi.org/10.1111/j.1471-8286.2006.01449.x>
- Dunn, P. O., & Braun, C. E. (1985). Natal dispersal and lek fidelity of sage-grouse. *Auk*, 102, 621–627.
- Dunn, P. O., & Braun, C. E. (1986). Late summer–spring movements of juvenile Sage-grouse. *Wilson Bulletin*, 98, 83–92.
- Dunne, J. A., Williams, R. J., & Martinez, N. D. (2002). Network structure and biodiversity loss in food webs: Robustness increases with connectance. *Ecology Letters*, 5, 558–567. <https://doi.org/10.1046/j.1461-0248.2002.00354.x>
- Dyer, R. J. (2007). The evolution of genetic topologies. *Theoretical Population Biology*, 71, 71–79. <https://doi.org/10.1016/j.tpb.2006.07.001>
- Dyer, R. J. (2014a). GSTUDIO: Spatial utility functions from the Dyer laboratory. R package version 1.2.-
- Dyer, R. J. (2014b). POPGRAPH: This is an R package that constructs and manipulates population graphs. R package version 1.4.
- Dyer, R. J. (2015). Population graphs and landscape genetics. *Annual Review of Ecology, Evolution, and Systematics*, 46, 327–342. <https://doi.org/10.1146/annurev-ecolsys-112414-054150>
- Dyer, R. J., & Nason, J. D. (2004). Population graphs: The graph theoretic shape of genetic structure. *Molecular Ecology*, 13, 1713–1727. <https://doi.org/10.1111/j.1365-294X.2004.02177.x>
- Dyer, R. J., Nason, J. D., & Garrick, R. C. (2010). Landscape modelling of gene flow: Improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology*, 19, 3746–3759. <https://doi.org/10.1111/j.1365-294X.2010.04748.x>
- Edminster, F. C. (1954). *American game birds of old and forest: Their habits, ecology, and management*. New York: Scribner.
- Emmons, S. R., & Braun, C. E. (1984). Lek attendance of male sage-grouse. *Journal of Wildlife Management*, 48, 1023–1028. <https://doi.org/10.2307/3801461>
- Evett, I. W., & Weir, B. S. (1998). *Interpreting DNA evidence: Statistical genetics for forensic scientists*. Sunderland, MA, USA: Sinauer Associates.
- Fedy, B. C., Row, J. R., & Oyler-McCance, S. J. (2017). Integration of genetic and demographic data to assess population risk in a continuously distributed species. *Conservation Genetics*, 18, 89–104. <https://doi.org/10.1007/s10592-016-0885-7>
- Galpern, P., Manseau, M., Hettinga, P., Smith, K., & Wilson, P. (2012). ALLELEMATCH: An R package for identifying unique multi-locus genotypes where genotyping error and missing data may be present. *Molecular Ecology Resources*, 12, 771–778. <https://doi.org/10.1111/j.1755-0998.2012.03137.x>
- Garroway, C. J., Bowman, J., Carr, D., & Wilson, P. J. (2008). Applications of graph theory to landscape genetics. *Evolutionary Applications*, 1, 620–630.
- Garroway, C. J., Bowman, J., & Wilson, P. J. (2011). Using a genetic network to parameterize a landscape resistance surface for fishers, *Martes pennanti*. *Molecular Ecology*, 20, 3978–3988. <https://doi.org/10.1111/j.1365-294X.2011.05243.x>
- Hagen, C. A., Connelly, J. W., & Schroeder, M. A. (2007). A meta-analysis of greater sage-grouse *Centrocercus urophasianus* nesting and brood-rearing habitats. *Wildlife Biology*, 13, 42–50. [https://doi.org/10.2981/0909-6396\(2007\)13\[42:AMOGSC\]2.0.CO;2](https://doi.org/10.2981/0909-6396(2007)13[42:AMOGSC]2.0.CO;2)
- Hanks, E. M., Hooten, M. B., Knick, S. T., Oyler-McCance, S. J., Fike, J. A., Cross, T. B., & Schwartz, M. K. (2016). Latent spatial models and sampling design for landscape genetics. *Annals of Applied Statistics*, 10, 1041–1062. <https://doi.org/10.1214/16-AOAS929>
- Holloran, M. J., Kaiser, R. C., & Hubert, W. A. (2010). Yearling greater sage-grouse response to energy development in Wyoming. *Journal of Wildlife Management*, 74, 65–72. <https://doi.org/10.2193/2008-291>
- Jacoby, D. M. P., & Freeman, R. (2016). Emerging network-based tools in movement ecology. *Trends in Ecology & Evolution*, 31, 301–314. <https://doi.org/10.1016/j.tree.2016.01.011>
- Kahn, N., St. John, J., & Quinn, T. W. (1998). Chromosome-specific intron size differences in the avian CHD gene provide an efficient method for sex identification in birds. *Auk: Ornithological Applications*, 115, 1074–1078.
- Knick, S. T., Dobkin, D. S., Rotenberry, J. T., Schroeder, M. A., Vander Haegen, W. M., & Van Riper, C. I. I. (2003). Teetering on the edge or too late? Conservation and research issues for avifauna of sagebrush habitats. *Condor*, 105, 611–634. <https://doi.org/10.1650/7329>
- Knick, S. T., & Hanser, S. E. (2011). Connecting pattern and process in greater sage-grouse populations and sagebrush landscapes. In S. T. Knick & J. W. Connelly (Eds.), *Greater Sage-Grouse: Ecology and conservation of a landscape species and its habitats. Studies in avian biology* (pp. 383–405), 38th edn. Berkeley, CA: University of California Press.
- Koen, E. L., Bowman, J., Garroway, C. J., & Wilson, P. J. (2013). The sensitivity of genetic connectivity measures to unsampled and under-sampled sites. *PLoS ONE*, 8, e56204. <https://doi.org/10.1371/journal.pone.0056204>
- Koen, E. L., Bowman, J., & Wilson, P. J. (2015). Node-based measures of connectivity in genetic networks. *Molecular Ecology Resources*, 16, 69–79.
- Lookingbill, T. R., Gardner, R. H., Ferrari, J. R., & Keller, C. E. (2010). Combining a dispersal model with network theory to assess habitat connectivity. *Geography and the Environment*, 20, 427–441.
- McKelvey, K. S., & Schwartz, M. K. (2005). DROPOUT: A program to identify problem loci and samples for noninvasive genetic samples in a capture-mark-recapture framework. *Molecular Ecology Notes*, 5, 716–718. <https://doi.org/10.1111/j.1471-8286.2005.01038.x>
- Naugle, D. E., Aldridge, C. L., Walker, B. L., Cornish, T. E., Moynahan, B. J., Holloran, M. J., ... Mayer, R. T. (2004). West Nile virus: Pending crisis for greater sage-grouse. *Ecology Letters*, 7, 704–713. <https://doi.org/10.1111/j.1461-0248.2004.00631.x>
- Naugle, D. E., Doherty, K. E., Walker, B. L., Copel, H. E., Holloran, M. J., & Tack, J. D. (2011). Sage-grouse and cumulative impacts of energy development. In D. E. Naugle (Ed.), *Energy development and wildlife*

- conservation in western North America (pp. 55–70). Washington: Island. <https://doi.org/10.5822/978-1-61091-022-4>
- Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103, 8577–8582. [www.pnas.org/cgi/doi/10.1073/pnas.0601602103](https://doi.org/10.1073/pnas.0601602103)
- Nychka, D., Furrer, R., Paige, J., & Sain, S. (2015). FIELDS: Tools for spatial data. <https://doi.org/10.5065/d6w957ct> (URL: <https://doi.org/10.5065/d6w957ct>), R package version 8.10, www.image.ucar.edu/fields.
- Oyler-McCance, S. J., Casazza, M. L., Fike, J. A., & Coates, P. S. (2014). Hierarchical spatial genetic structure in a distinct population segment of greater sage-grouse. *Conservation Genetics*, 15, 1299–1311. <https://doi.org/10.1007/s10592-014-0618-8>
- Oyler-McCance, S. J., Taylor, S. E., & Quinn, T. W. (2005). A multilocus population genetic survey of the greater sage-grouse across their range. *Molecular Ecology*, 14, 1293–1310. <https://doi.org/10.1111/j.1365-294X.2005.02491.x>
- Patterson, R. L. (1952). *The Sage-grouse in Wyoming*. Denver, CO, USA: Sage Books, for Wyoming Game and Fish Commission, Cheyenne, WY, USA.
- R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>
- Remington, T. E., & Braun, C. E. (1985). Sage-grouse food selection in winter, North Park, Colorado. *Journal of Wildlife Management*, 49, 1055–1061. <https://doi.org/10.2307/3801395>
- Roffler, G. H., Talbot, S. L., Luikart, G. H., Sage, G. K., Pilgrim, K. L., Adams, L. G., & Schwartz, M. K. (2014). Lack of sex-biased dispersal promotes fine-scale genetic structure in alpine ungulates. *Conservation Genetics*, 15, 837–851. <https://doi.org/10.1007/s10592-014-0583-2>
- Row, J. R., Oyler-McCance, S. J., Fike, J. A., O'Donnell, M. S., Doherty, K. E., Aldridge, C. L., ... Fedy, B. C. (2015). Landscape characteristics influencing the genetic structure of greater sage-grouse within the stronghold of their range: A holistic modeling approach. *Ecology and Evolution*, 5, 1955–1969. <https://doi.org/10.1002/ece3.1479>
- Rowland, M. M., Wisdom, M. J., Suring, L. H., & Meinke, C. W. (2006). Greater sage-grouse as an umbrella species for sagebrush-associated vertebrates. *Biological Conservation*, 129, 323–335. <https://doi.org/10.1016/j.biocon.2005.10.048>
- Sallaberry, A., Zaidi, F., & Melançon, G. (2013). Model for generating artificial social networks having community structures with small-world and scale-free properties. *Social Network Analysis and Mining*, 3, 1–13.
- Schrag, A., Konrad, S., Miller, S., Walker, B., & Forrest, S. (2011). Climate change impacts on sagebrush habitat and West Nile virus transmission risk and conservation implications for greater sage-grouse. *GeoJournal*, 76, 561–575. <https://doi.org/10.1007/s10708-010-9369-3>
- Schroeder, M. A., Aldridge, C. L., Apa, A. D., Bohne, J. R., Braun, C. E., Bunnell, S. D., ... Hilliard, M. A. (2004). Distribution of sage-grouse in North America. *Condor*, 106, 363–376. <https://doi.org/10.1650/7425>
- Schroeder, M. A., & Robb, L. A. (2003). Fidelity of Greater Sage-Grouse *Centrocercus urophasianus* to breeding areas in a fragmented landscape. *Wildlife Biology*, 9, 291–299.
- Schulwitz, S., Bedrosian, B., & Johnson, J. A. (2014). Low neutral genetic diversity in isolated Greater Sage-Grouse (*Centrocercus urophasianus*) populations in northwest Wyoming. *Condor: Ornithological Applications*, 116, 560–573. <https://doi.org/10.1650/CONDOR-14-54.1>
- Segelbacher, G. (2002). Noninvasive genetic analysis in birds: Testing reliability of feather samples. *Molecular Ecology Notes*, 2, 367–369.
- Semple, K., Wayne, R. K., & Gibson, R. M. (2001). Microsatellite analysis of female mating behaviour in lek-breeding sage-grouse. *Molecular Ecology*, 10, 2043–2048. <https://doi.org/10.1046/j.0962-1083.2001.01348.x>
- Smith, R. E. (2012). Conserving Montana's sagebrush highway: Long distance migration in sage-grouse. M.S. thesis, University of Montana, Missoula, MT, USA.
- Schwartz, M. K., Cushman, S. A., McKelvey, K. S., Hayden, J., & Engkjer, C. (2006). Detecting genotyping errors and describing American black bear movement in northern Idaho. *Ursus*, 17, 138–148. [https://doi.org/10.2192/1537-6176\(2006\)17\[138:DGEADA\]2.0.CO;2](https://doi.org/10.2192/1537-6176(2006)17[138:DGEADA]2.0.CO;2)
- Tack, J. D. (2009). Sage-grouse and the human footprint: Implications for conservation of small and declining populations. M.S. thesis, University of Montana, Missoula, MT, USA.
- Urban, D. L., & Keitt, T. (2001). Landscape connectivity: A graph-theoretic perspective. *Ecology*, 82, 1205–1218. [https://doi.org/10.1890/0012-9658\(2001\)082\[1205:LCAGTP\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2001)082[1205:LCAGTP]2.0.CO;2)
- U.S. Fish and Wildlife Service (2010). Endangered and threatened wildlife and plants; 12-month findings for petitions to list the Greater Sage-Grouse (*Centrocercus urophasianus*) as Threatened or Endangered; proposed rule. *Federal Register*, 75, 13910–14014.
- U.S. Fish and Wildlife Service (2015). Endangered and threatened wildlife and plants; 12-month finding on a petition to list Greater Sage-Grouse (*Centrocercus urophasianus*) as an Endangered or Threatened species; proposed rule. *Federal Register*, 80, 59858–59942.
- Van Oosterhout, C., Hutchinson, W. F., Wills, D. P., & Shipley, P. (2004). MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4, 535–538. <https://doi.org/10.1111/j.1471-8286.2004.00684.x>
- Walker, B. L., Naugle, D. E., & Doherty, K. E. (2007). Greater sage-grouse population response to energy development and habitat loss. *Journal of Wildlife Management*, 71, 2644–2654. <https://doi.org/10.2193/2006-529>
- Wallestad, R., & Eng, R. L. (1975). Foods of adult sage-grouse in Central Montana. *Journal of Wildlife Management*, 39, 628–630. <https://doi.org/10.2307/3800409>
- Wallestad, R., & Schladweiler, P. (1974). Breeding season movements and habitat selection of male sage-grouse. *Journal of Wildlife Management*, 38, 634–637. <https://doi.org/10.2307/3800030>
- Watts, D., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393, 440–442. <https://doi.org/10.1038/30918>
- Western Association of Fish and Wildlife Agencies [WAFWA] (2015). Greater sage-grouse population trends: an analysis of lek count databases 1965–2015. Report. Western Association of Fish and Wildlife Agencies, Cheyenne, WY, USA.
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16, 97–159.
- Wright, S. (1943). Isolation by distance. *Genetics*, 28, 114.

How to cite this article: Cross TB, Schwartz MK, Naugle DE, Fedy BC, Row JR, Oyler-McCance SJ. The genetic network of greater sage-grouse: Range-wide identification of keystone hubs of connectivity. *Ecol Evol*. 2018;00:1–19. <https://doi.org/10.1002/ece3.4056>