

INVITED REVIEWS AND SYNTHESSES

Application of network methods for understanding evolutionary dynamics in discrete habitats

GILI GREENBAUM*  and NINA H. FEFFERMAN†

*Department of Solar Energy and Environmental Physics and Mitrani Department of Desert Ecology, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion 84990, Israel, †Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, 37996 TN, USA

Abstract

In populations occupying discrete habitat patches, gene flow between habitat patches may form an intricate population structure. In such structures, the evolutionary dynamics resulting from interaction of gene-flow patterns with other evolutionary forces may be exceedingly complex. Several models describing gene flow between discrete habitat patches have been presented in the population-genetics literature; however, these models have usually addressed relatively simple settings of habitable patches and have stopped short of providing general methodologies for addressing nontrivial gene-flow patterns. In the last decades, network theory – a branch of discrete mathematics concerned with complex interactions between discrete elements – has been applied to address several problems in population genetics by modelling gene flow between habitat patches using networks. Here, we present the idea and concepts of modelling complex gene flows in discrete habitats using networks. Our goal is to raise awareness to existing network theory applications in molecular ecology studies, as well as to outline the current and potential contribution of network methods to the understanding of evolutionary dynamics in discrete habitats. We review the main branches of network theory that have been, or that we believe potentially could be, applied to population genetics and molecular ecology research. We address applications to theoretical modelling and to empirical population-genetic studies, and we highlight future directions for extending the integration of network science with molecular ecology.

Keywords: gene flow, network science, network theory, patch network, population genetics

Received 15 November 2016; revision received 3 February 2017; accepted 6 February 2017

Introduction

Many organisms in nature inhabit only discrete habitable patches within a continuous spatial matrix. This is mostly a result of physiological, behavioural and ecological constraints of the organism in question, and often also due to human-induced fragmentation processes. In such population structures, gene flow, selection and genetic drift interact to affect important evolutionary processes such as local and global adaptation (Kawecki & Ebert 2004), migration load (García-Ramos & Kirkpatrick 1997; Bolnick & Nosil 2007), gene swamping (Lenormand 2002) and genetic diversity loss

(Templeton 2006; Allendorf *et al.* 2012). The evolutionary consequences of such dynamics in discrete population structures have been of great interest in the population-genetic literature, and understanding the effect of gene flow between habitable patches has been the focus of many modelling efforts.

The classic continent–island model (Haldane 1930; Wright 1931) describes simple source–sink dynamics, while the full-island model (Wright 1931; Levene 1953) assumes discrete habitat patches where gene flow occurs simultaneously and equally between all patches (Fig. 1A). A more explicit spatial element was introduced by Kimura and Weiss (Kimura & Weiss 1964) with the stepping-stone models, where patches are ordered on a one-dimensional chain (Fig. 1B) or a two-dimensional lattice (Fig. 1C), gene flow occurring

Correspondence: Gili Greenbaum, Fax: +972 8 6596985; E-mail: gili.greenbaum@gmail.com

between adjacent patches. The stepping-stone models can also include a long-distance migration component (Kimura & Weiss 1964) with a designated migration parameter, usually much smaller than the migration rate between adjacent patches, describing migration between patches regardless of their spatial position, equally for all patches (Fig. 1B and C). The metapopulation dynamics framework has also been applied to study evolutionary dynamics in discrete habitats in a spatially explicit context, but relaxing the equidistance assumption of the stepping-stone model (Heino & Hanski 2001; Hanski & Heino 2003; Hanski *et al.* 2011). These spatially realistic metapopulation models (SRMM) allow for arbitrary positioning of patches in space, with gene flow occurring at decreasing-with-distance rates (Fig. 1D). Although simple in the characterizations of spatial relations between patches, the study of these and similar models has proven extremely useful for evolutionary theory.

However, interactions between patches in real populations are often nontrivial and cannot be simplified to such an extent, as in discrete habitats and fragmented landscapes migration often forms a complex gene-flow pattern which could be uncorrelated, to some extent, with geographic distance. Can such complexities be incorporated in our models to better explain and predict evolutionary processes? Can we develop new methodologies for studying populations with complex gene-flow patterns? Are there

evolutionary phenomena that we are failing to explain by relying on simplistic models? While these questions pose difficult challenges, the extensive mathematical discipline of network theory may provide an appropriate framework for addressing some of these issues.

Network theory is a relatively new discipline which has seen application in diverse fields, such as sociology (Easley & Kleinberg 2010), ecology (Proulx *et al.* 2005; Bascompte 2007; Greenbaum *et al.* 2015), real estate markets (Seiler *et al.* 2014), epidemiology (Keeling & Eames 2005; Fefferman & Ng 2007), criminology (Calvo-Armengol & Zenou 2004), animal behaviour (Croft *et al.* 2008; Hock & Fefferman 2011), evolutionary biology (Pickrell & Pritchard 2012; Greening & Fefferman 2014) and many others. This discipline, a branch of discrete mathematics, studies properties of mathematical constructs composed of discrete elements and connections of various types between them. The network framework allows exploration of complex topologies of interactions between discrete elements; methods and concepts for studying such topologies are continuously being developed (Wasserman & Faust 1994; Matousek & Nesetril 1998; Barabási & Albert 1999; Carrington *et al.* 2005; Hanneman & Riddle 2005; Bornholdt & Heinz 2006; Newman 2010; Sterbenz *et al.* 2011; Boccaletti *et al.* 2014; Lordan *et al.* 2014).

General migration-matrix models have been suggested in the population-genetics literature (Bodmer &

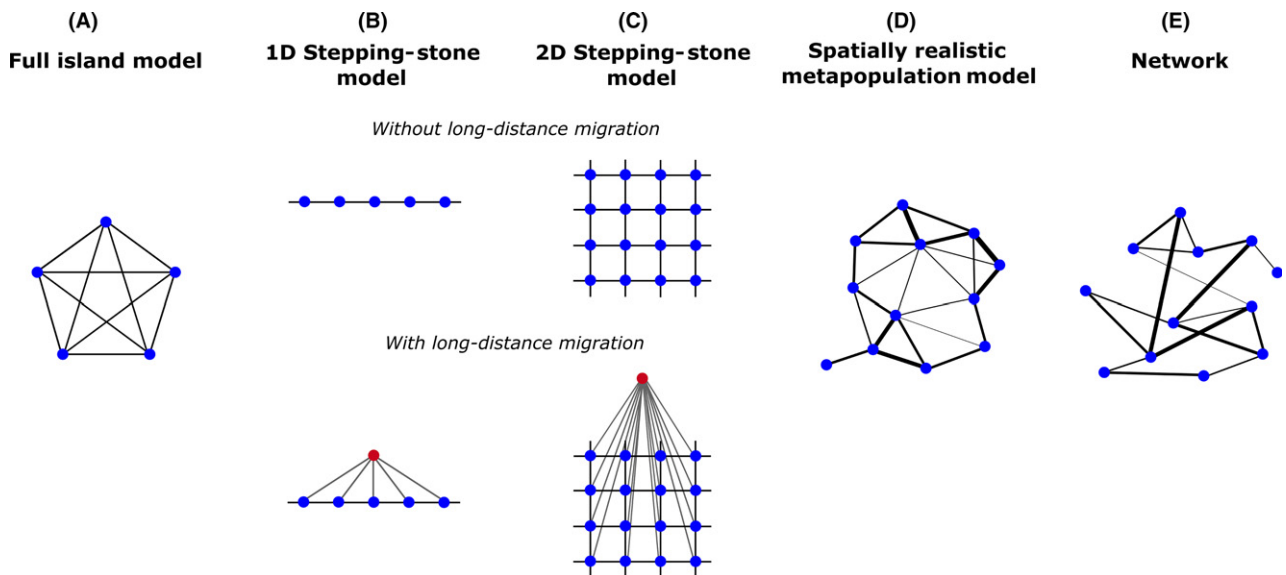


Fig. 1 Schematic depiction of discrete-habitat gene-flow models. Blue nodes represent habitat patches; black lines represent migration corridors, and line thickness represents gene-flow rate. In B and C, the red nodes represent the *point at infinity*, through which long-distance migration travels, and the grey lines represent the long-distance migration rates. In A, gene-flow rates are identical between all patches; in B and C, gene-flow occurs between adjacent patches and long-distance gene-flow occurs through the point at infinity; in D, patches may be positioned arbitrarily in two-dimensional space but gene-flow is constrained to follow distance-decreasing functions; in E, topology is unconstrained. Models A, B, C and D are special cases of the general network model, E.

Cavalli-Sforza 1968), and recently population geneticists and ecologists have started to realize that networks, where nodes represent discrete habitable patches and edges represent gene flows, may be used to account for contextually responsive gene-flow patterns between patches (Smouse 2000; Dyer & Nason 2004; Rozenfeld *et al.* 2008; Wagner & Fortin 2012; Dyer 2015). The previously described discrete models – such as the island model, stepping-stone models and SRMM – are particular cases of networks, but their interpretation may be constrained by the specific topology they consider. It is when gene-flow patterns cannot be assumed to be simple or ‘spatially realistic’ enough to be embedded into low-dimensional spaces (see Theoretical considerations in network modelling of discrete habitats section below) that network modelling is expected to be particularly important and provide novel insights. Network modelling can, in some sense, release population genetics from the necessity to devise specific models for specific topologies and allow us to move towards a more general framework.

In this paper, we will review the main branches of network theory that have been applied, or that we believe could potentially be applied, to population genetics and molecular ecology. The goal is to provide an overview of the implementation of these methodologies to population genetics in discrete habitats and to highlight the directions where network theory can further contribute to this research. We focus both on theoretical modelling and applied tools for empirical system-specific analyses.

Discrete habitats as networks

Network science is concerned with the study of mathematical constructs called *networks* or *graphs*. Networks are composed of discrete elements, called *nodes*, and

interactions or connections between these elements, called *edges*. It is often convenient to describe this construct using an *adjacency matrix*, where each matrix element describes the interaction between two nodes. Gene-flow patterns in discrete patchy habitats can therefore be naturally described by a network with habitat patches as nodes and migration corridors or gene-flow rates between the patches as edges (Fig. 2).

Theoretical considerations in network modelling of discrete habitats

Classic discrete models, such as the island and stepping-stone models, have looked into simple interaction-patterns between discrete habitat patches, and SRMM have looked into distance-dependent interactions. Although these models address habitats as discrete elements, they are mostly studied using continuous mathematical methods (e.g. Kimura & Weiss 1964; Hanski *et al.* 2011). The use of continuous methods to analyse discrete models can be justified when the underlying structure of the models can be topologically embedded into simple, continuous spaces (Note: an embedding of a space into another space is akin to positioning one space into another while preserving the relations between the different points in the original space). For example, the 1D stepping-stone model without long-distance migration, while being a discrete model, can be embedded into the continuous real line by identifying each patch (and therefore its location) with an integer, while preserving the order of the patches; this is an embedding of a discrete space into a continuous space. A similar embedding can be done with the 2D stepping-stone model without long-distance migration into the Euclidian plane. However, the 1D stepping-stone model with long-distance migration cannot be embedded into the real line as this model contains a node

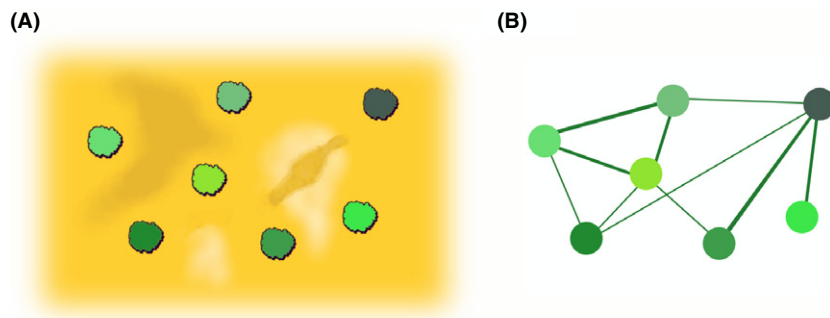


Fig. 2 Network description of a discrete habitat. (A) Discrete-habitat patches (green) in an inhabitable but traversable matrix (yellow). Both the habitats and matrix are heterogeneous, shown by different colour shades. (B) Network representation of the population, with nodes as habitat patches and edges as migration rates/corridors. Selection pressures in the different patches are shown as colour shades, and gene-flow rates are shown as edge thickness.

which is equally distant to every other node (the red node in Fig. 1B), while no point in the continuous real line shares this property. This discrete topology can still be embedded into a different simple continuous space, namely the real line with a *point at infinity* added to it (in mathematical topology, this is the ‘one-point compactification’ of the real line; Alexandroff 1924). In a similar manner, the 2D stepping-stone model with migration cannot be embedded into the Euclidian plane, but can be embedded into the Euclidian plane with a point at infinity. While the SRMM offer much more flexibility in the relative positioning of patches, the assumption that interpatch gene flows follow a distance-decreasing function precisely means that these models can be embedded into a two-dimensional Euclidean space; this is the ‘spatially realistic’ element of these models.

However, for many natural populations occupying discrete habitats, geographic distance is only one of several, often more significant, factors that shape gene-flow patterns; therefore, complex gene flow should be expected to be relatively common in nature. The topologies of the stepping-stone models and the SRMM can be described by networks (Fig. 1B–D) but, in general, networks (such as in Fig. 1E) cannot be embedded into simple continuous spaces. When such embeddings are not possible, it is crucial to account explicitly for spatial discreteness, as relying on continuous methods to describe spatial interactions may result in erroneous conclusions (Durrett & Levin 1994; Bascompte & Sole 1995; Shnerb *et al.* 2000). Therefore, it is becoming increasingly clear that we need to model discrete habitats with discrete methods (Urban *et al.* 2001, 2009; Butts 2009; Pocock *et al.* 2012; Cavanaugh *et al.* 2014). By framing evolutionary dynamics of discrete habitats in network theory terminology, both theoretical questions and empirical studies may benefit from new perspectives and existing tools.

Constructing habitat patch networks from empirical data

In network science, practical tools, concepts and methods have been developed to analyse complex empirical networks, for example linking patterns of neural processing across systems in the brain (Kinnison *et al.* 2012), or finding vulnerabilities that can lead to blackouts from failures in the power grid (Albert *et al.* 2004). The application of network theory to molecular ecology has, thus far, been predominantly to analyse empirical data, particularly in the context of landscape genetics (Dyer 2015; see also Table S1, Supporting information).

For such analyses, the first step is always to define the nodes and edges. Delineating nodes is often straightforward, as it should be congruent to the

ecological definition of the habitat patches. Edges represent gene-flow rates, and are more difficult to derive from empirical data. While edges could be approximated from nongenetic data, for example by combining mark–recapture methods, movement data, behavioural observations, life histories and landscape analyses to deduce migration and reproduction rates, gene-flow rates are most often inferred from interpopulation genetic-distance measures. Several measures have already been employed to construct networks, depending on the type of genetic data available (see Table S1, Supporting information): F_{ST} (Weir & Cockerham 1984), nucleotide distance (Tajima 1983), Goldstein distance (Goldstein *et al.* 1995), Jensen-Shannon divergence (Masucc *et al.* 2012) and the squared Euclidian distance between centroids as used in AMOVA (Dyer & Nason 2004). There has also been utilization of outputs of analyses such as principle component analysis (PCA) or ADMIXTURE (Alexander *et al.* 2009) to construct habitat patch networks (Paschou *et al.* 2014).

These procedures usually result in very dense pairwise matrices; that is, almost every pairwise relation between habitat patches is characterized by a nonzero value. To derive sparser networks, which are more workable, edge-inclusion criteria (or ‘edge pruning’) are typically used, targeting only the stronger or more informative relations. While the problem of determining how to formulate such criteria is currently being studied in general network science (Serrano *et al.* 2009; Radicchi *et al.* 2011; Dianati 2016), several criteria have been suggested and applied in the population-genetic literature: conditional independence (where we remove edges representing partial correlation coefficients that are small enough to imply that other edges in the network are sufficient to explain the total genetic covariance. Implemented in the software package Popgraph; Dyer & Nason 2004); the threshold below which the network becomes disconnected known as the percolation threshold (Rozenfeld *et al.* 2008); and the threshold resulting in the network with highest modularity (see communities in networks section below) value (Kininmonth *et al.* 2010). Of these, arguably, conditional independence is more appropriate for most applications, as the modularity criterion is suitable only for questions regarding the modular structure of the network, and the interpretation of the percolation threshold in these networks is relevant only if the question relates to the potential of an allele present in one network component to reach another component (see percolation and diffusion section below). However, as edge-inclusion may affect inferences made from network analyses, particularly the scale for which structure is maintained in the network after edges are removed (Serrano *et al.* 2009), perhaps a more conservative approach would be to test

several edge-inclusion thresholds or criteria, and develop methods to synthesize and interpret the results of several analyses. This has been implemented in other network-based population-genetic analyses (Greenbaum *et al.* 2016), and the topological field of persistent homology (Edelsbrunner & Harer 2008) may provide the relevant mathematical framework for developing such an approach.

Addressing evolutionary dynamics in discrete habitats using networks

Networks may represent a variety of systems, from genes to populations, from brain neurons to the Internet, and network theory has therefore branched out, with different branches addressing different issues related to network topologies and processes. In this section, we will review those branches that have been or may be applied to population-genetic questions in discrete habitats. Table 1 summarizes the main network terms in this review, along with their population-genetic equivalents and key literature. The Table S1, Supporting information lists studies that have developed or implemented network methodologies to study population genetics in the context of discrete habitat patches. Box 1 and Box 2 present a concrete example of a network formulation of evolutionary dynamics in habitat patch networks with heterogeneous selection.

Centrality measures

In many network applications, the major interest is in identifying nodes that are of importance for a particular function or process. A variety of centrality measures have been developed to evaluate different aspects of nodes' importance: *degree centrality* for local importance in the node's neighbourhood; *betweenness*, *flow betweenness* and *random-walk betweenness* for centrality in flow or diffusion processes; *closeness* and *random-walk closeness* for speed of information transfer; *eigenvector centrality* for the influence of a node on the network (see Newman 2010 for definitions of these and other measures, and Landherr *et al.* 2010 for a critical review). These measures assign a numeric value to each node, allowing ranking of nodes according to their centrality, making identification of the most central nodes, for a particular process, possible. The different measures not only address different functions of the network, but they employ different assumptions on the behaviour of the studied function.

In molecular ecology, centrality measures can be used, particularly in empirical population studies, to investigate the importance of certain habitat patches to different evolutionary processes. In many ecological

studies, and particularly in conservation, it may be important to know which habitats and which migration corridors are essential for maintenance of genetic diversity in the population as a whole, which constitute major pathways for gene flow, and which affect potential for local adaptation. Network analysis has been applied to identify central habitat patches in many systems and taxa, including mammals (Garroway *et al.* 2008, 2011; Ball *et al.* 2010; Creech *et al.* 2014; Fiset *et al.* 2015), frogs (Munwes *et al.* 2010; Naujokaitis-Lewis *et al.* 2013), invertebrates (Janes *et al.* 2014; Triponez *et al.* 2015), seagrass (Rozenfeld *et al.* 2008), trees (Richards *et al.* 2009; Herrera-Arroyo *et al.* 2013) and annual plants (Sexton *et al.* 2016). In these studies degree, betweenness, and eigenvector centralities, and occasionally closeness and flow centralities, have been used to locate patches central to gene flow in the network, highlighting subpopulations of ecological or conservation interest.

Although many centrality measures have been applied in such studies, not much attention has been given to the different functions these measures address. The appropriate measure to be used in a particular study should depend mainly on the ecological or evolutionary question of interest. Degree centrality (at the level of the node) can be used as a measure of connectivity of a patch (the number of patches the patch is connected to), but it only quantifies centrality at the local scale, and may be misleading for more global questions, such as for conservation prioritization. A more global centrality measure is eigenvector centrality, which quantifies the influence of nodes on the entire network by considering not only the connection of the nodes but also the centrality of the nodes to which it is connected (more precisely, it ranks nodes based on the attraction of random walks at the stationary state). This centrality measure can perhaps be useful for questions regarding local adaptation and migration load, particularly in theoretical and modelling studies which involve selection.

Most often, when applying centrality measures to empirical habitat patch networks, interest lies in identifying patches essential for gene flow in the network as a whole; the class of 'flow measures' is more likely to be correlated with gene flow than other classes. Many studies have applied *betweenness*, the most commonly used flow centrality measure, which counts the number of shortest paths connecting the rest of the network that pass through each given node. However, this measure is designed for studying information that is supervised and spreads only along shortest paths in the network, as is most often the case in social or computer networks where information is consciously directed. As gene flow in natural populations is not confined to shortest paths, nor is it consciously supervised, this measure, as well

Table 1 Main network terms mentioned in this paper

Network term	Network description	Example of ecological/ evolutionary interpretation	Relevant papers, where appropriate (Network/Biology)
Node (vertex)	Discrete entity	Habitat patch/deme	—
Edge (link)	Connections/interactions between nodes	Migration rate/migration corridor	—
Network (graph)	A set of nodes with connecting edges	Patch networks/mosaics	Newman 2010; /Pascual-Hortal & Saura (2006), Rozenfeld <i>et al.</i> (2008), Urban <i>et al.</i> (2009), Baranyi <i>et al.</i> (2011)
Weighted network	A network with numerical weights assigned to each of the edges	Patch networks with varying levels of gene flow between patches	Newman (2010)/this paper
Directed network	A network where edges have a direction from one node to the other	Patch networks with asymmetric gene flow between patches	Newman (2010)/Morrissey & de Kerckhove (2009)
Adjacency matrix	A matrix describing a network, in which unconnected nodes are set to 0 and connected nodes are set to 1 (or the weight of their connecting edge)	—	Newman (2010)/—
Node degree	The number of edges (or sum of edges weights) connected to the node	Connectivity of a patch	Newman (2010)/Estrada & Bodin (2008)
Centrality	A class of measures used to assign values to nodes according to their position in the network structure. There are various types of centrality measures used to capture different types of values/structures	Patches of particular importance for particular processes, for example gene flow, disease spread, local adaptation	Landherr <i>et al.</i> (2010); Newman (2010)/Rozenfeld <i>et al.</i> (2008)
Community	A group of nodes that are densely connected within the group and sparsely connected to nodes outside the group	Subpopulation/region	Girvan & Newman (2002); Newman & Girvan (2004)/Fortuna <i>et al.</i> (2009), Fletcher <i>et al.</i> (2013)
Component (maximal)	A group of nodes connected between themselves and not to any other node in the network	A maximal connected set of patches	Newman (2010)/Holstein <i>et al.</i> (2014)
Diffusion process	A process that spreads between adjacent nodes	Spread of alleles or mutations through a patch network	Newman (2002a)/Thomas <i>et al.</i> (2012); Neuwald & Templeton (2013)
Percolation threshold	A threshold (e.g. for the mean degree) above which a diffusion process may cover a significant portion of the network	A theoretical threshold for patch connectivity above which alleles or mutation are expected to spread to many patches	Broadbent & Hammersley (1957); Cohen & Havlin (2010)/—
Multilayer	A generalization of a network which consists of several layers, each of which is itself a network of nodes and intralayer edges, with an additional set of interlayer edges connecting nodes in different layers	Hierarchical population structure	Kivelä <i>et al.</i> (2014)/—
Multiplex	A multilayer where all layers contain the same network	Populations with complex life histories, epistasis	Kivelä <i>et al.</i> (2014)/—
Hypergraph	A generalization of a network, where instead of edges connecting pairs of nodes, there are hyperedges connecting arbitrary many nodes	Ecological Genomics	Berge & Minieka (1973)/Weighill & Jacobson (2015)

Box 1. Modelling selection-migration dynamics in networks

Here, we will consider discrete habitats with complex gene-flow patterns and heterogeneous selection pressures in the different habitat patches (see Fig. 2). The representation of discrete habitats as network adjacency matrices is convenient for formulizing evolutionary dynamics. We will present a simple model for selection-migration dynamics using a network with self-loops (edges that connect a node to itself).

The model describes n habitats, where in each habitat the population experiences different selection pressures. Gene flow between patches i and j is given by migration rate m_{ij} , representing the proportion of the population at patch i replaced by migrants arriving from patch j . Self-loops (i.e. the diagonal of the migration matrix) represent the proportion of the populations that do not migrate, that is $m_{ii} = 1 - \sum m_{ij}$. The model assumes a haploid population and tracks the frequency of one specific allele, f_i^t . The effect of selection on a patch with selection coefficient s and frequency f is given by a function g describing the change of frequency in one generation, $f^{t+1} = g(s, f^t)$ (e.g. for soft selection $g(s, f^t) = \frac{f^t(s+1)}{f^t s + 1}$; Wallace 1975). Selection is assumed to occur immediately after dispersal. With these definitions, we can now specify the recursion describing our model:

$$f_i^{t+1} = \sum_{j=1}^n m_{ij} g(s_j, f_j^t) \tag{eqn 1}$$

Each term in the summation represents the frequency change attributed to individuals arriving from patch j (or individuals staying in patch i for the case of $j = i$), with each migrant group arriving from patch j weighted in proportion to m_{ij} . The network m_{ij} , along with the selection coefficients s_i , fully describes the model.

This model can be generalized to address temporal changes in selection and migration by having s and m_{ij} contain functions rather than fixed values. It can also address life-history traits that affect migration and selection by having g depend on selection coefficients in both the source and the sink nodes of migration edges, in proportion to the time spent in each, as well as the cost of migrating between patches. Box 2 shows an elaboration on the formulization of this model.

other ‘supervised’ flow measures, is less than optimal. Perhaps the most appropriate centrality measure in respect to gene flow is the *random-walk betweenness* (Newman 2005), which quantifies the amount of random walks between nodes that pass through each node. This measure better reflects the unsupervised and stochastic nature of gene flow between habitat patches.

Communities in networks

Many networks in nature have a modular structure, where certain groups of nodes are more connected among themselves than to other nodes in the network. Such substructures within a network are known as modules or *communities* (the name stems from human social networks, where such substructures are interpreted as social communities). While there is no single definition for a community, it is roughly thought of as a dense subnetwork (subgraph) within a network. The study of communities in networks has gained considerable momentum with the development of the *modularity* measure (Girvan & Newman 2002; Newman 2004b; Newman & Girvan 2004) which denotes quality values to community partitions of the networks. The basic idea for modularity evaluation is to compare the

intracommunity densities (number and weights of edges) with what would be expected in a random network with the same node degrees. This allows quantification of how more (or less) modular the network is from a similarly structured random network. The measure of modularity, as well as other techniques that have since been developed, allows for computationally efficient detection of communities even in large and complex networks (reviewed in Fortunato 2010).

Many discrete habitats in nature are also modular, and gene-flow patterns in such habitats often have hierarchical and modular genetic structures (e.g. McCauley & Eanes 1987; Fletcher *et al.* 2013; Viricel & Rosel 2014; Pisa *et al.* 2015). Such hierarchical structures are often difficult to observe, especially when the underlying migration patterns are complex. Traditionally, such structures are detected using the F-statistics framework (Wright 1950), but here only *a priori*, putative, structures can be tested, and complexity of migration is not accounted for. Therefore, some studies have adopted a network framework and utilized community detection procedures to detect structure at various hierarchical levels, whether by looking at networks of individuals (Cohen *et al.* 2013; Greenbaum *et al.* 2016) or networks of habitat patches (Fortuna *et al.* 2009; Kininmonth *et al.* 2010; Munwes *et al.* 2010; Albert *et al.* 2013; Fletcher *et al.* 2013; Peterman *et al.*

Box 2. Selection-migration dynamics and walks on networks

The selection-migration model in Box 1 describes the evolutionary dynamics in the patch network, but the recursive formulation does not provide much insight; moreover, it requires calculation of the dynamics in the entire system to follow dynamics in a single patch. A more tractable formulation can be achieved using the concept of random walks on networks. A walk on a graph (network) is a sequence of nodes such that each pair of adjacent nodes in the sequence is connected with an edge in the network (West 2001), thus capturing the notion of traveling on the network from node to node along the network's edges. The weight of a walk, in weighted networks, is defined as the product of the edges included in the walk (Newman 2004a).

One can partition the gene pool in node i at time t to genes arriving from different patches of origin (at time 0) and through different walks on the networks to node i . Suppose we follow a walk w of length t from node j to node i , where each edge describes the proportion of the population migrating along that edge or staying in the same patch in one generation. Each generation we will be following a smaller group of individuals as different parts of the group follow different edges, but the frequency of the allele in this small group we are following will be determined solely by the frequency at the patch of origin and by the selection experienced at habitat patches along the walks. The proportion of the gene pool determined by the genes 'following' this walk is the proportion of the original population that remains at the end of the walk, that is the weight of the walk, \bar{w} . The frequency of this part of the population is determined by the selection coefficients along the walk, $\vec{s}_w = (s_{t-1}, s_{t-2}, \dots, s_1)$, which we will denote as $G(\vec{s}_w, f_j^0) = g(s_{t-1}, g(s_{t-2}, g(\dots g(s_1, f_j^0) \dots)))$. As each genealogy in each walk is independent, the frequency at node i at time t is the sum over all relevant walks:

$$f_i^t = \sum_{w \in W_i^t} w G(s_w, f_w^0) \quad (\text{eqn 2})$$

where W_i^t is the set of all walks of length t terminating in node i , and f_w^0 is the frequency at time 0 in the first patch in walk w .

The main benefit of this reformulation is that it is no longer recursive, making it more explicit. This also allows teasing apart the contribution of patches, connections of patches or regions of the patch network to the allele frequency in any particular patch in a given time. For example, considering the walks in W_i^t that include a patch j (and the selection along these walks) allows for the evaluation of the effect patch j has on the allele frequency of another patch i , and this effect can be assessed for the equilibrium state by taking t to be large. How a particular edge or group of patches or connections affect the allele frequency can be addressed in a similar manner. This could be useful for evaluating the influence of specific patches, or groups of patches, on the potential of local adaptation and migration load on patches of interest, or to identify patches or migration corridors with high influence on local adaptation of other patches.

2016). Detected communities are interpreted as clusters of habitat patches that are more genetically similar between themselves than to other habitat patches. The detection of these structures is done without *a priori* partitioning or knowledge of the number of clusters present in the population, which is often advantageous in molecular ecology studies, where such information is unavailable, or when we do not want to confine the results to presumed subpopulations.

While the rather limited utilization of community procedures in molecular ecology has so far been mostly focused on revealing hierarchical structures in empirically sampled populations, theoretical studies aimed at understanding evolutionary dynamics in discrete habitat patches may benefit from adopting a network approach as well. There is currently an effort among network theorists to provide a more general

framework for studying modular networks (Fortunato 2010), particularly using generalized community models (Newman & Peixoto 2015; Zhang *et al.* 2016). Such models, adapted to the population-genetic context, may be used in the future to ask questions regarding the formation of complex modular structures through evolutionary dynamics, or about processes in modular habitat patch networks, such as spread of alleles or local adaptation.

Percolation and diffusion

Percolation theory in networks is concerned with the ability to find an available path through a network (i.e. if introduced on one side, what is the probability of being able to follow a path through the network to reach the other side successfully; Broadbent &

Hammersley 1957), which relies on understanding the formation of disconnected components in networks by different processes. Both the formulation of this question and the techniques used to address it have been approached in many ways.

One formulation of a percolation problem is concerned with the ability of networks to remain connected when nodes or edges are removed from the network. It has been observed that in many networks there is a sharp threshold, termed *percolation threshold*, above which the networks remains relatively intact with one component containing a significant proportion of the nodes in the network (this component is known as a *giant component*), and below which networks breaks down into many small disconnected components. This behaviour of a sharp transition between two regimes means that a system that depends on connectivity between different parts of the network may collapse rapidly from a functional state to a dysfunctional state, once the percolation threshold is breached. Such behaviour can be of severe consequence to populations with complex gene-flow patterns between habitat patches that depends on adequate gene flow; understanding where the percolation threshold lies may be crucial for conservation and management of such populations (Bascompte & Sole 1995; Cumming *et al.* 2010). The tailoring of this question to the narrower formulation of 'invasion percolation', which considers 'paths of least resistance in traversing the network' rather than the probability of an available path simply existing, may also be of use (Wilkinson & Willemsen 1983; Furuberg *et al.* 1988), as networks that are connected but are characterized by very high resistance between regions may also be of conservation concern. This has recently been addressed in landscape genetics (McRae *et al.* 2008; Schwartz *et al.* 2009; Kershnerbaum *et al.* 2014), but seemingly not in the context of patch networks.

A second formulation of a percolation problem that may prove useful in the study of evolutionary dynamics is concerned with the extent to which a diffusive process would spread through a network. This has been well studied in the context of epidemiology, where the probability and size of an outbreak is studied using percolation theory (Newman 2002a,b). Here, the percolation threshold delineates scenarios where an outbreak, infecting a large part of the population, may occur, and the probability and size of such occurrences is addressed. This epidemiological problem resembles the problem of spread of alleles, particularly novel mutations, in patch networks. Here, the probability and extent of reach of new mutations in a patch network depend on the structure of the network, the levels of gene flow and the selection pressures at the different habitat patches. While this problem is yet to be

addressed using network terminology, existing formulations in network epidemiology (a network describing epidemiologically relevant contacts between individuals) may prove extremely useful.

Although percolation addresses the probability of there being a path through a network and the emergence of a giant component, it does not address more nuanced questions of network-based flow that can be of significant interest when trying to anticipate dynamics in a network over time. Traditionally, continuous diffusion models are used to explore such dynamics (Crank 1975), and similar discrete network diffusion models have also been developed (e.g. Leskovec *et al.* 2007; López-Pintado 2008; Kasprzyk 2012). While our understanding of the general behaviour of diffusion processes in networks is still limited, models for studying these behaviours might prove useful to answer questions such as: What is the expected time for a successful, locally adaptive mutation to reach another environmentally similar habitat? If a mutation is introduced at a particular patch, after a set duration, what is its likely distribution throughout the entire network (i.e. how far will it have permeated)? What are the transient spread dynamics and the ultimate stationary distributions (if any) for a mutation spreading to new patches and continuing to circulate among patches where it is already present (see, for example, Neuwald & Templeton 2013, where the temporal diffusion of alleles in a patchy habitat of collared lizards is tracked)? If the new mutation is beneficial only in certain patches, how will the diffusion dynamics affect the potential for local adaptation in these patches?

Multilayers, multiplexes and hypergraphs

While the above techniques may already provide a diversity of relevant tools for purposes of exploring discrete systems, they have all been limited in their representations of those systems to traditional networks, that is a set of nodes and the edges that connect pairs of them. Many useful generalizations of this concept have also been developed, most falling under two frameworks: *multilayer networks* (Kivelä *et al.* 2014) and *hypergraphs* (Berge & Minieka 1973). Multilayer networks consist of several interconnected layers, each containing a regular network (nodes within the same layer may be connected by *intra-layer* edges, while nodes in different layers may be connected by *inter-layer* edges). A multilayer network where the nodes in the different layers are the same entities is called a *multiplex network* (Lee *et al.* 2012). This generalization allows extending the applicability of network theory to systems that contain more than one network, such as in explorations of competition among viral strains transmitted in networks

where the nodes are hosts (Darabi Sahneh & Scoglio 2014), or the resilience of passenger rerouting in air travel with multiple carrier networks and random system failures (Cardillo *et al.* 2013).

Multilayer networks can be applied to the study of evolutionary dynamics in discrete habitats in several ways. First, multilayer networks can be used to model hierarchical population structures, where gene-flow patterns at local and regional scales are governed by different dynamics. In this formulation, a local network, labelled as a layer, describes the gene-flow patterns between patches in one region and consists of patches and intralayer edges, and inter-regional dynamics are described by adding interlayer edges describing gene flow between patches from different regions. Different dynamics at different spatial scales may lead to different patterns in inter- vs. intra-edges, which will result in different patterns of degree distributions.

Multiplex networks may be used to explicitly model selection and migration in organisms with different migration and selection patterns and different life stages and/or sexes. A multiplex of discrete habitats would consist of several layers, where in each layer, each habitat patch is represented by a node. The layers represent the different types of life stages (and/or different sexes) of the organism, and life stage-specific migration patterns and selection pressures may result in different networks at the different layers. The edges that connect the nodes between layers (only nodes representing the same patch may be connected between layers in a multiplex network) describe the demographic rates of flow from one life stage to another. The allele frequencies at each level of the multiplex indicate the allele frequency of a given stage in a given patch. Such multiplex networks may be used to address questions of centrality, percolation, and diffusion while taking into account all components of the population.

Multiplex networks can also be used to model selection acting on a suite of interdependent alleles. The gene-flow dynamics at a given locus can be described by the dynamics on one layer, while the selection coefficients for each layer are described by a different vector (for example as shown in Box 1). Allele frequencies are determined in each time step by evaluating the interaction of alleles in all layers (loci), taking into account the allele frequencies to determine the probability of co-occurrence of the interacting alleles (selection pressures may differ given presence of other alleles in the population). If gene-flow patterns are identical for all alleles, then the different layers may be identical; however, in the case of alleles affecting, for example, dispersal, gene-flow patterns may be different (e.g. an allele affecting a long-range dispersal trait should be formulated as a different patch network than an allele not affecting dispersal).

While a multiplex network with several layers may be sufficient to model several interacting alleles, this interaction can often be very complex and involve many alleles. Hypergraphs are mathematical constructs similar to networks; they consist of nodes and *hyperedges*, where hyperedges may connect more than a pair of nodes (i.e. a hyperedge may connect triplets, quadruples of nodes; Berge & Minieka 1973). Hypergraphs are often used in the field of genomics to represent interactions among several alleles for particular functions (Sole & Pastor-Satorras 2006; Tian *et al.* 2009; Weighill & Jacobson 2015). Thus, a system comprised of a multiplex network, describing gene-flow dynamics in different alleles, coupled with a hypergraph describing the interaction of the different alleles, provides a general network model for addressing questions in ecological genomics (Savolainen *et al.* 2013; Landry & Aubin-Horth 2014).

Work on these network extensions is by no means exhausted and novel mathematical techniques in this area are a focus of many recent publications. However, while more complicated in their mathematical characterization, already each of the measures of interest mentioned above (e.g. centrality measures, communities, random walks, and percolation and diffusion) has all been extended to capture analogous features in hypergraphs, multiplex networks and multilayer networks (Bonacich *et al.* 2004; Bradde & Bianconi 2009; Gao *et al.* 2011; Cellai *et al.* 2013; De Domenico *et al.* 2013; Gomez *et al.* 2013; Lu & Peng 2013; Battiston *et al.* 2014; Boccaletti *et al.* 2014; Solé-Ribalta *et al.* 2014).

Conclusion

Network methods have so far been primarily applied to detect and analyse characteristics of gene-flow patterns in discrete habitats. This significantly expands our molecular ecology toolkit for examining population structure, for example by allowing us to identify central patches, central corridors and hierarchical structures. However, one of the main challenges of molecular ecology is to connect these patterns to evolutionary and ecological processes, not merely to observe them. Molecular ecology, and conservation genetics in particular, will be able to considerably move forward when we will be able to assert predictions regarding the evolutionary consequences of real-world population structures. In this aspect, network theory is yet to be exploited, and we believe it has much more to offer in terms of theoretical modelling. Using a network framework to design population-genetic models (for example as in Boxes 1 and 2), we might be able to get a better handle on evolutionary dynamics.

To understand the role of patch networks in evolutionary dynamics, we would first need to evaluate the typical gene-flow topologies found in natural systems. These topologies may, in some circumstances, be approximated fairly well by models such as the stepping-stone or SRMM, but when geographically-independent factors influence migration and gene flow (e.g. heterogeneous matrix, as in Fig. 2; migration corridors; anthropogenic influences, when organisms 'hitchhike' on human transport networks), neglecting the complexity of gene flow may be problematic. In this context, it is important to remember that the evolutionary effect of gene flow is related to both distance and quantity – low amounts of long-distance gene flow can have a much larger effect than large amounts of short-distance migration, as demonstrated by the stepping-stone model (Kimura & Weiss 1964). Therefore, even if geographic distance is the main factor influencing gene flow, other factors inducing long-distance migration cannot always be neglected when considering evolution, and a complex topology may need to be assumed even in such cases.

In many fields, important insights have been gained by examining the peculiar characteristics of the network description of the systems in those fields, and in molecular ecology, we are only beginning to describe network topologies. As more and more natural discrete-habitat gene-flow networks will be described and analysed, we will be better able to characterize gene-flow topologies and get a clearer idea of the underlying processes forming and acting upon such networks. This will be an important step in understanding the role of gene-flow complexity in evolutionary dynamics, and network modelling will be crucial at this stage. Particularly, as was pointed out earlier, knowing to what extent natural gene-flow networks are simple or 'spatially realistic' would help us understand whether, and in what circumstances, previous modelling efforts were realistic, and how important it will be to further develop the integration of network theory into population genetics and molecular ecology.

Acknowledgements

The authors would like to thank Alan R. Templeton, Shirli Bar-David, Dina M. Fonseca and anonymous reviewers for their insightful comments. The collaboration between the co-authors was made possible thanks to the Aharon and Ephraim Katzir Fellowship granted to GG by the Batsheva de Rothschild Fund.

References

Albert R, Albert R, Albert I *et al.* (2004) Structural vulnerability of the North American power grid. *Physical Review E*, **69**, 25103.

- Albert EM, Fortuna MA, Godoy JA, Bascompte J (2013) Assessing the robustness of networks of spatial genetic variation. *Ecology Letters*, **16**, 86–93.
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.
- Alexandroff P (1924) Über die Metrisation der im Kleinen kompakten topologischen Räume. *Mathematische Annalen*, **92**, 294–301.
- Allendorf FW, Luikart GH, Aitken SN (2012) *Conservation and the Genetics of Populations*. Wiley-Blackwell, West Sussex, UK.
- Ball MC, Finnegan L, Manseau M, Wilson P (2010) Integrating multiple analytical approaches to spatially delineate and characterize genetic population structure: an application to boreal caribou (*Rangifer tarandus caribou*) in central Canada. *Conservation Genetics*, **11**, 2131–2143.
- Barabási A-L, Albert RR (1999) Emergence of scaling in random networks. *Science*, **286**, 509–512.
- Baranyi G, Saura S, Podani J, Jordán F (2011) Contribution of habitat patches to network connectivity: Redundancy and uniqueness of topological indices. *Ecological Indicators*, **11**, 1301–1310.
- Bascompte J (2007) Networks in ecology. *Basic and Applied Ecology*, **8**, 485–490.
- Bascompte J, Sole RV (1995) Rethinking complexity - modeling spatiotemporal dynamics in ecology. *Trends in Ecology & Evolution*, **10**, 361–366.
- Battiston F, Nicosia V, Latora V (2014) Structural measures for multiplex networks. *Physical Review E*, **89**, 32804.
- Berge C, Minieka E (1973) *Graphs and Hypergraphs*. North-Holland Publishing Company, Amsterdam.
- Boccaletti S, Bianconi G, Criado R *et al.* (2014) The structure and dynamics of multilayer networks. *Physics Reports*, **544**, 1–122.
- Bodmer WF, Cavalli-Sforza LL (1968) A migration matrix model for the study of random genetic drift. *Genetics*, **59**, 565–592.
- Bolnick DI, Nosil P (2007) Natural selection in populations subject to a migration load. *Evolution*, **61**, 2229–2243.
- Bonach P, Holdren AC, Johnston M (2004) Hyper-edges and multidimensional centrality. *Social Networks*, **26**, 189–203.
- Bornholdt S, Heinz GS (Eds.) (2006) *Handbook of Graphs and Networks: From the Genome to the Internet*. John Wiley & Sons, Darmstadt, Germany.
- Bradde S, Bianconi G (2009) The percolation transition in correlated hypergraphs. *Journal of Statistical Mechanics: Theory and Experiment*, **7**, P07028.
- Broadbent SR, Hammersley JM (1957) Percolation processes: I. Crystals and mazes. *Mathematical Proceedings of the Cambridge Philosophical Society*, **53**, 629–641.
- Butts CT (2009) Revisiting the foundations of network analysis. *Science*, **325**, 414–416.
- Calvo-Armengol A, Zenou Y (2004) Social networks and crime decisions: The Role of Social structure in facilitating delinquent behaviour. *International Economic Review*, **45**, 939–958.
- Cardillo A, Zanin M, Gómez-Gardeñes J *et al.* (2013) Modeling the multi-layer nature of the European Air Transport Network: Resilience and passengers re-scheduling under random failures. *European Physical Journal: Special Topics*, **215**, 23–33.
- Carrington PJ, Scott J, Wasserman S (2005) *Models and Methods in Social Network Analysis*. Cambridge University Press, New York, New York.

- Cavanaugh K, Siegel D, Raimondi P, Alberto F (2014) Patch definition in metapopulation analysis: a graph theory approach to solve the mega-patch problem. *Ecology*, **95**, 316–328.
- Cellai D, López E, Zhou J, Gleeson JP, Bianconi G (2013) Percolation in multiplex networks with overlap. *Physical Review E*, **88**, 52811.
- Cohen R, Havlin S (2010) *Complex Networks: Structure, Robustness and Function*. Cambridge University Press, New York, New York.
- Cohen O, Barocas A, Geffen E (2013) Conflicting management policies for the Arabian wolf *Canis lupus arabs* in the Negev Desert: is this justified? *Oryx*, **47**, 228–236.
- Crank J (1975) *The Mathematics of Diffusion*. Clarendon press Oxford, Bristol, UK.
- Creech TG, Epps CW, Monello RJ, Wehausen JD (2014) Using network theory to prioritize management in a desert bighorn sheep metapopulation. *Landscape Ecology*, **29**, 605–619.
- Croft DP, James R, Krause J (2008) *Exploring Animal Social Networks*. Princeton University Press, Princeton, New Jersey.
- Cumming GS, Bodin O, Ernstson H, Elmqvist T (2010) Network analysis in conservation biogeography: Challenges and opportunities. *Diversity and Distributions*, **16**, 414–425.
- Darabi Sahneh F, Scoglio C (2014) Competitive epidemic spreading over arbitrary multilayer networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, **89**, 1–15.
- De Domenico M, Solé A, Gómez S, Arenas A (2013) Random walks on multiplex networks. arXiv preprint, arXiv:1306.0519.
- Dianati N (2016) Unwinding the hairball graph: Pruning algorithms for weighted complex networks. *Physical Review E*, **93**, 1–8.
- Durrett R, Levin SA (1994) The Importance of Being Discrete (and Spatial). *Theoretical Population Biology*, **46**, 363–394.
- Dyer RJ (2015) Population Graphs and Landscape Genetics. *Annual Review of Ecology, Evolution, and Systematics*, **46**, 327–342.
- Dyer RJ, Nason JD (2004) Population Graphs: the graph theoretic shape of genetic structure. *Molecular Ecology*, **13**, 1713–1727.
- Easley D, Kleinberg J (2010) *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, New York, New York.
- Edelsbrunner H, Harer J (2008) Persistent homology - a survey. In: *Contemporary Mathematics : Vol. 453. Surveys on Discrete and Computational Geometry: Twenty Years Later* (eds Goodman JE, Pach J, Pollack R), pp. 257–281. AMS, Providence, Rhode Island.
- Estrada E, Bodin Ö (2008) Using network centrality measures to manage landscape connectivity. *Ecological Applications*, **18**, 1810–1825.
- Fefferman N, Ng K (2007) How disease models in static networks can fail to approximate disease in dynamic networks. *Physical Review E*, **76**, 31919.
- Fiset J, Tessier N, Millien V, Lapointe FJ (2015) Phylogeographic structure of the white-footed mouse and the deer mouse, two lyme disease reservoir hosts in Québec. *PLoS ONE*, **10**, e0144112.
- Fletcher RJ, Revell A, Reichert BE *et al.* (2013) Network modularity reveals critical scales for connectivity in ecology and evolution. *Nature Communications*, **4**, 2572.
- Fortuna MA, Albaladejo RG, Fernández L, Aparicio A, Bascompte J (2009) Networks of spatial genetic variation across species. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 19044–19049.
- Fortunato S (2010) Community detection in graphs. *Physics Reports*, **486**, 75–174.
- Furuberg L, Feder J, Aharony A, Jøssang T (1988) Dynamics of invasion percolation. *Physical Review Letters*, **61**, 2117.
- Gao J, Buldyrev SV, Stanley HE, Havlin S (2011) Networks formed from interdependent networks. *Nature Physics*, **8**, 40–48.
- García-Ramos G, Kirkpatrick M (1997) Genetic models of adaptation and gene flow in peripheral populations. *Evolution*, **51**, 21–28.
- Garroway CJ, Bowman J, Carr D, Wilson PJ (2008) Applications of graph theory to landscape genetics. *Evolutionary Applications*, **1**, 620–630.
- Garroway CJ, Bowman J, Wilson PJ (2011) Using a genetic network to parameterize a landscape resistance surface for fishers, *Martes pennanti*. *Molecular Ecology*, **20**, 3978–3988.
- Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 7821–7826.
- Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 6723–6727.
- Gomez S, Diaz-Guilera A, Gomez-Gardeñes J *et al.* (2013) Diffusion dynamics on multiplex networks. *Physical Review Letters*, **110**, 28701.
- Greenbaum G, Hoffman O, Shalev O, Zelnik YR (2015) Complex networks in ecology. *Israel Journal of Ecology & Evolution*, **61**, 67–68.
- Greenbaum G, Templeton AR, Bar-David S (2016) Inference and analysis of population structure using genetic data and network theory. *Genetics*, **202**, 1299–1312.
- Greening BR, Fefferman NH (2014) Evolutionary significance of the role of family units in a broader social system. *Scientific Reports*, **4**, 3608.
- Haldane JBS (1930) A mathematical theory of natural and artificial selection. *Proceedings of the Cambridge Philosophical Society*, **26**, 230–220.
- Hanneman R, Riddle M (2005) *Introduction to Social Network Methods*. University of California, Riverside, Riverside, California.
- Hanski I, Heino M (2003) Metapopulation-level adaptation of insect host plant preference and extinction-colonization dynamics in heterogeneous landscapes. *Theoretical Population Biology*, **64**, 281–290.
- Hanski I, Mononen T, Ovaskainen O (2011) Eco-evolutionary metapopulation dynamics and the spatial scale of adaptation. *The American Naturalist*, **177**, 29–43.
- Heino M, Hanski I (2001) Evolution of migration rate in a spatially realistic metapopulation model. *The American Naturalist*, **157**, 495–511.
- Herrera-Arroyo ML, Sork VL, González-Rodríguez A *et al.* (2013) Seed-mediated connectivity among fragmented populations of *Quercus castanea* (Fagaceae) in a Mexican landscape. *American Journal of Botany*, **100**, 1663–1671.
- Hock K, Fefferman N (2011) Extending the role of social networks to study social organization and interaction structure of animal groups. *Annales Zoologici Fennici*, **48**, 365–370.

- Holstein DM, Paris CB, Mumby PJ (2014) Consistency and inconsistency in multispecies population network dynamics of coral reef ecosystems. *Marine Ecology Progress Series*, **499**, 1–18.
- Janes JK, Li Y, Keeling CI *et al.* (2014) How the mountain pine beetle (*Dendroctonus ponderosae*) breached the canadian rocky mountains. *Molecular Biology and Evolution*, **31**, 1803–1815.
- Kasprzyk R (2012) Diffusion in networks. *Journal of Telecommunications and Information Technology*, **2012**, 99–106.
- Kawecki T, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.
- Keeling MJ, Eames KTD (2005) Networks and epidemic models. *Journal of the Royal Society, Interface*, **2**, 295–307.
- Kershenbaum A, Blank L, Sinai I *et al.* (2014) Landscape influences on dispersal behaviour: a theoretical model and empirical test using the fire salamander, *Salamandra atra*. *Oecologia*, **175**, 509–520.
- Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, **49**, 561–576.
- Kininmonth S, van Oppen MJH, Possingham HP (2010) Determining the community structure of the coral *Seriatopora hystrix* from hydrodynamic and genetic networks. *Ecological Modelling*, **221**, 2870–2880.
- Kinnison J, Padmala S, Choi J-M, Pessoa L (2012) Network analysis reveals increased integration during emotional and motivational processing. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, **32**, 8361–8372.
- Kivelä M, Arenas A, Barthelemy M *et al.* (2014) Multilayer Networks. *Journal of Complex Networks*, **2**, 203–271.
- Landherr A, Friedl B, Heidemann J (2010) A critical review of centrality measures in social networks. *Business & Information Systems Engineering*, **2**, 371–385.
- Landry CR, Aubin-Horth N (2014) *Ecological Genomics: Ecology and the Evolution of Genes and Genomes*. Springer, Dordrecht, The Netherlands.
- Lee KM, Kim JY, Cho WK, Goh KI, Kim IM (2012) Correlated multiplexity and connectivity of multiplex random networks. *New Journal of Physics*, **14**, 1–14.
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, **17**, 183–189.
- Leskovec J, Adamic L, Huberman B (2007) The dynamics of viral marketing. *ACM Transactions on the Web*, **1**, 1–39.
- Levene H (1953) Genetic equilibrium when more than one ecological niche is available. *The American Naturalist*, **87**, 331–333.
- López-Pintado D (2008) Diffusion in complex social networks. *Games and Economic Behavior*, **62**, 573–590.
- Lordan O, Sallan JM, Simo P (2014) Study of the topology and robustness of airline route networks from the complex network approach: a survey and research agenda. *Journal of Transport Geography*, **37**, 112–120.
- Lu L, Peng X (2013) High-Order random walks and generalized Laplacians on hypergraphs. *Internet Mathematics*, **9**, 3–32.
- Masucc AP, Arnaud-Haond S, Eguíluz V, Hernández-García E, Serrão E (2012) Genetic flow directionality and geographical segregation in a *Cymodocea nodosa* genetic diversity network. *EPJ Data Science*, **1**, 1–11.
- Matousek J, Nešetřil J (1998) *Invitation to Discrete Mathematics*. Oxford University Press, Oxford, UK.
- McCauley DE, Eanes WF (1987) Hierarchical population structure analysis of the milkweed beetle, *Tetraopes tetraophthalmus* (Forster). *Heredity*, **58**, 193–201.
- McRae BH, Dickson BG, Keitt TH, Shah VB (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology*, **89**, 2712–2724.
- Morrissey MB, de Kerckhove DT (2009) The maintenance of genetic variation due to asymmetric gene flow in dendritic metapopulations. *The American Naturalist*, **174**, 875–889.
- Munwes I, Geffen E, Roll U *et al.* (2010) The change in genetic diversity down the core-edge gradient in the eastern spadefoot toad (*Pelobates syriacus*). *Molecular Ecology*, **19**, 2675–2689.
- Naujokaitis-Lewis IR, Rico Y, Lovell J, Fortin MJ, Murphy MA (2013) Implications of incomplete networks on estimation of landscape genetic connectivity. *Conservation Genetics*, **14**, 287–298.
- Neuwald JL, Templeton AR (2013) Genetic restoration in the eastern collared lizard under prescribed woodland burning. *Molecular Ecology*, **22**, 3666–3679.
- Newman MEJ (2002a) Spread of epidemic disease on networks. *Physical Review E*, **66**, 1–11.
- Newman MEJ (2002b) Random graphs as models of networks. In: *Handbook of Graphs and Networks: From the Genome to the Internet* (eds Bornholdt S, Heinz GS), pp. 34–68. John Wiley & Sons, Darmstadt, Germany.
- Newman MEJ (2004a) Analysis of weighted networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, **70**, 1–9.
- Newman MEJ (2004b) Detecting community structure in networks. *The European Physical Journal B*, **38**, 321–330.
- Newman MEJ (2005) A measure of betweenness centrality based on random walks. *Social Networks*, **27**, 39–54.
- Newman MEJ (2010) *Networks: An Introduction*. Oxford University Press, Oxford, UK.
- Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Physical Review E*, **69**, 26113.
- Newman MEJ, Peixoto TP (2015) Generalized communities in networks. *Physical Review Letters*, **115**, 88701.
- Paschou P, Drineas P, Yannaki E *et al.* (2014) Maritime route of colonization of Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 9211–9216.
- Pascual-Hortal L, Saura S (2006) Comparison and development of new graph-based landscape connectivity indices: towards the prioritization of habitat patches and corridors for conservation. *Landscape Ecology*, **21**, 959–967.
- Peterman WE, Ousterhout BH, Anderson TL *et al.* (2016) Assessing modularity in genetic networks to manage spatially structured metapopulations. *Ecosphere*, **7**, 1–16.
- Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, **8**, e1002967.
- Pisa G, Orioli V, Spilotros G *et al.* (2015) Detecting a hierarchical genetic population structure: the case study of the Fire Salamander (*Salamandra salamandra*) in Northern Italy. *Ecology and Evolution*, **5**, 743–758.
- Pocock MJO, Evans DM, Memmott J (2012) The robustness and restoration of a network of ecological networks. *Science*, **335**, 973–977.
- Proulx SR, Promislow DEL, Phillips PC (2005) Network thinking in ecology and evolution. *Trends in Ecology & Evolution*, **20**, 345–353.

- Radicchi F, Ramasco JJ, Fortunato S (2011) Information filtering in complex weighted networks. *Physical Review E*, **83**, 1–9.
- Richards CM, Volk GM, Reilley A A *et al.* (2009) Genetic diversity and population structure in *Malus sieversii*, a wild progenitor species of domesticated apple. *Tree Genetics and Genomes*, **5**, 339–347.
- Rozenfeld AF, Arnaud-Haond S, Hernández-García E *et al.* (2008) Network analysis identifies weak and strong links in a metapopulation system. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 18824–18829.
- Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature Reviews. Genetics*, **14**, 807–820.
- Schwartz MK, Copeland JP, Anderson NJ *et al.* (2009) Wolverine gene flow across a narrow climatic niche. *Ecology*, **90**, 3222–3232.
- Seiler M, Collins A, Fefferman N (2014) Strategic mortgage default in the context of a social network: an epidemiological approach. *Journal of Real Estate Research*, **35**, 445–475.
- Serrano MA, Boguñá M, Vespignani A (2009) Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 6483–6488.
- Sexton JP, Hufford MB, Bateman A *et al.* (2016) Climate structures genetic variation across a species' elevation range: a test of range limits hypotheses. *Molecular Ecology*, **25**, 911–928.
- Shnerb NM, Louzoun Y, Bettelheim E, Solomon S (2000) The importance of being discrete: life always wins on the surface. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 10322–10324.
- Smouse PE (2000) Reticulation inside the species boundary. *Journal of Classification*, **17**, 165–173.
- Sole RV, Pastor-Satorras R (2006) Complex networks in genomics and proteomics. In: *Handbook of Graphs and Networks: From the Genome to the Internet* (eds Bornholdt S, Schuster HG), pp. 145–167. John Wiley & Sons, Darmstadt, Germany.
- Solé-Ribalta A, De Domenico M, Gómez S, Arenas A (2014) Centrality rankings in multiplex networks. In: *Proceedings of the 2014 ACM Conference on Web Science*, pp. 149–155. ACM, New York, New York.
- Sterbenz JPG, Çetinkaya EK, Hameed MA *et al.* (2011) Evaluation of network resilience, survivability, and disruption tolerance: analysis, topology generation, simulation, and experimentation. *Telecommunication Systems*, **52**, 705–736.
- Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.
- Templeton A (2006) *Population Genetics and Microevolutionary Theory*. John Wiley & Sons, Hoboken, New Jersey.
- Thomas M, Demeulenaere E, Dawson JC *et al.* (2012) On-farm dynamic management of genetic diversity: The impact of seed diffusions and seed saving practices on a population-variety of bread wheat. *Evolutionary Applications*, **5**, 779–795.
- Tian Z, Hwang T, Kuang R (2009) A hypergraph-based learning algorithm for classifying gene expression and arrayCGH data with prior knowledge. *Bioinformatics*, **25**, 2831–2838.
- Triponez Y, Arrigo N, Espíndola A, Alvarez N (2015) Decoupled post-glacial history in mutualistic plant-insect interactions: Insights from the yellow loosestrife (*Lysimachia vulgaris*) and its associated oil-collecting bees (*Macropis europaea* and *M. fulvipes*). *Journal of Biogeography*, **42**, 630–640.
- Urban D, Keitt T, Ecology S (2001) Landscape connectivity: a graph-theoretic perspective. *Ecology*, **82**, 1205–1218.
- Urban DL, Minor ES, Treml EA, Schick RS (2009) Graph models of habitat mosaics. *Ecology Letters*, **12**, 260–273.
- Viricel A, Rosel PE (2014) Hierarchical population structure and habitat differences in a highly mobile marine species: The Atlantic spotted dolphin. *Molecular Ecology*, **23**, 5018–5035.
- Wagner HH, Fortin M-J (2012) A conceptual framework for the spatial analysis of landscape genetic data. *Conservation Genetics*, **14**, 253–261.
- Wallace B (1975) Hard and soft selection revisited. *Evolution*, **29**, 465–473.
- Wasserman S, Faust K (1994) *Social Network Analysis: Methods and Applications*. Cambridge Univ. Press, Cambridge, UK.
- Weighill DA, Jacobson DA (2015) 3-way networks: application of hypergraphs for modelling increased complexity in comparative genomics. *PLoS Computational Biology*, **11**, e1004079.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- West D (2001) *Introduction to Graph Theory*. Prentice Hall, Englewood Cliffs, New Jersey.
- Wilkinson D, Willemsen JF (1983) Invasion percolation: a new form of percolation theory. *Journal of Physics A: Mathematical and General*, **16**, 3365.
- Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.
- Wright S (1950) The genetical structure of populations. *The Annals of Human Genetics*, **15**, 323–354.
- Zhang P, Moore C, Newman MEJ (2016) Community detection in networks with unequal groups. *Physical Review E*, **93**, 1–15.

G.G. and N.H.F. both developed the concept and ideas in this review. G.G. primarily wrote the paper, and significant contributions were made by N.H.F. Both authors gave final approval for publication.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Partial list of papers explicitly using network approaches to describe or analyze population-genetics of populations in discrete habitats (in Alphabetical order).