A replicated network approach to "Big Data" in Ecology

Athen Ma[1,*] - athen.ma@qmul.ac.uk

David A. Bohan[2] - david.bohan@inra.fr

Elsa Canard[3] - elsa.canard@inra.fr

Stephane Derocles[2] - stephane.derocles@inra.fr

Clare Gray[4,5] – clare.gray@imperial.ac.uk

Xueke Lu[1,6] – xueke.lu@qmul.ac.uk

Sarina Macfadyen[7] - sarina.macfadyen@csiro.au

Gustavo Q. Romero[8] - gqromero@unicamp.br

Pavel Kratina[4,*] - p.kratina@qmul.ac.uk

*Corresponding authors

1. School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK
2. UMR 1347 Agroécologie. AgroSup/UB/INRA, Pôle GESTion durable des ADventices. 17 rue Sully, BP 86510 21065 Dijon CEDEX, France
3. INRA, Agrocampus-Ouest, Université de Rennes 1, UMR1099 IGEPP, Rennes Cedex, France
4. School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, UK
5. Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot, Berkshire SL5 7PY, UK
6. School of Engineering, The University of Warwick, Coventry CV4 7AL, UK
7. CSIRO, Black Mountain, Acton, 2601, Australia
8. Departamento de Biologia Animal, University of Campinas, Campinas, Brazil

SUMMARY

Global environmental change is a pressing issue as evidenced by the rise of extreme weather conditions in many parts of the world, threatening the survival of vulnerable species and habitats. Effective monitoring of climatic and anthropogenic impacts is therefore critical to safeguarding ecosystems, and it would allow us to better understand their response to stressors and predict long-term impacts. Ecological networks provide a biomonitoring framework for examining the system-level response and functioning of an ecosystem, but have been, until recently, constrained by limited empirical data due to the laborious nature of their construction. Hence, most experimental designs have been confined to a single network or a small number of replicate networks, resulting in statistical uncertainty, low resolution, limited spatio-temporal scale and over-simplified assumptions.

Advances in data sampling and curation methodologies, such as Next Generation Sequencing (NGS) and the Internet "Cloud", have facilitated the emergence of the "Big Data" phenomenon in Ecology, enabling the construction of ecological networks to be carried out effectively and efficiently. This provides to ecologists an excellent opportunity to expand the way they study ecological networks. In particular, highly replicated networks are now within our grasp if new NGS technologies are combined with machine learning to develop network building methods. A replicated network approach will allow temporal and spatial variations embedded in the data to be taken into consideration, overcoming the limitations in the current "single network" approach.

We are still at the embryonic stage in exploring replicated networks, and with these new opportunities we also face new challenges. In this paper, we discuss some of these challenges and highlight potential approaches that will help us build and analyse replicated networks to better understand how complex ecosystems operate, and the services and functioning they provide; paving the way for deciphering ecological big data reliably in the future.

I.   INTRODUCTION – A NEED TO DETECT ECOSYSTEM CHANGE

Large scale monitoring of ecosystems has become ever more important in the face of global environmental change driven by the activity of man and changing climate. Current approaches to large scale monitoring include the use of proxies or indicators, such as chemical indicators to evaluate the level of environmental pollution risks (e.g. Arshad and Martin, 2002; Schoenholtz et al., 2000) or biotic

64   measures for assessing the condition of an ecosystem (e.g. Dale and Beyeler, 2001;
65   Davies and Jackson, 2006). In the past 30 years, satellite- or aircraft-based remote
66   sensing techniques have also led to radical changes to our understanding of the
67   planet's environments, allowing large scale and long-term processes, such as
68   desertification, algal blooms and deforestation, to be followed (e.g. Lawton et al.
69   2001; Perennou et al. 2018; van der Werf et al. 2008). However, these two approaches
70   cover the ecological extremes of what we would like to monitor. At one extreme,
71   classical ecological biomonitoring is often limited to an individual species in a
72   specific ecosystem; the data are therefore difficult to extrapolate to complex systems
73   over large spatial scales. At the other extreme, remote sensing is explicitly large
74   scale, covering many ecosystems and biomes simultaneously, but these data often
75   lack ecological details and mechanisms that govern patterns of change.

76   Accelerating global and local changes prompt an urgent need to better
77   understand responses of entire ecological networks and to predict the effects of
78   future perturbations (Gray et al., 2014; Kratina et al. 2014). There are structural
79   properties in ecological networks that reveal their underlying organisational
80   principle and evolutionary and compensatory mechanisms in response to
81   disturbance (Cohen 1977; Woodward et al., 2010; Lu et al., 2016). Some of these
82   structures can be further linked to ecological functioning and services provided by
83   an ecosystem (Thompson et al., 2012). Moreover, ecological network structures can
84   be related to the long-term dynamics of an ecosystem, such as resilience and
85   robustness (Dunne et al. 2002; Memmott et al. 2004, Ledger et al. 2013, Oliver et al.
86   2015). Network approaches to large scale monitoring have the potential to provide
87   forewarning of ecosystem degradation more informatively than existing
88   biomonitoring and remote sensing (e.g. Dakos and Bascompte 2014, Jiang et al.
89   2018).

90   In this middle ground between the ecological extremes of current
91   biomonitoring and remote sensing approaches, a network-based approach to large
92   scale, generic monitoring of ecosystems would be a valuable tool. We argue in
93   support of networks as a generic analytical approach that can be applied in any
94   domain of ecological research. Ecological networks can be examined for any
95   ecosystem with little modification of either construction or analytical methodology.
96   New molecular methods and machine learning approaches to sampling species are
97   starting to deliver the volumes of data necessary to monitor ecosystems. However,
98   analytical methods that can be applied to such highly replicated network big data
99   are little developed and their application lags behind the empirical advances. Here,
100   we summarize the historic empirical research on ecological networks and the

101   common analytical approaches used to analyse them. We highlight their
102   assumptions and shortcomings in the face of emerging "big" spatially and
103   temporarily replicated network data. Finally, we review new promising avenues for
104   future analyses, including the application of food web profiling, null models, and
105   weighted and multilayer networks. Overall we aim to summarise and highlight the
106   new approaches to replicate big data in network ecology and their use in evaluating
107   the changes and variation in network structure in order to detect and gauge
108   environmental impact.

109

A. The problem – Replication

111   The majority of past studies on ecological networks have been based on
112   unreplicated sampling designs mainly because of the excessive research time and
113   effort required to assemble complex, highly resolved networks. These "single
114   network" approaches have been used to study the structure and functioning of
115   ecosystems, and much of our understanding of network topology has been
116   characterised from such "snapshot" data (Table 1). Networks were constructed from
117   data aggregated across multiple time points and multiple sites. For example, Evans
118   et al. (2013) quantified a network of multiple ecosystem services from a single
119   organic farm in the UK. The farm network was comprised of 1502 unique
120   interactions, and 560 taxa of plants and animals (Pocock et al., 2012). However,
121   aggregation of data based on the presence or absence of species and expert
122   knowledge of interactions can link taxa that do not occur at the same time and space
123   and does not account for the fact that species pairs are located on a continuum of
124   very weak to very strong interactions. Moreover, the assembly of a local community
125   is often driven by local environmental conditions, species states, historical
126   contingencies and stochastic assembly processes (Poisot et al., 2015). This means that
127   an ecosystem can have multiple equilibrium states, with regional species pool
128   (Ricklefs, 1987) and a diversity of interactions varying in time and space.
129   Consequently, species co-occurrences in aggregated data do not always reveal an
130   interaction between two species (Olesen et al., 2011; Freilich et al. 2018) due to
131   differences in phenology (Vázquez, 2005), ontogeny, consumer-resource body size
132   ratios (Tsai et al., 2016) or low encounter rate (Canard et al., 2012). Some interactions
133   may also be dependent on the presence of a third species or interaction between
134   other species (Golubski and Abrams, 2011; Poisot et al., 2011), or be limited to certain
135   environmental conditions (Poisot et al., 2011).

136   Replication of ecological networks is becoming more common and it has the
137   potential to improve the understanding of the magnitude of community changes

138 following environmental perturbations (Tylianakis and Morris, 2017). There are
139 recent examples of where replication has been used. For example, Tylianakis et al.
140 (2007) pioneered the field by constructing 48 replicated networks along a gradient of
141 habitat modification (from agricultural habitats to forests), and assessed how the
142 disturbance and modification of habitats by humans affected the structure of species
143 interactions as well as ecosystem function. They showed that species richness
144 remained constant, but the vulnerability of the networks decreased while
145 community evenness increased along the gradient of habitat modification. Morris et
146 al. (2015) used a similar approach to construct 20 replicated networks of cavity-
147 nesting Hymenoptera and their parasitoids and kleptoparasites along an
148 environmental gradient to demonstrate the effect of elevation on network structures.
149 Dézerald et al. (2013) constructed 365 independent networks of metazoans
150 inhabiting natural microecosystems (bromeliads) in tropical forests in French
151 Guiana, and showed that key metrics, such as linkage density and nestedness, varied
152 across a gradient of canopy openness. Nielsen and Totland (2014) showed that the
153 structural properties of mutualistic plant-pollinator networks in heavily managed
154 boreal forests were conserved after habitat degradation, highlighting the resilience of
155 species interactions to external disturbance.

156   Construction of replicated ecological networks is often constrained by the
157 sampling effort required, as they are currently assembled in the same way as non-
158 replicated networks, but with replication across a range of environmental conditions
159 (Morris et al., 2015; Nielsen and Totland, 2014; Tylianakis et al., 2007). This is highly
160 resources costly and laborious. However, advances in computing, storage and
161 processing that have occurred alongside the expansion of the Internet, and machine
162 learning approaches will increasingly result in more "big data" becoming available
163 (see Buttigieg et al., 2018). Big data is the 'catch-all' name given to very large
164 databases that are accessed and analysed using bioinformatics approaches. In some
165 cases, the analysis of biological big data has taken the form of simple data mining,
166 looking for obvious correlations between biological variables. More recently,
167 sophisticated approaches to hypothesising and testing relationships from big data
168 have emerged. For example, statistical and logical machine learning approaches
169 have been used to build metagenomic and metabolomics interaction networks to
170 identify genes or gene products associated with cancer or cell growth (e.g. Kourou et
171 al., 2015). These techniques have also been applied to build ecological networks (e.g.
172 Bohan et al., 2011b; Tamaddoni-Nezhad et al., 2013; Vacher et al. 2016), both from
173 classical ecological sample data and molecular ecological data, and will pave the
174 way for generating replicated networks more quickly and economically.

175   Researchers have traditionally focussed on resolving networks that are
176   relevant to the delivery of ecosystem services (e.g. pest control, pollination, or
177   nutrient cycling). In many communities, highly abundant species are key to the
178   delivery of a service, and as a result, less attention has been paid to documenting
179   rare species and their interactions. These rare species can fulfil essential ecosystem
180   functions due to their rare trait combination (Mouillot et al. 2013). As we move
181   towards more sophisticated approaches for gathering large amounts of empirical
182   data (e.g. sequencing of environmental samples), it is envisioned that there will be
183   higher levels of variation present in the data as more rare species and their
184   interactions are recorded. With the emergence of big data in ecology, the research
185   challenge will be to process the vast quantities of data that are being produced and
186   to interpret their ecological significance (Woodward et al., 2014). This calls for more
187   effective analytical tools for examining replicated ecological networks (e.g. Poisot et
188   al., 2012) to complement the development and advances in ecological big data.

189

190   **Table 1** Definitions of common terms used in network analyses.

Glossary	Definition
Background noise	Random natural variation in a network, which leads to the detection of false positive connections. Replicated networks allow discriminating this background noise from the "real" differences between networks caused by an ecosystem disturbance.
Snapshot network	Single networks built on unreplicated sampling designs, often constructed from data aggregated across multiple time points and multiple sites.
Network profiling	Techniques that search for patterns in networks. These methods characterize the frequency distribution of individual elements (e.g. network motifs) in order to identify their causes and consequences.
Network backbone	Despite their typical link redundancy, networks seem to self-organize into robust, disassortative structures termed backbones. The backbone of a weighted network is composed from important individual nodes and their links and discriminates nodes and links (e.g. energy flux) that are insignificant in statistical sense.

Core	A cohesive structure of closely interconnected nodes in a network. Large cores increase system redundancy by buffering external disturbance.
Periphery	A structure of loosely connected nodes in a network.
Higher-level system properties	General patterns, dynamics and properties of an ecological community that can be predicted without understanding all the underlying ecological details.
Apparent competition	An increased abundance of one prey species negatively affects other prey species via increased abundance of shared predator.

191

B. A potential solution - Replicated networks through next generation sequencing

A replicated network approach would allow a better understanding of how community structures and ecosystem functions are interlinked, and we have already seen their advantages in a small number of studies. In 64 plant-pollinator networks, Kaiser-Bunbury et al. (2017) examined the reproductive performance of the ten most abundant plant species to describe pollination across four restored and four unrestored mountaintop communities. The study showed that the estimates of interaction specialisation (Blüthgen et al., 2006) were lower in restored networks than in unrestored networks, suggesting that pollinators mediate less specialised interactions in the restored environments. More importantly, the differences in network structure reflect direct positive effects on pollination through fruit production of native plants. There are other ecological functions and associated ecosystem services that are more complex to examine than parasitism or pollination, such as the case of weed regulation by carabid beetles (Bohan et al., 2011). This ecosystem service, as an example, relies on prey-predator interactions between weed seeds and carabid beetles that are very difficult to observe in nature and therefore would benefit from advanced molecular tools.

New molecular tools will provide an excellent opportunity to build ecological networks rapidly and efficiently, once methods are optimized and appropriate bioinformatics are developed and applied (Evans et al., 2016; Derocles et al. 2018). Most of the molecular methods available are based on high-throughput parallel sequencing, called "next-generation sequencing" (NGS), where a whole community from various kinds of field samples is screened and analysed (e.g. environmental

216 DNA obtained from samples of soil or water, mixture of cells or tissues from traps,
217 such as arthropod pitfall traps). DNA-based approaches do not rely on taxonomic
218 expertise and some of these techniques can produce millions of DNA sequences
219 within a relatively short time frame (Evans et al., 2016), and these approaches have
220 advantages of being much faster than existing biomonitoring methods (Ji et al.,
221 2013).

222   While researchers are starting to use NGS to construct replicated networks,
223 current lack of replication means that it is not possible to identify or filter out the
224 variation (noise) generated during the network assembly process, and this can
225 potentially lead to an inaccurate interpretation of the results. In particular, NGS
226 requires a bioinformatics pipeline for species identification (Kitson et al., 2016; Toju
227 et al., 2013). Because accuracy of the identification strongly depends on the quality of
228 DNA sequences and the completeness of the species database (Evans et al., 2016),
229 any sample contamination can adversely affect the results (Piñol et al., 2014).

230   As we move towards gathering multiple datasets for a replicated network
231 approach, utilising NGS will provide a means for generating big data and
232 constructing large numbers of networks more efficiently. The major advantage of
233 this is that it will be possible to discriminate natural variation encapsulated in a
234 network from the variation due to disturbance, both allowing the "real" differences
235 among the ecological networks to be evaluated alongside ecosystem responses to
236 perturbation. This is the first step to a next-generation of biomonitoring that allows
237 understanding and prediction of how environmental change affects ecological
238 networks and ecosystems through time and space.

239

240 ## II.   HISTORICAL PERSPECTIVE ON NETWORK ANALYSIS

241   A key objective in network ecology is to relate network structure to the
242 properties and dynamics of an ecological system so as to better understand the
243 underlying ecological meanings and implications of environmental disturbances. For
244 instance, changes in precipitation and temperature regimes alter the strength of
245 biotic interactions and reorganize the structure of a network (Blois et al., 2013;
246 Woodward et al., 2016). Flooding can also alter key network properties, reducing
247 taxon richness, food chain length, and increasing the proportion of species at basal
248 trohic level, with the degree of habitat isolation after flooding (McLaughlin et al.
249 (2013). Petchey et al. (2010) examined the effect of temperature on ecological
250 networks by incorporating the relative activation energies of attack rates and
251 handling times, and showed how diet breadth is linked to network connectance (i.e.

252 the number of realized links as fraction of all possible). Networks can also make
253 explicit indirect interactions that govern non-intuitive system behaviours (Rossberg
254 2013) and alter the whole ecosystems (e.g. Montoya et al., 2009). More mechanistic
255 understanding of networks is then essential for successful biodiversity management
256 and restoration.

257   Historically analyses have focused on simple networks, where all network
258 parameters have been measured, analysed analytically and simulated. These simple
259 networks have included isolated and remote ecosystems (e.g. Bear Island in the high
260 Arctic), human-simplified (e.g. crop-pest) networks with several dominant species or
261 networks that aggregated many species into several functional groups
262 (Summerhayes and Elton 1923). Big data has led to the recognition that ecological
263 communities form complex networks with many sensitive parameters that can make
264 deterministic modelling very difficult. Whereas for simple, un-replicated networks
265 we can analytically track links among all species in a community ("low-level
266 properties"), for complex and replicated networks require characterization by what
267 have been termed "higher-level system properties". This later describes the general
268 patterns, dynamics and properties of an ecological community that can be predicted
269 without understanding all the underlying ecological details.

270

A. Low-level network properties: Common metrics of simple networks

273   There are numerous metrics for characterizing networks (Costa et al., 2007),
274 that have been applied to ecological networks (Bersier et al., 2002; Blüthgen et al.,
275 2008; Ings et al., 2009; Thompson et al., 2012; Tylianakis et al., 2010; Vázquez et al.,
276 2009). Linkage density describes the number of interaction links per node and it
277 indicates a diet specialization averaged across the entire trophic network (Tylianakis
278 et al. 2007). Connectance evaluates how many of the possible links in a network are
279 present. Connectance has been found to increase with more generalized foraging in a
280 community (Van Veen et al. 2008) and it has been linked to robustness of an
281 ecosystem and the response of a network to perturbations (Briand 1983; Blüthgen,
282 2010). The trophic similarity among nodes (e.g. species) and links (e.g. feeding
283 pathways) has been used as a measure of redundancy (Cohen and Briand 1984). The
284 efficiency of a network characterizes how closely are nodes connected to each other
285 and the distribution of those connections in a network (Latora and Marchiori, 2001).
286 The cumulative frequency (degree) distribution represents the probability of finding
287 highly connected species (Ledger et al., 2012b, 2013; Montoya et al., 2009). The
288 topological position of nodes, or groups of nodes sharing common biological traits,

289 can be used to identify those that contribute most to network structure (e.g. Guimerà
290 and Nunes Amaral, 2005; Olesen et al., 2007). These and other metrics are commonly
291 used to characterize individual "snapshot" networks. However, there are recent
292 examples of replicated networks used to identify the effect of disturbance on these
293 network characteristics (Box 1).

294 Given a known distribution of species and interactions within a network, the
295 behaviour of the whole community can be described (Ings et al., 2009; Lewinsohn et
296 al., 2006). 'Modular' structures, where groups of nodes interact among themselves
297 more densely than with other nodes, have been reported for antagonistic networks.
298 Modularity or compartmentalisation is common among abundant microorganisms
299 that dominate energy transfer in soil ecosystems (Mulder et al., 2006). Composition
300 of the modules and the distribution of 'redundant' links among modules would
301 indicate the likelihood of cascading effects, whereby changes in one module or
302 network propagate to the next in a 'domino-like' manner (Krause et al., 2003;
303 Stouffer and Bascompte, 2011).

304 'Nested' structuring has often emerged from mutualistic networks, like plant–
305 pollinator webs (Thebault and Fontaine, 2010). Nestedness (Bascompte et al., 2003)
306 describes networks where nodes with few connections tend to be linked to a subset
307 of nodes interacting with more connected nodes. Consequently, most interactions
308 appear asymmetric and organised around a group of highly interconnected nodes.
309 However, both a modular structure in mutualistic networks (Olesen et al., 2007) and
310 a nested structure in antagonistic networks (Cagnolo et al., 2011) have been reported.
311 Profound nestedness has also been observed in randomly generated networks, and
312 thus should not necessarily be taken as evidence for specialisation (Blüthgen, 2010;
313 Blüthgen et al., 2007).

314 Common metrics and population dynamic models have been used to examine
315 stability of simple networks (e.g. Layer et al., 2010, 2011). It has been suggested that
316 nested patterns might stabilise communities (Bastolla et al., 2009), though this
317 approach has faced criticism of its underlying metrics (Blüthgen, 2010) and model
318 assumptions, particularly those related to functional responses (see Holland et al.,
319 2006) and interspecific competition (Benadi et al., 2012). More recently, James et al.
320 (2012) have shown that connectance, rather than nestedness, best explains variation
321 in stability. Despite this ongoing debate, the underlying logic is identical to that
322 proposed in the 'insurance hypothesis' of positive biodiversity effects (Loreau et al.,
323 2001). Higher connectance is associated with higher functional redundancy, which
324 may dampen the functional consequences of stochastic variations or loss of one or a
325 few species (Reiss et al., 2009; Tilman et al., 2006).

326

B. Higher-level properties: Considerations for the analysis of complex replicate networks

While analyses of simple "snapshot" networks have provided a framework to assess the effects of environmental change or disturbance on ecosystem structure and function, such approaches rely on assumptions that are not always fully considered. Firstly, the construction of ecological networks is largely carried out by pooling all interactions identified at several dates and in several locations into a single network, assuming that species interactions are invariant over time (at least during the study duration) and space. However, Derocles et al. (2014) showed that the structure of host-parasitoid food webs are much simpler at the local scale within a single field than at the global scale across the landscape, which included all crop fields of interest and uncultivated environments within the area. Despite this discrepancy, most analyses use aggregated data that overlooks the spatial variability in the landscape, which can significantly alter an ecological network structure (Tylianakis and Morris 2017).

Secondly, pooling data across different habitats into a single network assumes the equal distribution of species and their interactions throughout the space. However, this may not always be true as there exist species that can have disproportionally high number of interactions with other species only in some habitats (Evans et al., 2013). This assumption can therefore lead to inaccurate conclusions about ecological processes and the associated ecosystem services (*e.g.* natural pest control in Derocles et al., 2014). For instance, when two species share a common natural enemy, the presence of one species may negatively impact upon the dynamics of the other through an increase in predation risk or apparent competition (Holt, 1977; Muller and Godfray, 1997; Morris et al., 2004). In this context, an ecological network approach would allow the identification of potential cases of apparent competition through the detection of common natural enemies shared by prey species. This must be done cautiously as the focal food web could be an assemblage of multiple interactions (potentially disconnected either in time, space, or both), rather than an accurate representation of the ecosystem. In fragmented ecosystems, in particular, at least some of the species studied may not be able to move freely between habitats. Barriers in the landscape, which impede free movement of individuals, would therefore be ignored when a single network is built using data from several different habitats.

Thirdly, robustness analysis based on simulating species loss assumes that secondary extinctions are only driven by the direct negative effect of primary

363   extinctions, which have a bottom-up control on a network (Barbosa et al., 2017;
364   Sanders and van Veen, 2012). However, cascading extinctions can also be the result
365   of much more complex effects, sometimes driven by a top-down control (Sanders et
366   al., 2015) or indirect effect such as apparent competition (Holt, 1977; Muller and
367   Godfray, 1997; Morris et al., 2004). Such top-down controls or indirect effects require
368   the information on species dynamics and so are less likely to be detected with
369   "snapshots" of the ecosystem structure that contain a single network alone.

370   Empirical evidence for how ecological networks vary through time and space
371   are still sparse. It is often unknown whether differences observed in network
372   structure are due to an ecosystem disturbance or are a consequence of temporal and
373   spatial dynamics. Including such dynamics in network analyses would greatly
374   enhance our understanding of ecosystem functioning (Heleno et al., 2014). While the
375   robustness analysis performed on "snapshot" networks relies on species extinction
376   to identify cascading secondary extinction, it has been demonstrated that
377   incorporating dynamics in ecological networks had a significant impact on the
378   robustness of the networks, as the ecological function of a species can disappear
379   even before the species itself goes extinct (Säterberg et al., 2013). Therefore,
380   cascading secondary extinctions may result from a simple decrease in the abundance
381   of a given species because of the prey-predator dynamics constantly taking place in
382   the ecosystem. Indeed, dynamical models tested both on theoretical and observed
383   networks have shown that the species most likely to go extinct first are not the one
384   with an increased mortality rate, *i.e.* the species that would be considered as
385   "primary extinction" in classical robustness analysis (Srinivasan et al. 2007). The
386   incorporation of spatial and temporal dynamics into the network ecology is still in
387   its infancy (Tylianakis and Morris 2017).

388

389   ## C. Dynamics: Assessing disturbance in complex replicated
390   networks

391   Analyses of complex networks have been increasingly used to examine the
392   community-wide impacts of external disturbance. Degree distribution, modularity
393   and the inherently nested structure of mutualistic networks appear to confer
394   resistance to perturbations (Fortuna and Bascompte, 2006). Analyses of pollination
395   networks indicate a more rapid loss of interactions than of species, via extinction,
396   following habitat destruction (Aizen et al., 2012). This would imply that rare plants
397   can better resist habitat destruction than common plants that, in turn, support a large
398   numbers of generalist pollinators (Aizen et al., 2012). However, despite there being
399   this type of recurring response of networks to perturbation, the effects of global and

400   regional changes on network structure are often variable and depend on local
401   environmental conditions, and can include interactive effects of multiple stressors
402   (Tylianakis et al., 2008). Moreover, the heterogeneous organisation of complex
403   ecological networks is yet to be fully understood as novel network structures and
404   their relationship with ecosystem stability and functioning are still emerging
405   (Allesina and Pascual, 2008; Garcia-Domingo and Saldaña, 2008; Lu et al., 2016;
406   Lurgi et al., 2012; Stouffer and Bascompte, 2011).

407   Certain network structures appear to be relatively little influenced by
408   disturbance, due to non-random, scale-free distributions of links (Albert et al., 2000;
409   Barabasi, 2009; Parrott, 2010). Such structured networks may include a few hubs,
410   each having disproportionally high number of connections and potentially "small-
411   world" properties, where the mean path-length between any two nodes is shorter
412   than expected by chance (Watts, 1999; Watts and Strogatz, 1998). Scale-free and
413   small-world mycorrhizal networks, as found in Douglas-fir forests (Beiler et al.,
414   2010), are relatively resistant to random perturbations (removal or death of
415   individual trees) because of the small probability of losing the whole hub. In
416   contrast, the targeted removal of hub trees, via logging or pest infestation, would
417   substantially fragment the mycorrhizal network and slow down forest regeneration
418   (Beiler et al., 2010; Bray, 2003).

419   Several studies have used complex networks, replicated across time or space,
420   to determine the impacts of environmental disturbance (see Table 2 for a summary
421   of approaches). The measurement of disturbance or management intervention is
422   either explicitly incorporated into a survey design or is assessed across an implicit
423   gradient. Commonly, network robustness is evaluated to determine how a given
424   network attribute or property is altered as species or functional groups are removed
425   from the network (e.g. Genini et al., 2010). Studies have also examined the effects of
426   primary species loss by measuring the rate of cascading secondary extinctions
427   induced as consumers are left without resources and thus go extinct themselves, and
428   so on (Memmott et al., 2004; Montoya et al., 2006; Pocock et al., 2012). Dunne et al.
429   (2002) simulated cascading extinctions by removing species from 16 food webs and
430   quantified the number of secondary extinctions that followed. The scenarios
431   included random removal of species, removing species with the most links to other
432   species, or removing species with the least links to other species. When similar
433   approaches were applied to food webs in detritus-based systems, it was perhaps
434   unsurprising that the most connected species were the most important for food web
435   resistance (Calizza et al., 2015). The more recent development of meta-networks has
436   started to incorporate the idea of spatial movement as a way of compensating for

437   local species extinction within regional species pools that persist. For example,

438   Mougi (2017) found that a moderate level of spatial coupling between habitats can

439   attenuate some disturbance events due to species emigration. Whilst the complexity

440   and realism of network approaches have increased (e.g. Säterberg et al., 2013), some

441   fundamental limitations still remain. In particular, much of the contribution of

442   individual species dynamics to the network have been ignored, and we cannot yet

443   predict either what roles species remaining in networks might assume or those

444   functions might be maintained in the face of species extinctions (e.g. adaptive

445   trophic behaviour, Valdovinos et al., 2010).