Going Big

Stefano Allesina

Many biological systems are large: the gene-regulatory network of *Drosophila melanogaster* is composed of more than 15,000 genes interacting through activation and inhibition; over a thirty-year period, Charles Robertson documented more than 15,000 distinct interactions between 456 plants and their 1420 pollinators in a single location in western Illinois (Robertson 1929; Memmott and Waser 2002); the spread of influenza in Chicago and surrounding areas is mediated by a large social network comprising about ten million nodes, connected by billions of interactions.

These large biological systems can be represented as networks of interactions between agents (genes, individuals, populations): empirical biological networks invariably contain many agents with few interactions, and few with very many. This type of network structure makes it difficult to describe the systems using mean-field approximations, in which every agent is approximated by a typical, average agent. Empirical network structures are also very different from what we would expect under simple models (such as the Erdős–Rényi random graph, in which every node has the same probability of connecting to every other)—nodes in biological networks have been coevolving for long periods of time, and thus network structure is believed to reflect the underlying biological processes of the system.

An underappreciated aspect of these large networks is that they are quite variable: When measuring an arctic pollination network in Greenland, Olesen et al. (2008) found that the number of interactions between plants and pollinators varied dramatically from day to day. Similarly, were we to build the food webs of two lakes a few kilometers apart, we would find substantial differences. However, the dynamical processes occurring on these networks seem to be fairly robust to their ever-changing nature: Every day in Chicago somebody is born, somebody dies, somebody breaks up with their boyfriend—yet we do not expect these events to dramatically alter the number of cases of influenza in a given year. This suggests that, although we can only measure *a* network, rather than *the* network of interactions, all sampled network structures for a given system belong to a well-defined statistical ensemble, in which some unknown but important quantities are kept constant.

Historically, theoretical ecologists have devoted much effort to the study of small dynamical systems. This year marks the 94th anniversary of the work of Volterra (1926) on predator–prey interactions (the same set of equations had been published by Lotka (1925) the year before). These seminal contributions led to the birth of the field of theoretical ecology and spurred the development of ever more refined sets of equations describing the interaction between two populations. Interestingly, after almost a century, there is still debate over which variants of these equations can best model how species interact (Arditi and Ginzburg 2012). Besides the great influence of the pioneering work by Lotka and Volterra, the field concentrated on small systems also out of mathematical and empirical convenience. From a mathematical point of view, studying even the simplest system of nonlinear, autonomous differential equations describing the growth and decrease of interacting populations becomes prohibitively difficult when we include more than three populations. For larger systems, we can rely on numerical integration on a computer, but again it is difficult to understand these systems well enough to know what to look for in the rich output of such simulations. In the context of empirical studies, measuring interaction strengths is very complex and quite debated (Berlow et al. 2004); accurately measuring many interactions among many species seems to be outside the realm of possibility at the moment.

The study of small ecological systems has produced many beautiful and obviously useful results. I am thinking especially of the successes in modeling tightly coupled systems, such as pathogen–host dynamics and laboratory experiments (e.g., Costantino et al. 1997, Yoshida et al. 2003, Dai et al. 2012, among many). Small, treatable systems such as these are also great for developing intuition about the behavior of complex systems in general and can be seen as building blocks of larger systems. However, I sometimes feel that the focus on small systems could have serious side effects for the discipline as a whole. For example, much of our understanding of competitive interactions descends from the careful study of two interacting populations, and yet the beautifully simple result that intraspecific competition must exceed interspecific competition for populations to coexist does not extend to more than two competitors (Barabás et al. 2016). In addition, some documented dynamics leading to coexistence (such as intransitive competition) cannot even be studied for less than three competitors (Sinervo and Lively 1996, Allesina and Levine 2011), and focusing on two species at a time does not allow for the study of the community structure that emerges when pooling all interactions together (Barabás et al. 2016).

Summarizing, much progress in ecology has been made by taking the limit of $n \rightarrow 2$ for communities of interacting populations. Being somewhat of a contrarian, I cannot help but wonder what ecology would look like today, had we instead taken the limit $n \rightarrow \infty$, starting from infinitely many species and working our way down.

This might seem a strange idea, but it has a strong parallel in the development of physics. The celebrated equations by Newton can accurately describe the motion of the Earth circling the Sun. Take three bodies along with their mass, position, and velocity, however, and you will end up with a system of equations that cannot be solved analytically. This three-body problem has important consequences. For example, although many "proofs" of the stability of our solar system have appeared in the literature, we currently believe the motion of the planets to be chaotic, with large-scale computations being the only means to produce accurate results (Laskar 2013). However, when physicists are confronted with very many particles, for example gas molecules in a room, they do not write equations of motions for each and every one of them and integrate the equations in a computer. Rather, they rely on a completely different theory that, by describing the system statistically, provides important insights on its dynamics. In this way, one can show that fundamentally statistical quantities, such as temperature and pressure, play a key role in determining the system's behavior.

Continuing the analogy with physics, one might ask what the equivalent of temperature and pressure for ecological systems would be—and especially what kind of a toolbox one would need to be able to determine which quantities have the largest effect on the fate of large ecological dynamical systems.

In my explorations around these themes, I stumbled upon the theory of random matrices. This branch of mathematics is concerned with the characterization of the distribution of eigenvalues and eigenvectors of (typically infinitely) large matrices whose coefficients, rather than being fixed numbers, are random variables (Bai and Silverstein 2010). Several characteristics make results in random matrices directly applicable to the study of biological systems. First, this is a theory for large matrices, and networks and matrices are two ways of representing the same problem. Second, although it is difficult to precisely measure interactions in biological systems (and these numbers would be subject to change anyway), one might hope that describing them with distributions would be simpler and more natural. Third, and most importantly, many results in random matrix theory are *universal* (Tao and Vu 2010): the exact distributions for the coefficient do not matter, provided that some quantities (e.g., mean, variance, and correlation) are kept constant. Hence, one can take a very complex network problem and use the machinery of random matrix theory to determine which quantities have the largest influence on eigenvalue distribution. Furthermore, the universality property guarantees that these conclusions will be robust.

Many ecological problems (notably, local asymptotic stability (May 1972), metapopulation capacity (Hanski and Ovaskainen 2000), and spread of infectious diseases (Van Mieghem et al. 2009) can be turned into problems regarding the eigenvalues of certain matrices, so that one can directly translate mathematical results into ecological insights. Though the first application of random matrix theory in ecology dates back to the seminal work by May (1972) on complexity and stability, ecologists have rarely used these tools in the past forty years. Right now, the timing for an exploration of these topics and their potential for biology could not be more perfect: In mathematics, random matrix theory is experiencing a phase of exponential growth, with many active researchers producing new, fundamental results and methods at a very high rate (e.g., Rogers 2010, O'Rourke and Renfrew 2014, Aljadeff et al. 2015). Thanks to these advances, one can analyze ever more complex, structured random matrices, unveiling which broad-scale properties of ecological systems strongly influence dynamics (Allesina and Tang 2015, Allesina et al. 2015, Grilli et al. 2015). Ecologists should engage with the mathematical community, proposing new challenges and classes of random matrices that are of ecological interest. This dialogue between mathematicians and ecologists could both further the study of random matrices and produce beautiful results in the realm of ecology.

Besides the opportunity granted by the development of this theory in mathematics, the availability of new, high-throughput ecological data should spur the development of a toolbox capable of harnessing the information contained in these datasets of unprecedented quality and size. It is not surprising that a random matrix approach immediately appealed to the community of scientists working on microbial communities and metagenomics (Coyte et al. 2015), as our ability to deal with communities containing hundreds of "species" using more traditional approaches is quite limited.

In summary, ecological systems are large, network-structured, and variable, and we need to build a toolbox that can handle these complex systems in a simple and natural way. The theory of random matrices is a promising research avenue, with many active researchers, and fundamental, new results published every year. Using this and similar techniques, we can pinpoint important quantities that largely drive the dynamics of ecological systems.

The need for good ecological theories capable of dealing with the complexity of natural communities is growing every day, be it for managing ecosystems, or to understanding how to preserve or restore essential ecosystems services. With these growing needs in mind, I maintain that in the next century ecology should go big or go home.

References

Aljadeff, J., D. Renfrew, and M. Stern. 2015. Eigenvalues of block structured asymmetric random matrices. Journal of Mathematical Physics 56:103502.

Allesina, S., J. Grilli, G. Barabás, S. Tang, J. Aljadeff, A. Maritan, and A. 2015. Predicting the stability of large structured food webs. Nature Communications 6(7842).

Allesina, S., and J. M. Levine, 2011. A competitive network theory of species diversity. Proceedings of the National Academy of Sciences of the United States of America 108:5638–5642.

Allesina, S., S. Tang, 2015. The stability–complexity relationship at age 40: A random matrix perspective. Population Ecology 57:63–75.

Arditi, R., and L. R. Ginzburg. 2012. How species interact: Altering the standard view on trophic ecology. New York: Oxford University Press.

Bai, Z., and J. W. Silverstein. 2010. Spectral analysis of large dimensional random matrices. Berlin: Springer.

Barabás, G., M. S. Michalska-Smith, and S. Allesina. 2016. The effect of intra- and interspecific competition on coexistence in multispecies communities. American Naturalist, in press.

Berlow, E. L., A.-M. Neutel, J. E. Cohen, P. C. De Ruiter, B. Ebenman, M. Emmerson, et al. 2004. Interaction strengths in food webs: issues and opportunities. Journal of Animal Ecology 73:585–598.

Costantino, R. F., R. A. Desharnais, J. M. Cushing, and B. Dennis. 1997. Chaotic dynamics in an insect population. Science 275:389–391.

Coyte, K. Z., J. Schluter, and K. R. Foster. 2015. The ecology of the microbiome: Networks, competition, and stability. Science 350:663–666.

Dai, L., D. Vorselen, K. S. Korolev, and J. Gore. 2012. Generic indicators for loss of resilience before a tipping point leading to population collapse. Science 336:1175–1177.

Grilli, J., G. Barabás, and S. Allesina. 2015. Metapopulation persistence in random fragmented landscapes. PLOS Computational Biology 11:e1004251.

Hanski, I., and O. Ovaskainen. 2000. The metapopulation capacity of a fragmented landscape. Nature 404:755–758.

Laskar, J. 2013. Is the solar system stable? In: B. Duplantier, S. Nonnenmacher, and V. Rivasseau, eds. Progress in mathematical physics. Vol. 66. Chaos. Berlin: Springer, 239–270.

Lotka, A. J. 1925. Elements of physical biology. Baltimore: Williams and Wilkins.

May, R. M. 1972. Will a large complex system be stable? Nature 238:413–414.

Memmott, J., and N. M. Waser. 2002. Integration of alien plants into a native flower-pollinator visitation web. Proceedings of the Royal Society of London B: Biological Sciences 269:2395–2399.

Olesen, J.M., J. Bascompte, H. Elberling, and P. Jordano. 2008. Temporal dynamics in a pollinator network. Ecology 89:1573–1582.

O'Rourke, S., and D. Renfrew. 2014. Low rank perturbations of large elliptic random matrices. Electronic Journal of Probability 19:1–65.

Robertson, C. 1929. Flowers and insects: Lists of visitors to four hundred and fifty-three flowers. Carlinville, IL: C. Robertson.

Rogers, T. 2010. Universal sum and product rules for random matrices. Journal of Mathematical Physics 51(9):093304.

Sinervo, B., and C. M. Lively. 1996. The rock-paper-scissors game and the evolution of alternative male strategies. Nature 380:240–243.
Van Mieghem, P., J. Omic, and R. Kooij. 2009. Virus spread in networks. IEEE/ACM Transactions on Networking 7:1–14.
Volterra, V. 1926. Fluctuations in the abundance of a species considered mathematically. Nature 118:558–560.
Tao, T., and V. Vu. 2010. Random matrices: universality of ESDs and the circular law. Annals of Probability 38:2023–2065.
Yoshida, T., L. E. Jones, S. P. Ellner, G. F. Fussmann, and N. G. Hairston. 2003. Rapid evolution drives ecological dynamics in a predator–prey system. Nature 424:303–306.