Chaos, Solitons and Fractals 108 (2018) 94–103

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

Complexity and heterogeneity in a dynamic network

David Lambert [a], Fabio Vanni [b,*]

[a] Center for Nonlinear Science, University of North Texas, USA
[b] Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy

ARTICLE INFO

Article history:
Received 18 September 2017
Revised 20 December 2017
Accepted 16 January 2018

Keywords:
Complexity
Phase transition
Network formation
Network value
Stochastic process
Critical phenomena
Long-range correlations

ABSTRACT

We present an approximate analytical solution for the connectivity of a network model with a "non-simultaneous" linking scheme. This model exhibits node-space correlations in the link distribution, anomalous fluctuations in the time series of the connectivity variable, and a finite-size effect: the maximum number of links occurs away from the critical value of the system parameter. We derive an exact Master Equation for this model in the form of an infinitesimal time-evolution operator. Fluctuations are much more important than the mean-field approximation predicts, which we attribute to the heterogeneity in the model. Finally, we give a sketch of possible real world applications where the value of a network is related to the number of links.

1. Introduction

Complexity is a polymorphous concept, with definitions that vary from one discipline to another. Herein, we will refer to a system as complex if it exhibits spontaneous emergent phenomena over a small range of values of the free parameter(s) of the system. One of our working hypotheses is that complexity is closely tied to heterogeneity. The results below indicate that node heterogeneity is instrumental in determining the degree of interconnectedness in a model for network dynamics.

Our model demonstrates that finite-size effects can be extremely important, especially in systems that display phase transitions. Furthermore, the interplay between risk and profit indicated by the model leads one to the conclusion that there is an optimal size for types of networks that obey the same general principles, for example economic, biological, and sociological groups large and small [1].

Since we work with a network with a dynamic topology, it is the fluctuations and heterogeneity of the network that are most relevant to the behaviors observed. Indeed, at the parameter values at which the network is complex, these fluctuations in the degree of interconnectedness become extraordinarily large, comparable in size with their allowed range.

This emphasis on the dynamical aspects of network properties represents a departure from the standard approaches to studying networks. Most social and financial networks have been studied with an emphasis on their characteristic topological features, especially the *patterns of connection* (often referred to as complexity) between their elements [2,3]. For example, financial economists have largely discussed the benefits of interactions among financial intermediaries. Some degree of interconnectedness is crucial to the proper functioning of financial systems, as no single institution can access the full range of available capital and investment opportunities in the economy. Connections among financial institutions are also assumed to facilitate risk sharing, decrease the uncertainty faced by individual agents, and so allow agents to offer better services to the economy.

On the other hand, complexity is also regarded as a source of system breakdown [4]. In particular, increasing interconnectedness in the market in terms of contracts among financial institutions comes at the price of increasing inaccuracy in the estimation of systemic risk [5]. So, in financial markets, the challenge for market participants, policymakers, and regulators is to find a balance between the benefits of interconnectedness and its potentially harmful destabilizing effects [6,7].

In the present paper, we examine a model from a new network class based on agent preferences, namely preferred degree networks [8], where the number of links continuously fluctuates and the system has a non-trivial steady state distribution. In this class of networks, the system undergoes a phase transition from

* Corresponding author.
E-mail addresses: fabio.vanni@santannapisa.it, FabioVanni@my.unt.edu (F. Vanni).

https://doi.org/10.1016/j.chaos.2018.01.024

a sparsely connected network to a densely connected one. The intermediate critical condition is the condition of maximal population heterogeneity among the nodes and the system shows an abrupt onset of anomalously large fluctuations in the network connectivity, which emphasizes the heterogeneity of the units' preferences. Our analytical results reveal that these large fluctuations are poorly accounted for by the mean-field approximation of the model.

This class of networks differs from the standard literature of pairwise network formation [9] in that a link is not a bilateral compromise between pairs of nodes. An asymmmetry that is in contrast with the statistical symmetry present in other common methods of generation of random graphs, such as via a linking probability matrix.

This class of models is based on the initiative of single units, which according to their preferences can propose to generate or destroy a connection with another node (which can accept or reject the proposition). This change of perspective introduces a dramatic difference in the dynamics of networks displaying the emergent phenomena typical of complex systems.

The model which we analyse has two types of units [10]: generators and destroyers of links. We propose an analytical solution for the equilibrium mean value of the connectivity, which exhibits a phase transition. We provide a closed-form equation for the phase transition thereby characterizing the correspondent critical point.

These results are obtained using a relativity new approach to stochastic processes which makes use of mathematical notations reminiscent of the quantum mechanical formalism [11], and a mean-field approximation to the Fokker–Planck equation for the probability density of links. This approach has been turned out to be crucial, since no purely analytical prediction has been obtained so far for this model, see [8–10,12,13]), where authors have used a more standard approach to stochastic processes. In order to provide a microscopic description of the network, we write an approximate Langevin equation, which allows one to characterize the network in terms of emergent properties at criticality through the study of spatial and temporal correlations. Additionally, we highlight some limitations of the mean-field approximation in capturing the heterogeneity of the nodes in their dynamics of creating and destroying links. These corrections are non-negligible when the system is at its critical point.

As further hallmarks of complexity, we stress the importance of finite-size effects, observing how finite networks produce more links away from the critical condition. Indeed, real-world systems' statistical properties are affected by finite-size and other truncation effects which can play an important role in defining the complexity of networks in terms of the effects of systems in constrained situations such as a limited number of units in the system (i.e., small groups), which is related to coordination issues [14]. This gives rise to a paradigm of emergent properties of groups including the fact that larger team sizes lead to an increasing need for coordination that can limit the efficiency of group members, drawing attention to the optimal connectivity condition as a function of the global size of the network.

The model we describe is an expository model, having the purpose of highlighting and explaining the most crucial mechanisms underlying the phenomena of complex evolving systems as discussed in many disciplines, in particular economics and finance [15–18].

So, without any predictive intention, we set forth an abstract example of a system which gains value according to its interconnectedness, and bears a cost depending on the number of active nodes (i.e., generators of links). The resulting profitability shows a signature of complexity in terms of finite-size network effects: small groups reach a maximal profitability far from the critical point of maximal heterogeneous population, but they tend to suffer less uncertainty of the expected connectivity. As the size of the network increases, we imagine that the system tends to organize itself near a critical point where the network has its maximal profitability; however, this point is also associated with a very high uncertainty (connection volatility). In this state, the system can be more vulnerable to possible systemic failure since it spends some part of its time in an unprofitable state. In terms of social and economic policy, minimizing the importance of heterogeneity also in mathematical terms (by using the mean-field approximation) leads one to drastically underestimate the size of fluctuations at the critical point, which could lead to an underestimation of the risk in the system.

2. The generators–destroyers model and its analytical description

Among the possible models in the class of unit-driven networks, we select the most simple heterogeneous case where we have a bipartite graph in which two types of nodes exist: one group of nodes aims to create new links every time they are selected, the other group aims to cut a link with previously connected partners.

The Generators–Destroyers model is a model for the intergroup link dynamics between a group of link generators and a group of link destroyers directly derived from the introvert-extrovert model (XIE) as introduced and studied in [8,10]. In the following calculations, it is assumed that time is continuous (in the sense that there is no fixed minimum time between changes in the number of links), one can think of this as an event-based approach. Links are bi-directional, and there can be at most one link between any two units. Neither self-links nor intragroup links are considered, as these subgraphs quickly go to a static equilibrium state, thus the graphs produced come from a subset of the set of simple graphs. The generators create links as long as there is at least one destroyer available to which it has no link. Destroyers destroy links until they are not linked to any generators.

It is convenient to represent a graph in terms of a matrix, called the adjacency matrix. This matrix is formed by enumerating the vertices of the graph, the i, jth entry of the matrix is the number of links from vertex i to vertex j. In the model under consideration, the links are bi-directional, and there is at most one for every pair of vertices, so the adjacency matrix is symmetric (a bi-directional link consists of one unidirectional link in each direction) and consists of ones and zeros (either a link is occupied or it is not). Since the graph is dynamical in the generators–destroyers model, so is its associated adjacency matrix.

As mentioned in the introduction, since the standard tools for studying stochastic processes have not sufficed to find an analytical solution for the phase transition of the present system, as a possible path towards an analytical solution to this model, the authors found useful to give to the dynamics of the system a physical interpretation in terms of an operator formalism like that used for the harmonic oscillator in quantum mechanics. In the following theoretical treatment, we model the dynamics as due to the action of an infinitesimal stochastic time-evolution operator H on the adjacency matrix. One can imagine such an operator as a sum of simpler operators, one for each element of the adjacency matrix.

In the present mathematical description we focus on an analytical treatment in terms of the average number of links and link distribution, instead of the degree-distribution and average degree as in [9]. In order to derive an equation of motion for the link distribution, we write a Fokker–Planck equation for this model starting from the formalism of creation and annihilation operators whose basic notions are elucidated in Appendix A. The only free parameter in this representation is the total rate of events, i.e., an overall constant in the following equation:

$$H = \frac{1}{N}\sum_{i=1}^{N_1}\sum_{j=1}^{N_0}\left(L_{ij}^{\dagger} - S_{ij}^{\dagger}\right)\frac{1}{N_0 - \sum_{k=1}^{N_0} L_{ik}^{\dagger}L_{ik}}S_{ij} + \left(S_{ij}^{\dagger} - L_{ij}^{\dagger}\right)\frac{1}{N_1 - \sum_{\ell=1}^{N_1} S_{\ell j}^{\dagger}S_{\ell j}}L_{ij}. \tag{1}$$

In this equation, H is the infinitesimal time-evolution operator, which can be used to generate the master equation of the associated non-homogeneous birth-death linking process (for more details on this formalism and its connection with master equations see Appendix A). The symbols with daggers represent creation operators, those without denote annihilation operators. N_0 is the number of destroyers of links, while N_1 is the number of generators of links, and $N = N_1 + N_0$ is the total number of units. The L operators have to do with setting or erasure of ones in the element of the adjacency matrix given by the positive integers i, j. Similarly, the S operators set or erase zeros in the element of the adjacency matrix corresponding to their subscripts.

The first term in Eq. (1) corresponds to the conversion of zeros to ones, and is associated with the actions of the ith generator on the jth destroyer. The second term corresponds to the conversion of ones to zeros, and is associated with the action of the jth destroyer on the ith generator. It is necessary to include the A type of operator because there is a conserved quantity in the system: the total number of ones and zeros in the relevant portion of the adjacency matrix (this is equal to N_0N_1 and is also the maximum total number of links).

This conservation law makes it possible to view absent links as undergoing a process of creation and destruction dual to that of ordinary links. The model has a symmetry between these shadow links (hence the S symbol in the operators) and the ordinary links. If one exchanges ordinary links for shadow links and generators for destroyers, the model is unchanged.

In words, Eq. (1) represents the following process. A node is selected uniformly randomly from the set of N nodes (thus the factor of $\frac{1}{N}$ in front of the sum). If the node is a generator of links, then a destroyer is chosen uniformly randomly from among the destroyers to which the generator is not already linked (this yields the factor of $\frac{1}{\text{\# of unlinked destroyers}}$). If there are no such destroyers, then nothing occurs. Otherwise a link is created between the selected generator and destroyer (producing the terms $(L_{ij}^{\dagger} - S_{ij}^{\dagger})S_{ij}$). Since this is how the simulations run, Eq. (1) contains the full microscopic description of the dynamics of the model (up to possible differences in implementation, such as discreteness versus continuity of time).

We are interested in the ratio of the difference between the number of ordinary links and the number of unoccupied (shadow) links to the maximum total number of links, formally:

$$\ell = \frac{1}{N_1N_0}\sum_{i=1}^{N_1}\sum_{j=1}^{N_0}\left(L_{ij}^{\dagger}L_{ij} - S_{ij}^{\dagger}S_{ij}\right). \tag{2}$$

which is the (rescaled) *connectance*[1] observable related to the raw total number of links L through:

$$\ell = \frac{2L}{N_1N_0} - 1\,, \qquad \ell \in [-1, 1] \tag{3}$$

in terms of the control parameter:

$$\Delta = \frac{N_1 - N_0}{N}\,, \qquad \Delta \in [-1, 1] \tag{4}$$

which is the scaled population difference between the two groups (note that if we exclude the degenerate cases $N_1 = 0$ and $N_0 = 0$, $\Delta \in [-1 + \frac{1}{N}, 1 - \frac{1}{N}]$).

The next step is to recover an analytic estimation of the average connectance $\langle\ell\rangle$ for various values of the population parameter Δ, thus finding a closed-form solution for the shape of the phase transition plot for this kind of network, and explaining why Δ is referred to as a control parameter.

3. Phase transition and time evolution of the average connectance

From the microscopic Master Equation associated with Eq. (1) it is possible to provide the time evolution of the probability density function of the connectance and find the stationary condition. We use a mean-field approximation to obtain a Fokker–Planck equation and the associated drift-diffusion process, from which we recover an accurate prediction of the equilibrium value of the connectance and the corresponding phase transition. This precise understanding of the “first-order” equilibrium variable is obtained at the cost of misjudging the intensity of fluctuations (variance) at criticality where correlations among the units are not negligible.

The equation of motion of (the Fourier transform of) the probability density of the variable corresponding to the connectance operator (see Eq. (2)) is given by:

$$\frac{d\left\langle e^{i\lambda\ell}\right\rangle}{dt} = \frac{1}{N}\left\langle\sum_{i=1}^{N_1}\sum_{j=1}^{N_0}\left[\left(e^{\frac{i\lambda}{N_0N_1}} - 1\right)\frac{1 - \ell_{ij}}{N_0 + 1 - \sum_{k\neq j}\ell_{ik}} + \left(e^{-\frac{i\lambda}{N_0N_1}} - 1\right)\frac{1 + \ell_{ij}}{N_1 + 1 - \sum_{m\neq i}\ell_{mj}}\right]e^{i\lambda\ell}\right\rangle \tag{5}$$

and in the case of the mean-field approximation we assume $\ell_{i,j} = \ell$, since the ℓ_{ij} are all exchangeable and have mean ℓ.

$$\frac{d\left\langle e^{i\lambda\ell}\right\rangle}{dt} = \left\langle e^{i\lambda\ell}N_1N_0\left[\frac{1}{N}\left(e^{\frac{2i\lambda}{N_1N_0}} - 1\right)\frac{1 - \ell}{N_0 + 1 - (N_0 - 1)\ell} + \frac{1}{N}\left(e^{\frac{-2i\lambda}{N_1N_0}} - 1\right)\frac{1 + \ell}{N_1 + 1 + (N_1 - 1)\ell}\right]\right\rangle \tag{6}$$

Expanding in $\frac{1}{N_1N_0}$ and replacing $i\lambda$ by $\frac{\partial}{\partial x}$

$$\frac{\partial\left\langle e^{i\lambda x}\right\rangle}{\partial t} = \left\langle\frac{2}{N}\left[\frac{(N_1 - N_0)(1 - \ell^2) - 4\ell}{(N_0 + 1 - (N_0 - 1)\ell)(N_1 + 1 + (N_1 - 1)\ell)}\frac{\partial}{\partial\ell} + \frac{2}{N_1N_0}\frac{N + 2 - (N - 2)\ell^2}{(N_0 + 1 - (N_0 - 1)\ell)(N_1 + 1 + (N_1 - 1)\ell)}\frac{\partial^2}{\partial\ell^2}\right]e^{i\lambda\ell}\right\rangle \tag{7}$$

Via integration by parts and an inverse Fourier transform (see Appendix A),

$$\frac{\partial P(\ell, t)}{\partial t} = \frac{2}{N}\frac{\partial}{\partial\ell}\Big\{A(\ell)\,P(\ell, t)\Big\} + \frac{2}{NN_1N_0}\frac{\partial^2}{\partial\ell^2}\Big\{B(\ell)\,P(\ell, t)\Big\} \tag{8}$$

which is the Fokker–Planck equation for the Generators–Destroyers model, where the drift and the diffusion non-homogeneous coefficients are :

$$A(\ell) = \frac{(N_1 - N_0)(\ell^2 - 1) + 4\ell}{[N_0 + 1 - (N_0 - 1)\ell][N_1 + 1 + (N_1 - 1)\ell]} \tag{9}$$

$$B(\ell) = \frac{N + 2 - (N - 2)\ell^2}{[N_0 + 1 - (N_0 - 1)\ell][N_1 + 1 + (N_1 - 1)\ell]}. \tag{10}$$

The fundamental result of our study of the Generators–Destroyers model (and thus XIE networks) is the closed-form solution for the

[1] The connectance is defined as the link density of the network, which is the fraction of the number of actual links over the number of potential links between pairs of nodes. We use the term interconnectedness with the same meaning as we did in the introduction. In contrast, the term connectivity indicates a generic property of the network which gives a measure of the link density in the network.

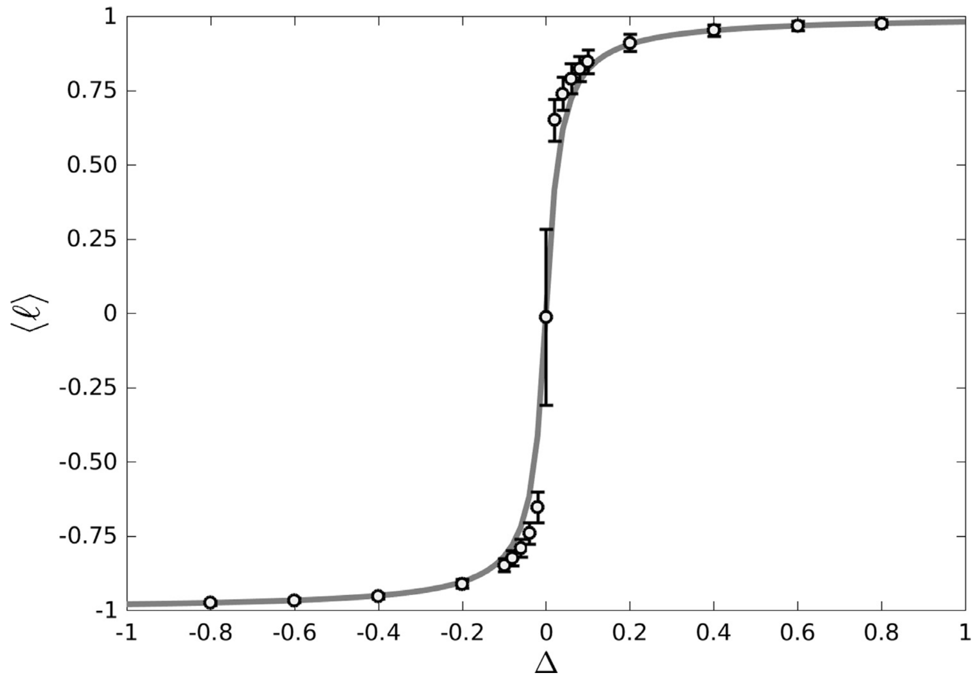

Fig. 1. The average connectance as a function of the scaled population difference Δ. The solid gray line represents the prediction of Eq. (11). The points come from the results of monte carlo simulations, with error bars indicating the standard deviation. This figure was done with $N = 100$ nodes.

equilibrium mean connectance as function of the population parameter $\Delta = \frac{N_1 - N_0}{N}$. From the stationary solution of the probability density in the Fokker–Plank equation Eq. (8), the equilibrium value $\langle \ell \rangle_{eq}$ is the value for which the drift term in Eq. (9) is equal to zero (in particular, its numerator, as it is not possible for its denominator to vanish). From the correspondent quadratic equation, we have:

$$\langle \ell \rangle_{eq.} = \begin{cases} \dfrac{\sqrt{\Delta^2 + (2/N)^2} - 2/N}{\Delta} & \text{for} \quad \Delta \neq 0 \\ 0 & \text{for} \quad \Delta = 0, \end{cases} \tag{11}$$

Note that Eq. (11) has a removable discontinuity at the critical value $\Delta = 0$, i.e., $N_0 = N_1$. This equation also makes clear that the phase transition is discontinuous in the limit as $N \to \infty$. We see in Fig. 1 how it predicts the average connectance of the model, where the bipartite network shows a transition, as a function of Δ, from a sparse condition to a dense one passing through a critical point with an abrupt jump between the two conditions (this demonstrates that Δ is a control parameter). The transition becomes steeper and steeper as we increase the size of the network, i.e., as $N \to \infty$, in fact in the thermodynamic limit we have only three equilibrium points given by $\langle \ell \rangle^{\infty}_{eq.} = \text{sign}(\Delta)$.

At this point, it is possible to write an equation of motion for single trajectories as an Ito process driven by a standard Wiener process W_t and described by the stochastic differential equation:

$$d\ell = -\frac{2}{N} A(\ell) \cdot dt + \sqrt{\frac{2}{N N_1 N_0} B(\ell)} \cdot dW_t, \tag{12}$$

which yields a mean-field approximation of the connectance time-series. Despite the fact that the equilibrium mean values of the connectance are in agreement with the mean-field predicted values, the standard deviation at the critical point is much larger than the mean-field prediction. Even though away from the critical point the fluctuations are also in agreement with the theory, one expects higher-order corrections there as well. In summary, the Fokker–Planck equation Eq. (8) accurately describes the real network for off-critical values of the population parameter and for large networks ($N \to \infty$).

4. The mean-field simplification and the heterogeneity issue close to criticality

The simulated network model shows an anomalous behavior not captured by our derivation. As already discussed in [10] and measured in [9], the system shows persistent fluctuations even for very large networks, thus indicating the presence of correlations among the units. Indeed, at criticality the connectance shows a variance of the fluctuations of the order of one. This fact is not described by the Fokker–Planck equation Eq. (8) nor the trajectory Eq. (12), which instead predict that the intensity of fluctuations will vanish as one increases the number of units. The principal suspect of this mathematical shortcoming is the use of the mean-field approximation.

The main idea of the mean-field simplification is that for each unit we replace the state variable (the node's degree) with the average number of links per node across the system. In this approximation, each unit interacts with the "average individual". This implies that some local interactions are disregarded and the fluctuations among different individuals are assumed to be uncorrelated. However, in this kind of network, especially close to criticality, the links between pairs are correlated, since different nodes can have very different numbers of links.

In our case, the mean-field approximation is entirely encompassed in replacements of the form

$$\frac{1}{1 - \Delta + 2\varepsilon - 2\varepsilon \sum_{k \neq j} \ell_{ik}} \to \frac{1}{1 - \Delta + 2\varepsilon - (1 - \Delta - 2\varepsilon)\ell} \quad \text{where } \varepsilon = \frac{1}{N}. \tag{13}$$

However, this replacement is not exact since the links among pairs of nodes can be very different from the average value in the sense that $\ell_{ik} = \ell \frac{\langle \ell_i \ell_k \rangle}{\langle \ell_i \rangle \langle \ell_k \rangle}$, and in general we have $\langle \ell_i \ell_k \rangle \neq \langle \ell_i \rangle \langle \ell_k \rangle$, where ℓ_i indicates links leading to the node i. In the mean field approximation we replace the state variable (connectance) of each individual link with the average state variable, and, under this condi-

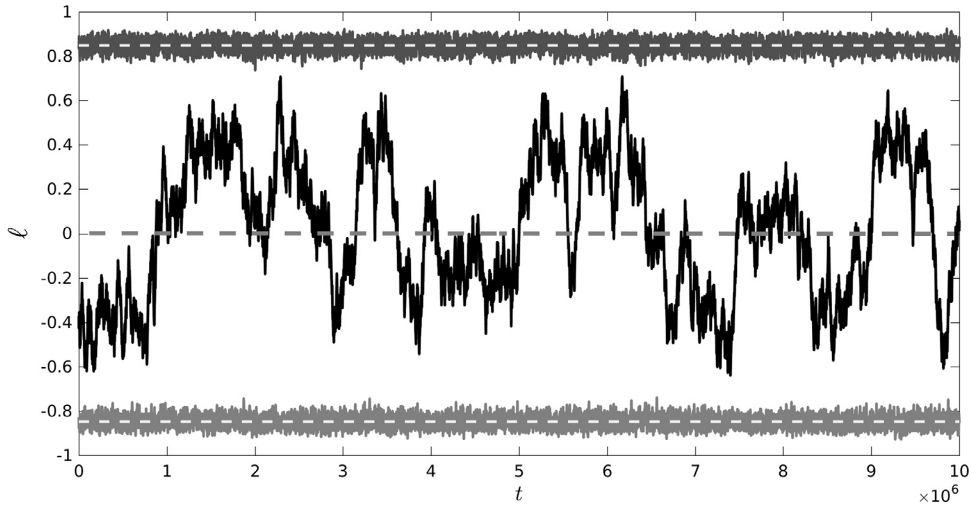

Fig. 2. Three different trajectories of the connectance ℓ for three different values of the population parameter Δ. The top dark gray time series is the case $\Delta = 0.1$, the central black time series is for the critical value of $\Delta = 0$, the bottom light gray series is for $\Delta = -0.1$. The dashed horizontal line is the zero connectance, Above this line the system is in a dense network state below the line it is in a sparse network state.

tion, some network aspects become unobservable across the system (Fig. 2) .

Nevertheless, the mean-field approximation is useful, since it allows for analytical results to be obtained in many cases. It is also often a justifiable approximation, in the sense that for large system sizes the vast majority of the probability mass is typically concentrated at a single value (or very small range of values) of the stochastic variable(s). However, if the system size is small, or the central limit theorem does not apply (for example, due to long-range correlations), then the mean-field approximation may fail. In our case, the mean-field approximation omits systematic corrections due to the fact that the random variables are dichotomous, or perhaps more to the point, due to the fact that the variance of the sums replaced is larger than that of the random variable by which they are replaced. Thus we would need a way to avoid neglecting the correlation between links of different pairs and to take into account the heterogeneity among nodes, which is especially prominant at criticality (when the variance is of order one and the nodes' correlations are very intense). This is the reason why at criticality the number of links explores a wide range of values over time, passing from an almost empty network to an almost complete network during the evolution, as shown in Fig. 3.

The failure of the mean-field approach is evident at the critical point, in fact, this inaccuracy is still present close to criticality as shown in Fig. 1, where we have small deviations of the analytical phase transition from the real transition. The corrections affect both the drift term and the diffusion term, where higher-order approximations should take in account the correlations. The corrections to the drift term should improve the equilibrium value of the connectance, and the corrections to the diffusion term should recover the non-vanishing variance at criticality.

5. Finite-size and small group effects

A crucial aspect of unit-driven network models is that finite networks exhibit certain universal properties and emphasize the presence of correlations in the system. An important aspect of the finite-size effect is related to the existence of an extreme Thouless effect in this model, i.e., a mixed phase transition at which the connectance parameter displays a discontinuous jump and there are large dynamical fluctuations. We have discussed the nature of persistent fluctuations at criticality even for $N \to \infty$ (also attested to by the steep jump in the analytical sigmoid function of Eq. (11)) in the thermodynamic limit as shown in Fig. 4.

Another key effect of finite network size on these networks is that the maximum number of links is obtained relatively far from the critical point, thus the critical condition is not the optimum case for the maximal number of links in the network. In the limit of infinite networks we have that the critical point coincides with the point where the maximum number of links are created in the network.

Let us write the equation for the average value of the total number of links in the network, using Eqs. (3) and (11). as:

$$\langle L\rangle_{eq.} = \begin{cases} \frac{N^2}{8}\left(1-\Delta^2\right)\left(\frac{\sqrt{\Delta^2+(2/N)^2}-2/N}{\Delta}+1\right) & \text{for} \quad \Delta \neq 0 \\ \frac{N^2}{8} & \text{for} \quad \Delta = 0 \end{cases} \tag{14}$$

where we have expressed the number of generators and destroyers respectively as $N_1 = (1+\Delta)N/2$ and $N_0 = (1-\Delta)N/2$. In the present model, a relevant finite-size effect consists of the fact that the critical point is not the point of maximal connectivity. In other words, the maximum number of links, for finite networks, happens for $\Delta > 0$, i.e., when the number of generators is larger than the number of destroyers. This is an effect which is important for small groups and becomes negligible for large groups, where the critical point becomes (in the limit) the point of maximal connectivity.

Proposition 1. *For a given size of the network N, let us define $S(N)$ to be the set of population parameters Δ for which the average number of links $\langle L\rangle_{eq}$ is maximum. The maximum points $S(N)$ are obtained always for $\Delta > 0$ in the case of finite-size systems and tend asymptotically to $\Delta = 0$ when $N \to \infty$.*

Proof. Now, let us define a function of two variables $f(\Delta, N) := <L>_{eq.}$, and study it as the objective function where we want to maximize f with respect to Δ treating N as given:

$$\max_{\Delta \in [-1,1]} f(\Delta, N), \tag{15}$$

where clearly the solution will depend on N: let the maximal value be $S(N)$ attained at $\Delta = S(N)$. The value function is defined by

$$V(N) = f(S(N), N) \tag{16}$$

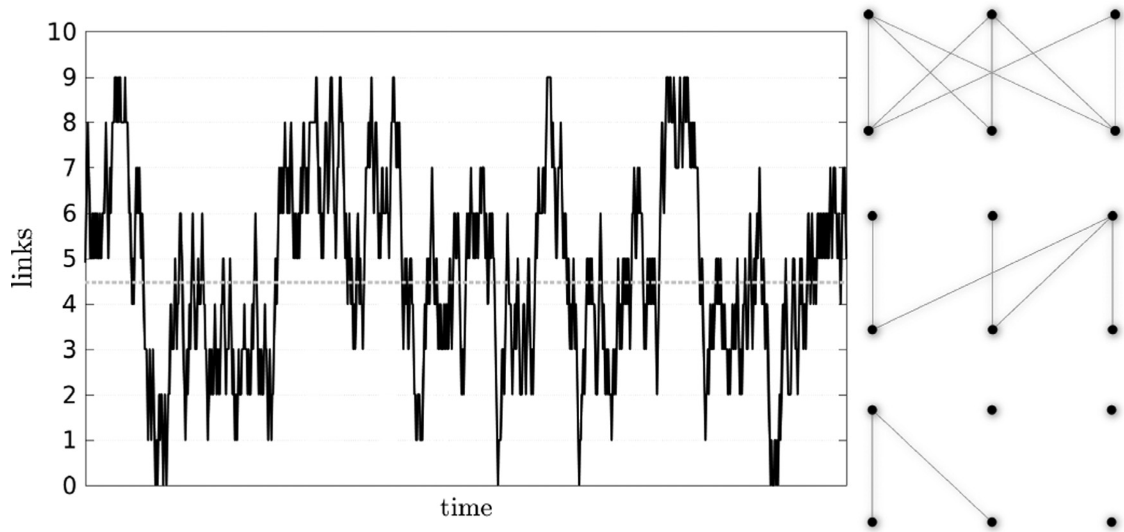

Fig. 3. Left: An example of a small network with $N = 10$ nodes at criticality (i.e., $\Delta = 0$). Right: Example states of the system with $N_1 = N_0 = 3$, showing highly connected, intermediate, and sparsely connected states.

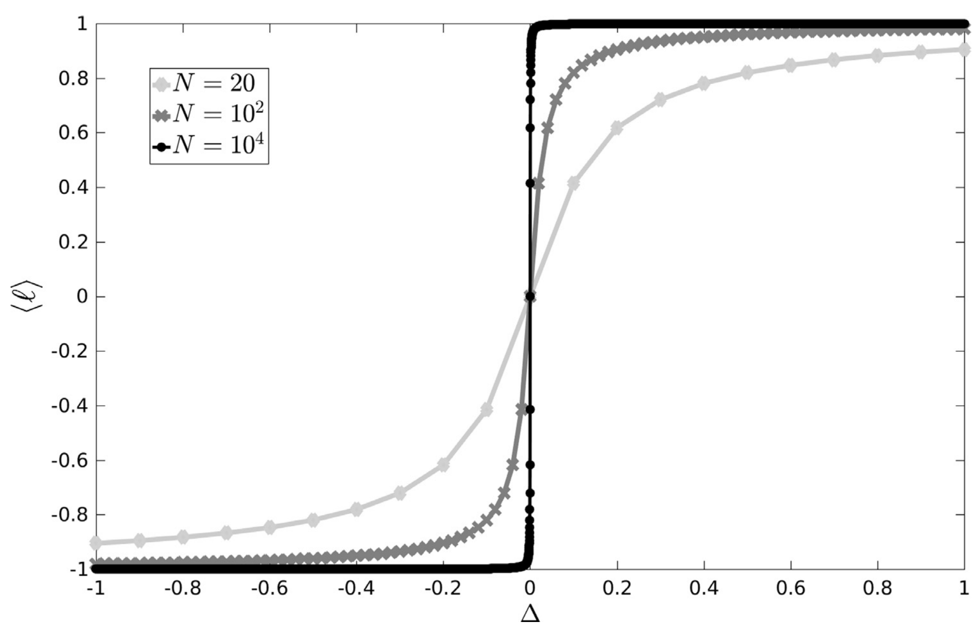

Fig. 4. Finite size effects on the equilibrium value of the connectance.

The Berge maximum theorem guarantees that $S(N) = \arg\max_{\Delta} f(\Delta, N)$ is a continuous function since f is a strictly quasi-concave function [19]. Then it follows from the equations:

$$S^4N^3+4S^3VN^2-2S^2N^3+8S^2VN+4SVN^2-16SV^2+N^3-8VN=0 \tag{17}$$

$$S^3N^3+3S^2VN^2-SN^3+4SVN+VN^2-4V^2=0 \tag{18}$$

that $S(N) = 0$ has no solution for finite N (note that $N = 0$ is not a valid value of N for the purposes of this model). Eq. (17) was obtained from the stationary solution to the drift part of Eq. (8), Eq. (18) was obtained by differentiating Eq. (17) with respect to N and setting $\frac{dV}{dN} = 0$. These equations thus exactly describe the variety on which $f(\Delta, N)$ attains its maximum (in the mean field approximation). As $N \to \infty$, one can see intuitively that $V \to \frac{N^2}{4}$. Plugging this into Eqs. (17) and (18) and solving in the asymptotic limit leads to the formulas

$$S(N) \sim \frac{1}{(4N)^{\frac{1}{3}}} \qquad V(N) \sim \frac{N^2}{4} - \frac{3}{4}\left(\frac{N}{2}\right)^{\frac{4}{3}}. \tag{19}$$

Since $S(N)$ is asymptotically positive, nonzero for finite N, and continuous, it is positive for all finite N. □

The behavior of the expected number of links and the numerical results of the maximization problem are plotted in Figs. 5 and 6. Now let us discuss a plausible picture in which there is a purpose behind the construction of connections. In our plausible world, we consider a trade-off between the effort needed (resources) to create connections and the possible revenue which can be derived from the available connections. We identify the value of the network with the total number of occupied links. One can imagine this value as accounting for ways in which units receive benefits from having links, for example a positive feedback due to exchanged information, knowledge diffusion or a profit deriving

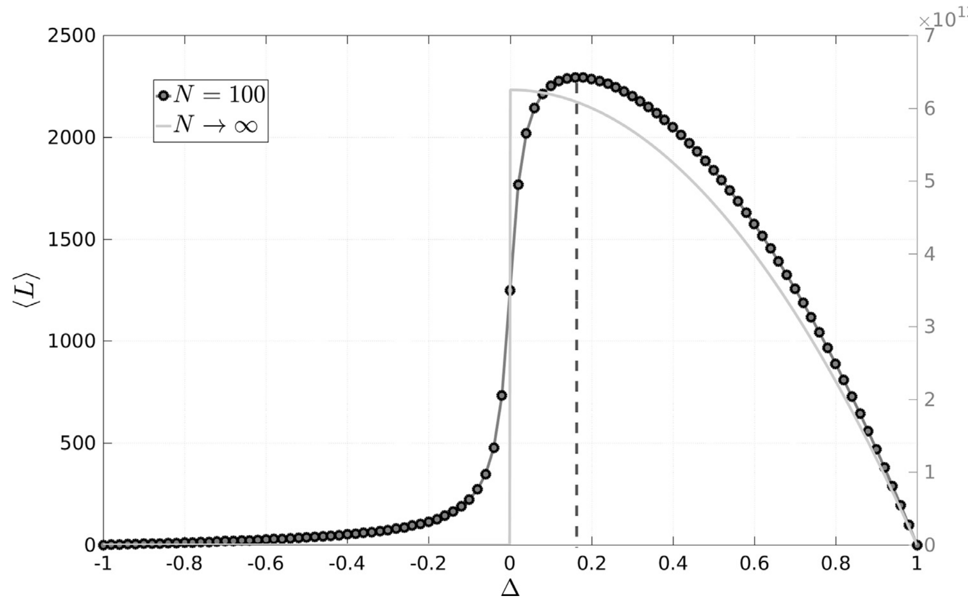

Fig. 5. Average number of actualized links in a network with $N = 100$ units, dotted dark gray points. We observe how the maximum number of links is obtained for $\Delta > 0$ (located at the dashed line), so the maximum point is not the critical point as in the case of an infinite network (here 1 million nodes) represented by the light grey curve.

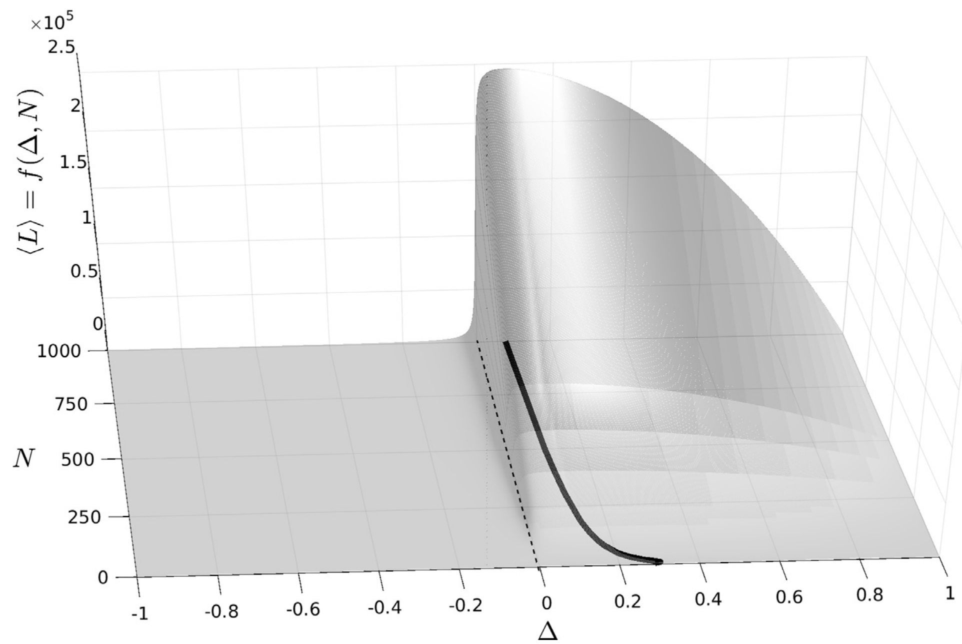

Fig. 6. A three dimensional plot of the expected number of links as a function of the scaled population difference Δ and the system size N. The black curve in the horizontal plane represents the loci of maximum points, $S(N)$, for which we have the highest number of links for each system size. We see that for small networks the maximum point is in the super-critical regime ($\Delta > 0$), as the network size increases the maximum points tend asymptotically to the critical value of $\Delta = 0$.

from an external investment realized through a financial loan. So, the occupied connections represent an opportunity to gain some form of payoff for the individuals.[2] We imagine that the cost of the network depends on the number of generators that make the effort to create links, i.e., each individual has a certain cost to behave as a generator, which are the units creating the connections.[3]

We define the profitability of the bipartite network to be the ratio of the total links L (sources of a possible gain) to the cost of having N_1 generators (source of a loss), the cost (in this simplified world) to create a link can be roughly thought of as proportional to the number of generator units. So the profitability index of the bipartite network is defined as:

$$\mathcal{P} = \frac{\langle L \rangle_{eq.}}{\alpha N_1} \tag{20}$$

which, in terms of benefit cost analysis, is the index given by the ratio of the total cash inflows to the initial investment. If the index

[2] In other terms [20], the utility function representing the benefit that a node obtains if it is connected to another, is considered constant since the neighbor selection is uniformly random.

[3] Let us note that the network cost is not the cost of creating a link which actually could be included in the neat utility discounted by a possible linking cost. Here the cost is meant as the effort to become and maintain the condition of being a link generator. Similarly, we could also consider the cost to cut a connection incurred by the nodes that destroy links, but since we consider linear-costs the net-value between cutting and creating a link, we take only the cost for the creation.

is larger than 1 the system is profitable. Our choice of a linear cost function is an arbitrary choice just for the purpose of discussing a simple situation. In fact, there is no universal law for cost curves, rather, they depend on the specific case under study. A popular example of performance and costs is given by Metcalfe's law in information and communication networks [21–23], e.g., Ethernet nodes and WWW servers, There are also many examples in economics [24,25] where network performance can influence market structure, firm behavior and economic outcomes.

At this point, taking the cost of links to be a linear function it is possible to find an analytical solution for the minimal condition to have a certain population parameter Δ for which the network is profitable.

Proposition 2. *For a given network size N and a cost-per-unit α, it is possible to satisfy the profitability condition $\mathcal{P} > 1$ only if the following minimal requirement is fulfilled:*

$$\frac{N-\sqrt{N}-1}{\sqrt{N}(N-1)} > \alpha \tag{21}$$

Proof. The profitability index, for a linear cost curve, is given by Eq. (20). The maximum of the profitability can be obtained with the first derivative test respect to Δ, it is

$$\Delta_{max} = \frac{N-\sqrt{N}-1}{\sqrt{N}(N-1)}. \tag{22}$$

There exists a least value of N for which the network can be profitable when $\Delta_m > \alpha$, i.e., when the population heterogeneity has its maximum point at a value larger than the connection cost per generator node α. The minimal requirement of the proposition follows. □

In general, there could be an interval of values of Δ for which the average network is profitable. In any case, if we always have $\Delta < \alpha$ the network is never profitable, for that given system size N.

We defined the profitability as an expected value, but we know there are fluctuations that can make the system more unstable in terms of network performance.

Let us focus our attention on the uncertainty of the profitability index, which comes from the anomalous behavior of link fluctuations. We observe that as $N \to \infty$ the maximum profit point is for $\Delta_{max} \to 0$ which is the critical point, where there are non-vanishing fluctuations. So, for very large networks the maximum profitability is the point of minimal effort (total cost), but also the point of largest risk, where risk is defined as the uncertainty of the profitability index. In small networks, Δ_{max} is far from criticality, though the effect of uncertainty on profitability is still present, because fluctuations still exist with size $\sim O(1/N)$. In Fig. 7 we summarize these results; here the system is profitable inside an interval where the network has varying uncertainty. In fact, despite the fact that the critical point is a profitable point, the large fluctuations make the system unprofitable for certain portion of the time. This introduces a certain degree of risk that, if not considered, can lead to negative effects on the stability of the network.[4]

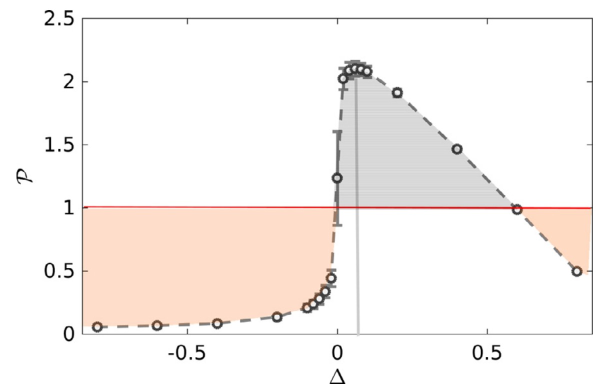

Fig. 7. Profitability of a network with $N = 100$ total nodes in a Monte Carlo simulation. The horizontal line represent the profitability threshold. For the points which are below the line the system is in an unprofitable state. Above the line, the system can be profitable since $\mathcal{P}\rangle 1$. The shaded-striped area represents the interval for which the network is profitable in this specific case of linear cost. The vertical line indicates the point of maximal profitability Δ_{min}. The error bars represent the uncertainty of the profitability since the system undergoes fluctuations due (in our mode) to the linking variance and are of the order of $\sigma(\mathcal{P}) = \sigma(\ell)N_0/2\alpha$. So, at criticality $\mathcal{P}$ has the largest uncertainty and in our example this means that the network goes through unprofitable periods during its temporal evolution. During those times the system can be more vulnerable to losses, for example, by external shocks.

In conclusion, the point of maximal profit does not necessarily coincide with the point of minimal risk. In the time series perspective, the system can end up in long time periods during which it is exposed to losses, and the whole network could become more susceptible to losses, triggering systemic crisis and possible collapses.

In terms of finite size effects as hallmarks of complexity, we showed that very large networks, while tending to a maximal benefit that happens at criticality $\Delta_{max} = 0$, can precipitate more easily a network breakdown due to the large uncertainty in the interconnectedness. Instead, for small groups, the network should make more effort to reach the maximal performance since $\Delta_{max} > 0$ but with a lower risk.

As a route to possible future applications, we note that the performance and the profitability of a network can have an important impact on the study of optimal system size. A better understanding of which would allow one to reach a given global wealth more efficiently, as well as shed light on the mechanism(s) by which large uncertainty arises as the system grows and the total connections increase. Consequently, one could rigorously prove that interconnections among network nodes are not an unalloyed good.

6. Conclusions

In this paper we presented an approximate analytical solution Eq. (11) of a minimal model belonging to a new class of agent-driven networks [8]. We exhibited a complex evolving system which starts from the most basic notion of "complexity", capturing at the very least that a system is made up of multiple interacting actors. The key feature in the network is the heterogeneity of the population since there are two types of units which are characterized by conflicting behaviors. The ones who create links and the ones who delete them as they act in the system. The control parameter Δ is the level of heterogeneity in the network, which indicates the difference between the number of units in each group. The most heterogeneous case is also the critical point of the phase transition in the system. At this point, the connectivity time series

[4] We are not considering the mechanism that makes Δ change, we rest on some external dynamics which mimics a collective society that tends to organize itself near the point of maximal profitability (highest wealth). Additionally, one can imagine that when the system is flourishing other participants want to join the system and more nodes are added to the network thus increasing the system size N. At this point, the system again moves towards the new point of maximal profitability. The result is that these mechanisms shift maximal point towards the critical point which is the point of minimal cost for the society but it is also the point of maximal risk where the network is very sensitive to systemic crises or external shocks. To make the mechanisms consistent, one should imagine a cost per unit α that changes when the system size increases; in fact, "resources" become less available and the cost becomes higher, since more effort is needed to generate connections. We arbitrarily choose $\alpha = N/4$ to be able to compare the profitability of different network sizes.

shows non-vanishingly-large fluctuations and strong correlations. In our mean-field solution we demonstrate that if we do not consider the interdependence among nodes, we decrease the importance of heterogeneity in complex networks.

Further work should be done to go beyond the mean-field approximation and to generalize the model pushing toward a new class of network formation methods which are agent-driven and not simultaneous pair connected. We also showed that using new mathematical tools analogous the quantum mechanical formalism, it is possible to solve stochastic dynamics in an alternative and sometimes more efficient way than using the standard mathematics for stochastic processes [9,10].

Finally, we depicted an abstract explanatory case in which the network shows a group effect in terms of profitability analysis in systems where the network value is related to the connectivity and the cost to the active units which create the connections. We show how small networks produce fewer connections but with lower uncertainty, while large networks have maximum profitability near the critical case with a high degree of uncertainty in the average payoff. Additional studies should focus attention on optimal system size in terms of reward in comparison to the available resources.

Simply by considering a different approach to the way units in the network connect, in this simple model we have found two fundamental ingredients of system complexity: interconnectedness and diversity. These are two fundamental concepts deeply discussed by the scientific and economic communities since the 2008 financial crisis [26–28]. These communities have been trying hard to find a relationship between systemic risk and network effects. Further investigation is necessary to improve the realism of this class of network through prediction-centered models calibrated to the details of the observed phenomena of interest. We expect that other characterizations of complex networks emerge when we observe the response of the network to perturbations and the way networks adapt and evolve during and after random disruptive events, which tend to change the network structure [29,30]. In this case, the dynamic nature of these agent-driven networks could give rise to interesting new anomalous behaviors in terms of the adaptability of complex networks to exogenous or endogenous shocks.

Acknowledgment

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 640772 - DOLFINS. The authors would like to thank ARO for support of the workshop that lead to this paper through grant W911NF-16-1-0461. DL would like to thank the ARO for support through grant No. W911NF-15-1-0245, and Welch for support through grant No. B-1577.

Appendix A. Infinitesimal time-evolution operator and master equation

The operator of Eq. (1) acts on a state of the system and returns the rate of change of that state with time, thus it deserves the name infinitesimal time-evolution operator. In the formalism described below, a state of the adjacency matrix can be decomposed into a basis in which each element has a definite value (1 or 0), the operator of Eq. (1) represented in this basis yields a master equation for the probability of a given set of values of elements in the adjacency matrix. What follows is an example of how to use this formalism in ways analogous to those used to derive Eqs. (5)–(8) in the main text.

First note that the creation $A†$ and annihilation A operators satisfy the canonical commutation relations

$$[A, A^\dagger] = 1 \qquad [A, \mathcal{O}] = 0 \qquad [A^\dagger, \mathcal{O}] = 0, \tag{A.1}$$

for all operators $\mathcal{O}$ that are not functions of A and $A†$. Where the brackets denote the lie algebraic commutator ($[A, B] = AB - BA$, for arbitrary operators A and B).

Example. Two independent states with unit growth rates.

Let the Hamiltonian analogue be

$$H = A^\dagger + B^\dagger - 2. \tag{A.2}$$

This operator describes the probabilistic increase of the occupation number of two states, A and B, and corresponds to Eq. (1). The time derivative of the expectation value of an operator $\mathcal{O}$ is given by:

$$\frac{d\langle\mathcal{O}\rangle}{dt} = \langle[\mathcal{O}, H]\rangle. \tag{A.3}$$

A basis for this space is given by states with definite occupation numbers of the two states. Letting $v^n w^m$ represent the state with occupation numbers n for the state A, and m for the state B, a general state of the system can be represented as:

$$\psi(t) = \sum_{m,n=0}^{\infty} P(n, m, t) v^n w^m \qquad \sum_{m,n=0}^{\infty} P(n, m, t) = 1, \tag{A.4}$$

where $P(n,\ m,\ t)$ is the (time-dependent) probability that the system is in the state with occupation numbers n for the state A and m for the state B. This corresponds to the fact that the probabilistic state of the adjacency matrix can be can be described as a (time-dependent) convex combination of basis states with a definite configuration of 1s and 0s.

In this basis, the operator H of Eq. (A.2) can be represented by

$$H = v + w - 2, \tag{A.5}$$

where v and w are the operators of multiplication by v and w, respectively. This operator acts on an arbitrary basis element by

$$Hv^n w^m = v^{n+1} w^m + v^n w^{m+1} - 2v^n w^m. \tag{A.6}$$

This leads to a master equation given by:

$$\frac{dP(n, m, t)}{dt} = P(n-1, m, t) + P(n, m-1, t) - 2P(n, m, t) \quad \text{where } P(-1, 0, t) = P(0, -1, t) = 0. \tag{A.7}$$

This process yields what was referred to in the text as the master equation associated with an infinitesimal time-evolution operator, and provides a connection with the standard approach to the study of stochastic processes.

Complete information on the probability distribution can be obtained via the expectation value of the characteristic function, as this is equivalent to taking the Fourier transform of the probability distribution function.

$$\begin{aligned}\frac{d\left\langle e^{i(\lambda_A A^\dagger A + \lambda_B B^\dagger B)}\right\rangle}{dt} &= \left\langle\left[e^{i(\lambda_A A^\dagger A + \lambda_B B^\dagger B)}, A^\dagger + B^\dagger - 2\right]\right\rangle \\ &= \left\langle e^{i\lambda_B B^\dagger B}\left[e^{i\lambda_A A^\dagger A}, A^\dagger\right] + e^{i\lambda_A A^\dagger A}\left[e^{i\lambda_B B^\dagger B}, B^\dagger\right]\right\rangle \\ &= \left(e^{i\lambda_A} + e^{i\lambda_B} - 2\right)\left\langle e^{i(\lambda_A A^\dagger A + \lambda_B B^\dagger B)}\right\rangle.\end{aligned} \tag{A.8}$$

The solution to this equation is

$$\left\langle e^{i(\lambda_A A^\dagger A + \lambda_B B^\dagger B)}\right\rangle = e^{t\left(e^{i\lambda_A} + e^{i\lambda_B} - 2\right)} P(\lambda_A, \lambda_B, t = 0). \tag{A.9}$$

Expanding the exponential on the RHS to second order in λ_A, λ_B yields

$$\left\langle e^{i(\lambda_A A^\dagger A + \lambda_B B^\dagger B)}\right\rangle \approx e^{t\left(i(\lambda_A + \lambda_B) - \frac{1}{2}\left(\lambda_A^2 + \lambda_B^2\right)\right)}, \tag{A.10}$$

which can be inverted in the fundamental case $P(\lambda_A, \lambda_B, t = 0) = 1$.

$$P(A^\dagger A = j, B^\dagger B = k) \approx \frac{1}{2t} e^{\frac{(j-t)^2 + (k-t)^2}{2t}} \tag{A.11}$$

Eqs. (A.8) –(A.11) correspond to Eqs. (5)–(8) in the main text, and are intended to explain the logic followed there.

In the main text we do not explicitly distinguish between the values of the operators, which are random variables, and the operators themselves, using the same notation for both. This abuse of notation is benign, since in places where the distinction matters it is possible to determine which notion is meant (and only one of the meanings is used in any given equation).

It is noteworthy that the variable $\xi = \frac{j-k}{2}$ undergoes free diffusion (Brownian motion) in this approximation (which is accurate up to discretization in the true distribution).

These rules and approximations were used to determine the Fokker–Planck equation Eq. (8) for the Generators–Destroyers model (an exact solution not being tractable in this case).

References

[1] Grigolini P. Call for papers: special issue on evolutionary game theory of small groups and their larger societies. Chaos, Solitons Fractals 2017;103:371–3. doi:10.1016/j.chaos.2017.06.020.

[2] Caldarelli G, Battiston S, Garlaschelli D, Catanzaro M. Emergence of complexity in financial networks. Complex Netw 2004:399–423.

[3] Schwarcz SL. Regulating complexity in financial markets. Wash UL Rev 2009;87:211.

[4] Bottazzi G, De Sanctis A, Vanni F. Non-performing loans, systemic risk and resilience in financial networks. Tech. Rep. Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy; 2016.

[5] Battiston S, Caldarelli G, May RM, Roukny T, Stiglitz JE. The price of complexity in financial networks. Proc Natl Acad Sci 2016;113(36):10031–6.

[6] Allen F, Gale D. Financial contagion. J Polit Econ 2000;108(1):1–33.

[7] Gai P, Haldane A, Kapadia S. Complexity, concentration and contagion. J Monet Econ 2011;58(5):453–70.

[8] Liu W, Schmittmann B, Zia R. Extraordinary variability and sharp transitions in a maximally frustrated dynamic network. EPL (Europhysics Letters) 2013;100(6):66007.

[9] Vanni F, Barucca P. Time evolution of an agent-driven network model. Tech. Rep. Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy; 2017.

[10] Bassler K, Liu W, Schmittmann B, Zia R. Extreme thouless effect in a minimal model of dynamic social networks. Phys Rev E 2015;91(4):042102.

[11] Baez J.C., Biamonte J. Quantum techniques for stochastic mechanics. arXiv:12093632 2012.

[12] Liu W, Jolad S, Schmittmann B, Zia R. Modeling interacting dynamic networks: I. preferred degree networks and their characteristics. J Stat Mech 2013;2013(08):P08001.

[13] Bassler KE, Dhar D, Zia R. Networks with preferred degree: a mini-review and some new results. J Stat Mech 2015;2015(7):P07013.

[14] Scholtes I, Mavrodiev P, Schweitzer F. From aristotle to ringelmann: a large-scale analysis of team productivity and coordination in open source software projects. Empirical Software Eng 2016;21(2):642–83.

[15] Dosi G. Economic coordination and dynamics: Some elements of an alternative "evolutionary" paradigm. Tech. Rep. LEM Working Paper Series; 2012.

[16] Dosi G, Virgillito ME. In order to stand up you must keep cycling: change and coordination in complex evolving economies. Struct Change Econ Dyn 2017.

[17] Arinaminpathy N, Kapadia S, May RM. Size and complexity in model financial systems. Proc Natl Acad Sci 2012;109(45):18338–43.

[18] Johnson NF, Jefferies P, Hui PM, et al. Financial market complexity. OUP Catalogue; 2003.

[19] Sundaram RK. A first course in optimization theory. Cambridge university press; 1996.

[20] Oliva G, Panzieri S. Modeling real networks with deterministic preferential attachment. In: Control & Automation (MED), 2011 19th Mediterranean Conference on. IEEE; 2011. p. 13–18.

[21] Metcalfe B. Metcalfe's law after 40 years of ethernet. Computer 2013;46(12):26–31.

[22] Zhang X-Z, Liu J-J, Xu Z-W. Tencent and facebook data validate metcalfe's law. J Comput Sci Technol 2015;30(2):246.

[23] Madureira A, den Hartog F, Bouwman H, Baken N. Empirical validation of metcalfes law: how internet usage patterns have changed over time. Inf Econ Policy 2013;25(4):246–56.

[24] Nagurney A, Siokos S. Financial networks: statics and dynamics. Springer Science & Business Media; 2012.

[25] Jentzsch N. The economics and regulation of financial privacy: an international comparison of credit reporting systems. Springer Science & Business Media; 2006.

[26] Haldane AG. Rethinking the financial network. In: Fragile stabilität–stabile fragilität. Springer; 2013. p. 243–78.

[27] Yellen J. Interconnectedness and systemic risk: Lessons from the financial crisis and policy implications: a speech at the american economic association/american finance association joint luncheon, san diego, california, january 4, 2013. Tech. Rep. Board of Governors of the Federal Reserve System (US); 2013.

[28] Plosser C.I. Improving financial stability: a speech at the university of chicago booth school of business, march 31, 2009. Tech. Rep.; 2009.

[29] Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. Epidemic processes in complex networks. Rev Mod Phys 2015;87(3):925.

[30] Jolad S, Liu W, Schmittmann B, Zia R. Epidemic spreading on preferred degree adaptive networks. PloS one 2012;7(11):e48686.