Vanni, Fabio; Barucca, Paolo

Working Paper

Time evolution of an agent-driven network model

LEM Working Paper Series, No. 2017/16

Provided in Cooperation with:
Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies

Suggested Citation: Vanni, Fabio; Barucca, Paolo (2017) : Time evolution of an agent-driven network model, LEM Working Paper Series, No. 2017/16, Scuola Superiore Sant'Anna, Laboratory of Economics and Management (LEM), Pisa

This Version is available at:
https://hdl.handle.net/10419/174566

INSTITUTE
OF ECONOMICS

Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM
WORKING PAPER SERIES

Time evolution of an agent-driven network model

Fabio Vanni °
Paolo Barucca *

° Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy
* Department of Banking and Finance, University of Zurich, Switzerland, and London Institute for Mathematical Sciences, London, United Kingdom

2017/16 **June 2017**
ISSN(ONLINE) 2284-0400

Time evolution of an agent-driven network model

Fabio Vanni * [1] and Paolo Barucca† [2,3]

[1] *Scuola Superiore Sant'Anna, Institute of Economics, Pisa, Italy*
[2] *Department of Banking and Finance, University of Zurich, Switzerland*
[3] *London Institute for Mathematical Sciences, London, United Kingdom*

Abstract

We present a different approach to the study of networks where the formation of links is driven by unilateral initiative of nodes. First, we propose a mathematical description of the extreme introvert and extrovert model (XIE), a dynamic network model in which the number of links fluctuates over time according to the degree-preferences of nodes. Second, we introduce a generalization of the model in which intermittent states can make the evolution of connectivity slower. In this class of networks, system evolution has nodes which can either create or destroy links according to their target attitudes. This model belongs to a class of networks not based on the rationale of linking probability between pair of nodes, but it is based on the initiative of the single units when they act. We select the minimal case of a bipartite network where we have two groups of nodes, each group has nodes with a given capability to bear links. One group (high-target) is composed by nodes that create as many links as possible, the extroverts. The other group (low-target) is composed by nodes that delete as many links as possible, i.e. the introverts. We here provide a novel analytical formulation of the system evolution through coupled Master Equations for the two interacting populations, recovering the steady state degree distributions and a new analytical description of the transient dynamics to the equilibrium. Moreover, we provide numerical evidence supporting the existence of an extreme Thouless effect in this model, i.e. a mixed-phase transition at which the order parameter, here given by the network connectivity, displays a discontinuous jump and large dynamic fluctuations. Fluctuations are also shown to be connected to a peak in degree correlation at the transition, corresponding to same size populations of extroverts and introverts.

Keywords: Network formation, temporal evolution, stochastic process, financial networks, critical transitions
JEL codes: D85, C62, G21

Introduction

The conceptual framework of network formation can well describe complex systems which exhibit an intricate structure of evolving connections among the units.

*E-mail: fabio.vanni@santannapisa.it. Corresponding author.
†E-mail: baruccap@gmail.com

The vast majority of the scientific literature on network formation has focused on models which are generated through an attachment probability approach where two random nodes have a certain bilateral chance to be connected, generally depending on both nodes. The Erdos-Renyi model based on a uniform linking probability, the Albert-Barabasi model [1] and the class of preferential attachment models, and the generalized random graphs known as hidden-variable or fitness models [5] adopt this kind of rationale. In the present paper we start from a different approach introduced by [11], where the system dynamically evolves according to the unilateral preferences of each agent. Passing from a pair-wise approach to a unit-wise one. This simple difference actually introduces a deep change in the network structure and its dynamics with emergent properties that arises for certain values of simple network parameters. In particular, the extreme introverts and extroverts model (XIE) has been shown [3] to be characterized by a critical behavior beyond Ehrenfest classification of phase transitions: the discontinuity displays an extreme Thouless effect, where both a discontinuous jump in the order parameter, and large dynamic fluctuations appear, two features of first order and second order phase transitions respectively.

Despite its recent introduction, unit-driven mechanisms are in some cases more realistic than pairwise ones. For instance, in social networks like Twitter, the decision over following another user is strictly unilateral, and the ratio between followers and followed users, i.e. out-coming and in-coming links, strongly characterizes the behavior and attitude of users [8].

Another interesting evidence for the presence of different agents attitudes which drives the dynamics of a systems can be largely found in financial networks analysis [6, 9, 10, 2], where various financial institution in the interbank market can be divided into different communities related to their financial management.

In recent years, financial systems have represented an important field of application for the network science, which naturally captures the linkage architecture of the financial agents and their relations. The interest has increased even more as consequence of recent financial crisis of the years 2007-2009, and a network approach is particularly crucial for assessing financial stability, since it is possible to have models and methodological tools which could describe, detect and eventually mitigate systemic risk [4]. There are different perspectives form which it is possible to address financial network analysis, mainly the research has focused on the one hand on topological structures and the dynamics of network formation, and the dynamic processes occurring on the network on the other.

Here, we study a unit-driven model of network formation through a new mathematical description based on dynamical Master Equation for the degree of the networks. The solution we found provides both the transient non-equilibrium dynamics and the steady state of the degree distribution. At the same time we recover the crucial emergent property with a phase transition behavior as the proportion between the two interacting groups changes [3].

Moreover, we investigate the limits of such approach and shed lights on a new important feature of such network that is the emergence of strong correlations among the agents when the system is at criticality, connected with the extreme Thouless effect already shown in the XIE model.

The paper is organized as follows: in Section 1 we introduce the unit-driven model and describe in detail its rationale and dynamics. In Section 2 we outline the new coupled master

equation dynamics (CMED) for the degree-distribution of extroverts and introverts, explaining how it is possible to couple them consistently. Further we present numerical results on the non-equilibrium dynamics, showing how CMED succeeds in predicting the full dynamical evolution of degree-distributions in the unit-driven model outside of the criticality. At the criticality degree correlations are shown to be strongly affect the CMED approach. In Section 3, we introduce a more realistic version of the agent-driven network based on XIE model where intermittency is present in the degree trajectories. Finally in Section 4 we summarize results, stressing the need for a robust methodology to model degree correlations and their role in mixed phase transitions and suggesting the possible applications of the model - or similar generalizations - in the context of systemic risk and financial stability.

1 Bipartite target network model

The network model is based on an intergroup dynamics where individual agents connect according to a degree target, i.e. tending to reach a preferred number of neighbors. In the simplified and extreme situation the nodes in the network are divided between a group of link generators and a group of link destroyers. The dynamics is defined as follows: a single agent is randomly chosen and it acts according to its nature (attitude) so pushing towards its target (desired) degree. At this level, each agent can add or create a connection per time, and the network is made up of only two types of agents in a bipartite graph framework: high-target agents are able only to generate links (when possible) and the low-target agents which can only destroy links among its neighbors (if any). In this extreme view, the agent in action overbear the attitude of chosen neighbor, which at its turn it will be able to act towards its preferred degree. This new perspective is crucial for the nature of fluctuations on the links dynamics as well as in the arising of emergent properties. The generators are agents with an infinite target degree, and destroyers are agents with a zero target degree. Moreover, links are bi-directional, and there can be at most one link between any two units. Neither self-links nor intragroup links are considered. The generators create links as long as there is at least one destroyer without a link. Destroyers remove links until there is at least one generator with a link. A node which has reached its preferred degree does not act in the network. Another difference with the model described in [3], is that we just study the bipartite version of the network, since the high target nodes if involved in the creations of links inside their groups gives only a fixed and constant values, not contributing to the dynamics of the network.

Let us define a bipartite graph as as a triplet $G = (\top, \perp, E)$ where $\top$ is the set of high-target nodes (extrovert units) and $\perp$ is the set of low-target nodes (introvert units), and $E \subseteq \top \times \perp$ is the set of links that can vary in the domain $[0, |\top| \times |\perp|]$ where $|\cdot|$ is the number of elements of the set.

Then, the time evolution of the this bipartite graph is an ordered sequence of M graphs $\{G_1, G_2, \ldots, G_M\}$ defined over N nodes, where each graph G_m in the sequence represents the state of the network, i.e., the configuration of links in the time-window $[t_m, t_m + \Delta t_m]$, $m = 0, \ldots, M-1$.

The time-varying graph G is fully described by means of a time-dependent adjacency matrix $A(t_m), m = 0, \ldots, M-1$, where $a_{ij}(t_m)$ are the entries of the adjacency matrix of the graph at

time t_m. Tuning the size of the time window used to build each snapshot, we obtain different representations of the system at different temporal scales.

In particular, in the limit case when $\Delta t \to 0$, we obtain an infinite sequence of graphs, where each graph corresponds to the configuration of connections at time instant t.

On the contrary if we set $\Delta t = T$, the time-varying graph degenerates into the corresponding aggregated graph, loosing all the temporal information and obtaining a static graph.

In order to write an equation of the degree distributions of the graph we split the analytical description in terms of the two groups.

In principle we should know at each time step the configuration of the network $\mathcal{N}$ described for example by the adjacency matrix. We actually study the degree distribution defined as:

$$\rho(k,t) = \sum_{\{\mathcal{N}\}} \sum_{i} \delta(k - k_i) P(\mathcal{N}, t). \tag{1}$$

where δ is the Kronecker delta, and $P(\mathcal{N}, t)$ is the probability to find our network in the configuration $\mathcal{N}$ at time t, where each node has degree k_i.

In our particular case, we expect to have a master equation that takes in account the add-cut links prescription of the target-model in which the equation to find the $\top$ system in the state k (the degree) at time $t+1$ is:

$$\begin{aligned}
&\rho_{\top}(k, t+1) = \rho_{\top}(k,t) + \\
&+ \text{Prob}_t(\text{pick a } \top \text{ node with degree } k-1) \\
&- \text{Prob}_t(\text{pick a } \top \text{ node with degree } k) \\
&- \text{Prob}_t(\text{pick a } \perp \text{ node cutting a link with one of } \top \text{ nodes with degree } k) \\
&+ \text{Prob}_t(\text{pick a } \perp \text{ node cutting a link with one of } \top \text{ nodes with degree } k+1)
\end{aligned}$$

since high-target nodes can only generate links and low-target nodes can only destroy links.

It useful to describe the model in a more simplified version in order to obtain a solvable macroscopic equation. First let us define the average number of high-target nodes as

$$\langle n_1(k,t) \rangle := \rho_{\top}(k,t) \cdot N_1$$

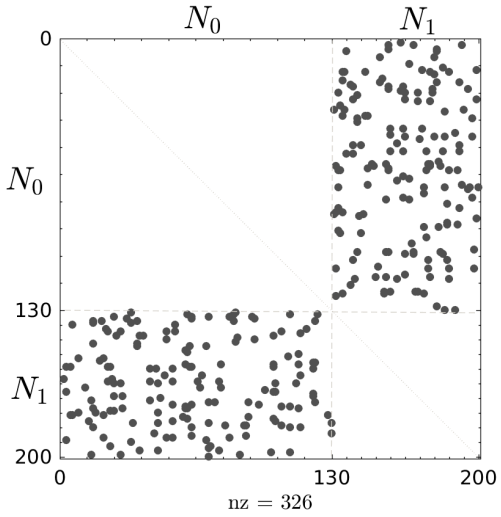
, and the number of low-target nodes as

$$\langle n_0(k,t) \rangle := \rho_{\perp}(k,t) \cdot N_0$$

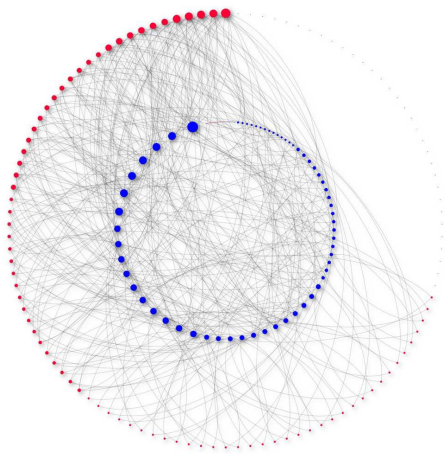
where $N_1 = |\top|$ and $N_0 = |\perp|$, and $N_1 + N_0 = N$ which indicates the total number of nodes in the whole network which is constant.

In the minimal case of two extreme groups, after a transient the network, dynamics of links between different groups is fluctuating. So in the steady state the dynamics of the network is fully described by the bipartite graph between the two groups of agents. So the bipartite model fully describes the evolution of cutting-adding of links as shown in Fig.1, where a bipartite structure is clear for off-diagonal blocks in the adjacency matrix.

In this regime and using the mean field approximation it is possible to write the evolving



(a) Snapshot at a given time of the adjacency matrix of the bipartite target network. Off-diagonal blocks are symmetric since we do consider links to be undirected, total number of links are $nz = 326$.



(b) Graphical representation of the bipartite target network at a given time, inner circle nodes are the high target nodes N_1. Node sizes in the graph is proportional to their degrees.

Figure 1: The temporal bipartite target network ,at a given time step, has a typical adjacency matrix (a) where only the off-diagonal blocks are involved in the dynamics. The same bipartite structure can be visualized in (b) as a core (high target nodes) represented by blue circles and the periphery (low target nodes) represented by red circles. Arcs are the links between the two groups made up of $N_1 = 130$ and $N_0 = 70$.

network in terms of birth-death process with time-dependent transition rates. However, the system always reach a steady state, since it is a finite and closed [3].

In the bipartite version, let us describe the change of states of the network from the "point of view" of the high-target degree nodes, N_1. Let $\langle n_1(k, t_n)\rangle$ the average number of high-target degree units with degree k at the time step t_n. It is possible to write the following approximated rate equation:

$$\langle n_1(k, t_n + 1)\rangle = \langle n_1(k, t_n)\rangle + \tag{2}$$

$$+ \frac{N_1}{N}\frac{\langle n_1(k-1, t_n)\rangle}{N_1} - \frac{N_1}{N}\frac{\langle n_1(k, t_n)\rangle}{N_1} + \tag{3}$$

$$- \frac{N_0}{N}\frac{k\langle n_1(k, t_n)\rangle}{N_0}\frac{\epsilon}{k_0} + \frac{N_0}{N}\frac{(k+1)\langle n_1(k+1, t_n)\rangle}{N_0}\frac{\epsilon}{k_0} \tag{4}$$

The right side gives the combination of factors for which we have $\langle n_1(k, t_n + 1)\rangle$ average number of high target nodes with degree k at time step $t_n + 1$.

- The first row eq.(2) are the average number of units at the previous time step.
- The second row eq.(3) represents the chance for gaining or loosing a connection due to the fact that one high-target unit has been randomly drawn among N, this happens with a frequency of N_1/N. Then there is the chance to pick a high-target node with a certain degree with frequency $\langle n_1(k, t_n)\rangle/N_1$.
- The third row eq.(4) represents the chance for gaining or loosing a connection due to the fact that one low-target unit has been randomly drawn among N total nodes in the

network. This happens with a frequency of N_0/N. Then each high-target node has k links connected to k different (no-multilinks) low-target nodes. So the rate to pick one of this low-target-node is k/N_0 and this must be independently replicated for other $\langle n_1(k,t_n)\rangle$ high-target nodes. At this point, the drawn low-target node will cut one over all their total neighbors and we need to know how many links that particular low-target node has, and how many connections among its neighbors belong to one over the $\langle n_1(k,t_n)\rangle$ high-target nodes. Due to the difficulty of this last task, we use a mean-field approximation where we imagine the probability to cut one k-high-target nodes is $1/k_0$ where $k_0 = \langle k\rangle_0$ represents the average degree of low-target nodes. In principle we should take only those low-target units which have at least one connection (not isolated nodes). The correction factor takes in account this issue and considers the fact that there is a decrease of the cutting rate when an isolated low-degree node has been picked [1].

The process description in eq.(2)-(3)-(4) is valid for $k = 1 \dots N_0 - 1$, and the boundary conditions at $k = 0$ and $k = N_0$ are:

$$\langle n_1(0,t_n+1)\rangle = \langle n_1(0,t_n)\rangle - \frac{\langle n_1(0,t_n)\rangle}{N} + \frac{\langle n_1(1,t_n)\rangle}{Nk_0}\epsilon \tag{5a}$$

$$\langle n_1(N_0,t_n+1)\rangle = \langle n_1(N_0,t_n)\rangle + \frac{\langle n_1(N_0-1,t_n)\rangle}{N} - \frac{N_0\langle n_1(N_0,t_n)\rangle}{Nk_0}\epsilon \tag{5b}$$

In order to get the full description of the model, we have to know the number of empty low-target nodes where in the master equation appears as:

$$\epsilon = \frac{N_0 - \langle n_0(0,t_n)\rangle}{N_0} \tag{7}$$

that is the rate to not pick an empty low-target node over the total N_0 low-target nodes.

At this point we can evaluate $\langle n_0(0,t_n)\rangle$ directly from a correspondent master equation for the low-target nodes with a symmetric reasoning. We define as "shadow links" the total number of possible links minus the actual links. In some sense they are the "imaginary" links that are possible in our network. In this perspective low-target nodes are generators of shadow-links and high-target nodes are destroyers. In particular we have that if we have a total number of L links in the networks, then the shadow-links are $N_0N_1 - L$. Per each low-target nodes, the shadow degree is $s = N_1 - k$ with $k = 0 \dots N_1$.

It is possible to symmetrically write the master equation for low-target nodes in terms of

[1] The third row eq.(4) can be also seen in terms degree distribution of the neighboring node for uncorrelated networks $P_0(q) = \frac{k\langle n_1(k,t)\rangle}{N_1\langle k\rangle_\infty}$, that is the frequency with which end nodes of a randomly chosen link have degree q. It is the probability that among the high-target nodes, a neighbor of the selecetd low-target node has degree k (i.e. assuming neighbor degrees are not correlated). After that we do need to multiply by the rate to select a low-target node N_0/N and the correction factor to not pick an empty node ϵ.

shadow links as:

$$\langle n_0(s,t_n+1)\rangle = \langle n_0(s,t_n)\rangle + \tag{8}$$

$$+\frac{N_0}{N}\frac{\langle n_0(s-1,t_n)\rangle}{N_0} - \frac{N_0}{N}\frac{\langle n_0(s,t_n)\rangle}{N_0} + \tag{9}$$

$$-\frac{N_1}{N}\frac{s\langle n_1(s,t_n)\rangle}{N_1}\frac{\tilde{\epsilon}}{s_1} + \frac{N_1}{N}\frac{(s+1)\langle n_0(s+1,t_n)\rangle}{N_1}\frac{\tilde{\epsilon}}{s_1} \tag{10}$$

The average number of high-target s-isolated nodes (full degree nodes) is indicated with $\langle n_1(k = N_0,t_n)\rangle$. The dual correction factor can be taken as

$$\tilde{\epsilon} = 1 - \frac{\langle n_1(k=N_0,t_n)\rangle}{N_1} = 1 - \frac{\langle n_1(s=0,t_n)\rangle}{N_1} \tag{11}$$

the boundary conditions are:

$$\langle n_0(s=0,t_n+1)\rangle = \langle n_0(0,t_n)\rangle - \frac{\langle n_0(0,t_n)\rangle}{N} + \frac{\langle n_0(1,t_n)\rangle}{Ns_1}\tilde{\epsilon} \tag{12a}$$

$$\langle n_0(s=N_1,t_n+1)\rangle = \langle n_0(N_1,t_n)\rangle + \frac{\langle n_0(N_1-1,t_n)\rangle}{N} - \frac{N_1\langle n_0(N_1,t_n)\rangle}{Ns_1}\tilde{\epsilon} \tag{12b}$$

where we have the fundamental relation for the interdependent average degrees as:

$$s_0 = N_1 - k_1\frac{N_1}{N_0} \tag{13}$$

At this point the model can be described by the coupled master equations of the two interacting populations.

2 Statistical properties and degree dynamics of the model

The master equation for the degree distribution can be recovered from the rate equation considering that $\rho_1(k,t_n) = \langle n_1(k,t_n)\rangle/N_1$ where $\sum_{k=0}^{N_0}\langle n_1(k,t_n)\rangle = N_1$, $\forall t_n = 0,1,2,\ldots$:

$$\begin{aligned}\rho_1(k,t+1) = \rho_1(k,t) &+ \\ &+\Gamma^+[k-1]\,\rho_1(k-1,t) - \Gamma^+[k]\,\rho_1(k,t) + \\ &-\Gamma^-[k,t]\,\rho_1(k,t) + \Gamma^-[k+1,t]\,\rho_1(k+1,t)\end{aligned} \tag{15}$$

that represents a general non-homogeneous birth-death process that describes the add-cut link procedure in the network. The transition probability $\Gamma^+[k-1]$ is the birth rate from the state $k-1$ to k, and the transition probability $\Gamma^-[k+1,t]$ is the time-dependent death rate from the state $k+1$ to k.

From the "point of view" of the high-target units, the transition rates are can be written as:

$$\Gamma^{+}[k-1] = \frac{N_1}{N}\frac{1}{N_1} \tag{16}$$

$$\Gamma^{-}[k,t] \approx k \cdot \frac{N_0}{N}\frac{1}{N_0}\frac{\epsilon}{\langle k\rangle_0} \tag{17}$$

The equation follows the reflecting boundary conditions:

$$\Gamma^{+}[k = N_0] = 0 \tag{18}$$

$$\Gamma^{-}[k = 0] = 0 \tag{19}$$

which takes in account that a node cannot have less than zero link, and it has the limitation of a full connected node.

The master equation of the degree distribution remind of a generalize birth-death process where the rate are time dependent. Following the authors [3] the systems reach a state where the transition rates become homogeneous recovering the detailed balance condition property. Anyway we use only the fact that the master equation is restricted under reflecting boundary conditions within a closed systems, that is enough to guarantee the existence of a stationary distribution regardless of the number of links at the beginning of the process [7]. In our simulations we took the initial network to be empty, i.e. $n_1(k, t = 0) = N_1\delta_{0,k}$. So the only parameter that has an impact on the statistical structure of the network is the proportion between the two populations.

Finally let us consider that this kind of description is based on a mean-field approximation based of $\langle k\rangle_0$ for which we expect to have a good approximation on the first order of the degree-distribution, for higher order moments it would be necessary to more considerations on the cut rate probability.

2.1 Degree distribution and average degree

As results of our mathematical interpretation of the model we compare the degree distributions obtained by the Montecarlo simulation of the model with the distributions obtained as numerical results of the coupled master equations. As expressed in the appendix A, the distributions obtained by the master equation approach are bounded poisson distributions.

According to the values N_0 and N_1 the network behaves differently, so we define as control parameter of the system the quantity

$$h = \frac{N_1 - N_0}{N_1 + N_0}$$

that is the proportion variable between the two groups.

In Fig.2 the degree distributions of the network is plotted for several values of N_1 and N_0. It is evident as the model prediction is exact when considering the first moment of the distribution (the expected value of the degree). This is consequence of the first order approximation made in the master equation representation with a mean field approach. As regards with the entire

degree distribution it is exact for all the values of h but when we are at the values for which $N_1 = N_0$: here the degree distribution departs from the poisson distribution and and it is closer to a fatter distribution which can be represented as a normal distribution with mean equal to the exact average degree $\mu = \langle k \rangle_1$ but variance of the order of the support of the distribution $\sigma^2 \sim N_0^2$.

All these results are in line with the solutions obtained in [3] via numerical convergence of a fixed-point procedure; the consequent degree distributions have been compared numerically with the master equation solution. Compared to the true model distributions we obtained the same Kullback-Leibler divergence up to a factor of 10^{-6} as analyzed in the appendix A.

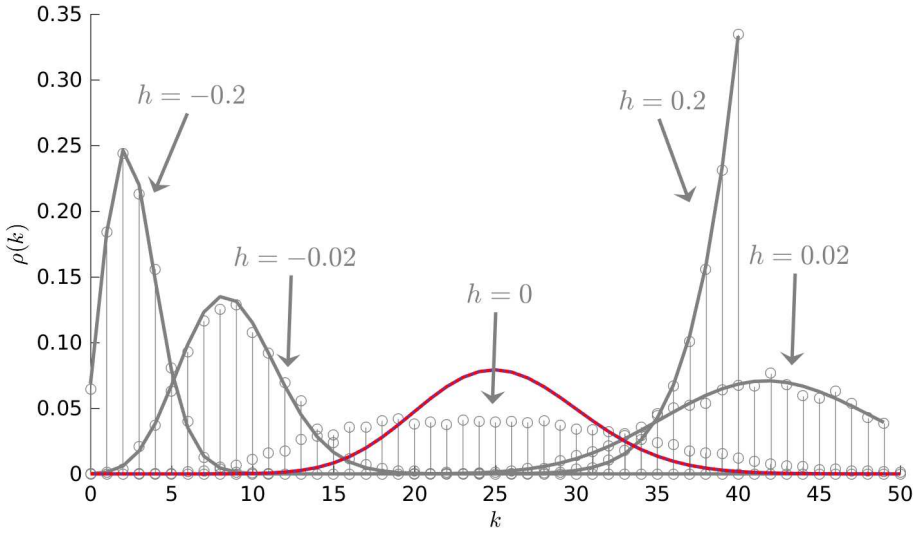


Figure 2: In a network of $N = 100$, with ensemble averages of 1000 samples. Where $h = (N_1 - N_0)/N$ is the population control parameter. The stem plot represents the degree distribution from the model simulations of the target network, the continuous gray lines are degree distributions obtained directly from the coupled master equations. The two distributions always coincide on the shape and are centered exactly at the average degree values. An exception happens for populations such that $N_1 = N_0$ ($h = 0$), here the approximated distributions (dashed line) is less fat then the real distribution. This is a direct consequence of the mean-field approximation made in the master equation representation of the model.

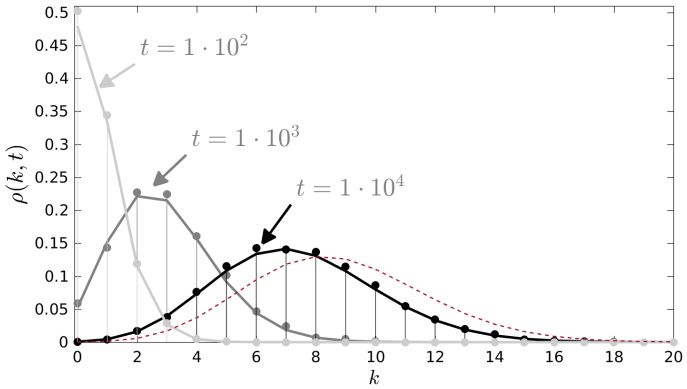
Using the interacting master equations we can also obtain and plot the temporal evolution of the degree distributions and consequently the average time needed to reach the stationary distributions, as well illustrated in Fig.3.

The same type of results are obtained in the evaluation of the number of empty low target nodes and the full high target ones, see Fig.4.

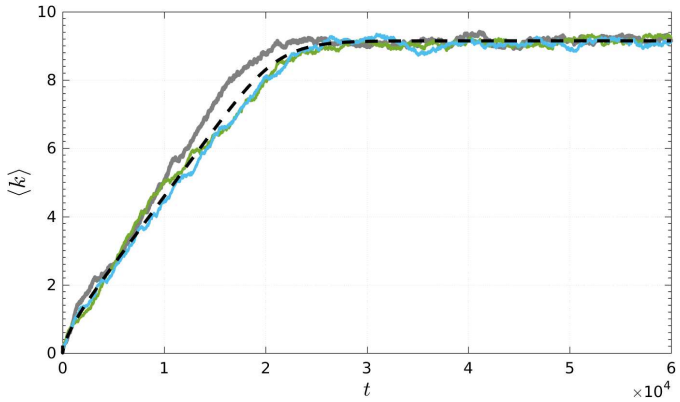
Once we obtained numerically the value of the average degree and the number of empty nodes we can recover the shape of the degree distribution to be a bounded poisson distribution as shown in the appendix A:

$$\rho_1(k) = \frac{1}{e^{\chi} \frac{\Gamma(N_0+1,\chi)}{\Gamma(N0+1)} - 1} \frac{\chi^k}{k!} \qquad , \; k = 0 \ldots N_0 \tag{20}$$

where $\chi = \frac{N_1}{N_0 - n_{00}} k_1$, and $n_{00} = \langle n_0(0, t_n) \rangle$ is the average number of empty low-target nodes.



(a) Stem plot distributions are from the montecarlo simulations of the model, the continuous lines are the distributions from the master equation. Different time snapshots shows the agreement between the two approaches. The dashed curve represent the stationary distribution.



(b) The three lines are three different realizations of the model versus the black dashed line representing the time-evolution of the average degree in the master equation numerical solutions. The dotted curve represents the stationary degree distribution.

Figure 3: Time evolution of the degree distribution (a) and the average degree (b) of the bipartite network simulations and its mathematical counterpart of coupled master equations for the case of $N_1 = 51, N_0 = 49$ where $h = -0.02$. We notice how good is the approximations both maintaining the distribution shape both detecting the right average convergencing time towards the equilibrium distribution.

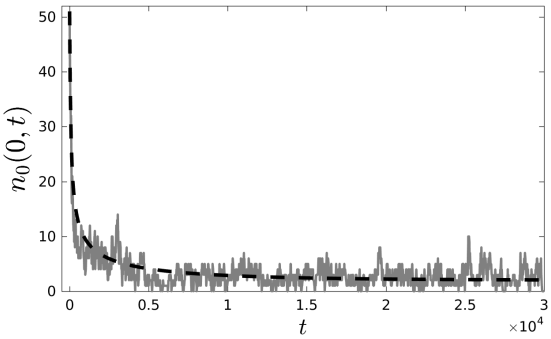
2.2 Phase transition and criticality

Other than the degree, we can better characterize the network formation defining the *effective degree*:

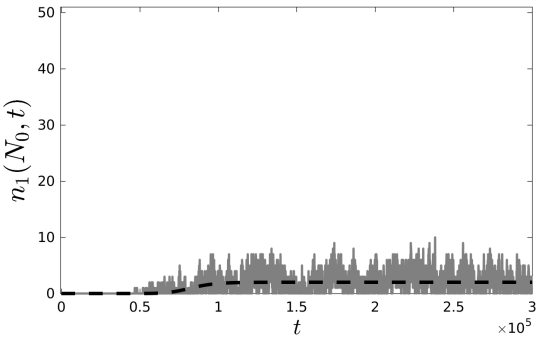
$$\xi = \frac{k}{N_0}, \quad \in [0, 1] \tag{21}$$

which is the normalized degree; it measure the number of links created by an high-target node over the total number of possible links it can make (i.e. N_0). The average over all the high-target nodes gives just the network degree *efficiency* $\langle \xi \rangle$: values close to 0 indicates that the high target nodes create very few links respect to the possible links they can make. On the opposite, values of the efficiency close to 1 indicates an high capability for each high target node to generate links with the other group nodes. The passage between these two extreme situations exhibit a typical phase transition where we have an abrupt change of efficiency when $N_1 = N_0$.

We can plot and reproduce the phase transition in Fig.5. Exploring the average value of



(a) Time evolution for the number of low-target empty nodes towards the equilibrium value $n_{00} = N_0 - N_1 = 2$ since $N_0 = 51$ and $N_1 = 49$.



(b) Time evolution for the full high target nodes towards the equilibrium value$n_{11} = N_1 - N_0 = 2$ since $N_0 = 49$ and $N_1 = 51$.

Figure 4: In a network with $N = 100$ total number of nodes, two cases: (a) the convergence of empty low-target nodes in the case $N0 = 51$, and (b) the convergence of full high-target nodes in the case $N0 = 49$. Gray noisy curves are the montecarlo simulation of a single trajectory of the model, the black dashed lines are the estimations obtained by the master equations. For both cases an empty network is the initial condition. This explains the apparent different convergence times in the two cases.

the mean field variable ξ for all the possible values of the population in the two groups, we have a phase transition, where the critical point is the case $N_1 = N_0$, in which we have a big jump from an minimal-efficient network to a maximal-efficient network. The figure shows as the numerical solution of the coupled master equations approach gives an exact behavior of the phase transition which accounts only for the first moment of the distribution.

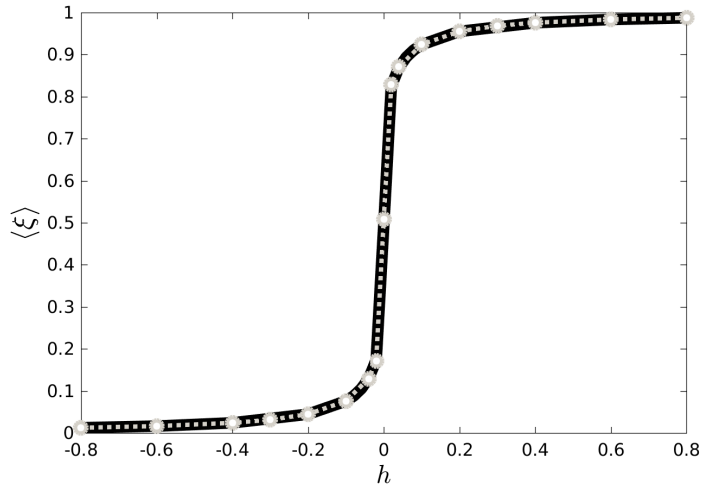


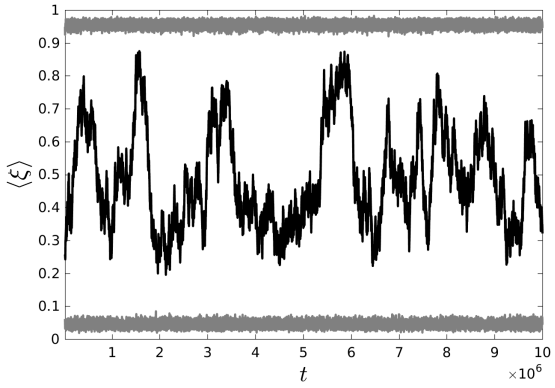
Figure 5: Phase transition of the mean value of the mean-field variable $\langle \xi \rangle$ (the network efficiency) over the possible values of N_1 and N_0 in a bipartite network with $N = 100$ total nodes.The bold black line is the curve corresponding to the model simulation, the dashed white line is the curve made up using the coupled master equations. The plot suggests the idea that, in a target-network, for a node being part of a group endogenously increase (o decrease) its ability to follow its preference.

Let us notice that the phase transition is a property of the network efficiency and not of the average degree by itself. Anyway they share the same emergent properties mostly based on a second order effect related to the variance of the average degree and the nodes correlation in the network.

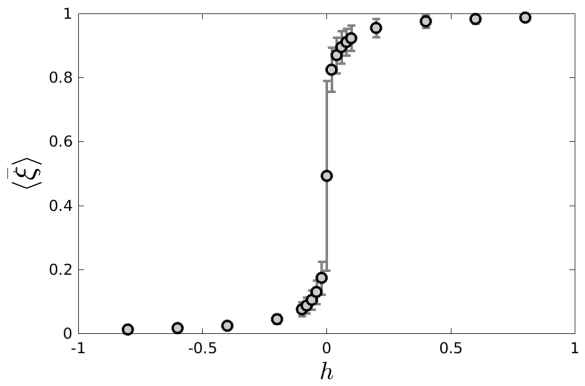
2.3 Second order effects and emergent properties

Despite the presence of the phase transition, the emergent property of the target network resides in higher moments of the degree distribution, those which are not satisfactorily addressed when the network is at criticality. The first important emergent property is the variance of the degree distributions, which exhibits an abrupt large value at criticality ($h = 0$). We want to give some heuristic explanation of this phenomenon.

The stochastic process associated to the the network dynamics can be also analyzed through the time series of the average degree as in Fig.6, which shows a given mean value and a standard deviation.



(a) Temporal behavior of the average degree or equivalently of the degree efficiency $\langle \xi \rangle = \langle k \rangle_1 / N_0$ where $N = 100$: the black central line represent the mean-field parameter ξ in the equal proportion case of $N_1 = N_0 = 50$ in which we have a "quasi" diffusion process which explores all the the states in the degree domain. The top gray line is the case $N_1 > N_0 = 45$, and the bottom gray line is for the case $N_1 < N_0 = 55$. In these two cases the brownian motion is attenuated by the reflecting boundary condition, resulting in much smaller fluctuations.



(b) The usual phase transition of the degree efficiency but this time we plot twice the standard devition of the relative time series. It is evident the huge amplitude of the fluctuations at criticality, so defining the onset of an emergent propery of the network when the systems is at the critical point $N_1 = N_0$.

Figure 6: The average degree and the relative efficiency variable can also be described in terms of their time series along the simulation time as in (a). This is equivalent to evaluating several times the average degree for different trajectories at a given time. The resulting variable is $\overline{\langle k \rangle}$ that is mean degree, in (b) we also plotted the error bars which corresponds to the standard deviation of average degree along the time.

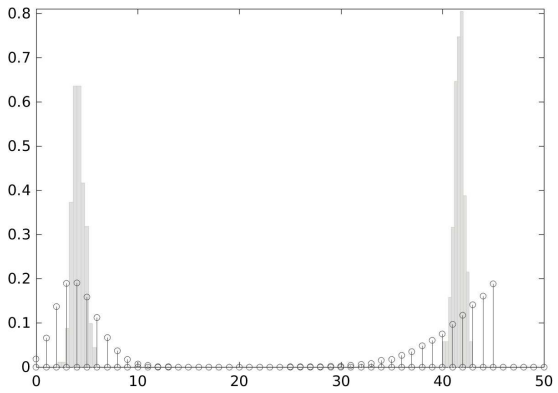
Given the ergodicity of the process we can freely interchange the mean values over time and the average over different trajectories. Finally we can build the distribution of the average degree and the degree distribution. We do plot those probabilities on Fig.7.

The two distributions, as their moments, are related by convolution of the distributions of each node, since $\langle k \rangle_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} k_i$, defining a sort of a sampling distribution for the mean. The expected value results to be

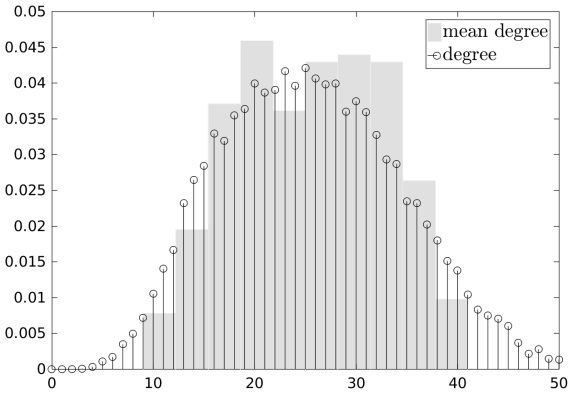
$$E[\overline{k}] = \langle k \rangle \tag{22}$$

and the variance is:

$$Var(\overline{k}) = \frac{\sigma^2}{N_1} + \rho \frac{N_1 - 1}{N_1} \sigma^2 \tag{23}$$



(a) The two distributions on the left are for $N_1 = 45, N_0 = 55$ in the sub-critical case, the two right distributions are for $N_1 = 55, N_0 = 45$ in the super-critical regime.



(b) The special case of criticality ($N_1 = N_0$), where the two distributions are both fat with same mean and different variance

Figure 7: comparison between two distributions: stem black lines indicates the ordinary degree distribution $\rho(k)$ and the gray bars indicate the sampling distribution of the average degree. In (a) there is the case in which $N_1 \neq N_0$ and in (b) the case $N_1 = N_0$ at criticality.

where σ^2 is the variance of the degree distribution, and the $Var(\bar{k})$ is the variance of average degree distribution (the sampling distribution). If the nodes were independent, the average correlation factor ρ would be zero. Actually from the empirical studies of the model we can estimate the value of the correlation factor ρ, just comparing the variances of the two distributions evaluated via simulations

$$\rho = \frac{Var(\bar{k})}{\sigma^2} \frac{N_1}{N_1 - 1} - \frac{1}{N_1 - 1} \tag{24}$$

which for large networks the average correlation ratio between the average of the sample mean distribution over the variance of the degree distribution. What is obtained is well expressed in Fig.8, where strong correlations arise among the nodes at criticality i.e. $N_1 = N_0$. This is an evidence on how it is not possible ignore the presence of nodes correlation in the mathematical description of the model. For example, the model prescription prevents the node to go beyond its capacity, so that if an low-target node is already full, an high target node is forced to exclude it from its candidate neigbors. On the contrary, the master equation which describes the network dynamics has been addressed using a first-order mean field approximation. In evaluating the cutting rates we only consider first-order effect ingoring any possible correlation among the nodes in selecting the nodes to connect with.

Despite these limitations of the master equation representation, we can reach the conclusion that another important emergent property of the bipartite target network is the onset of strong correlations among the agents at criticality.

As important variable in the network, we consider the total number of links in the network, which we call network linkage, in order to distinguish it from the related to the degree of nodes. These variable are connected though the relation:

$$L \equiv \langle k \rangle_1 N_1 = \langle k \rangle_0 N_0 \tag{25}$$

that is the total number of links in the bipartite network.

The phase transition property and the temporal evolution behaviors of the linkage are iden-

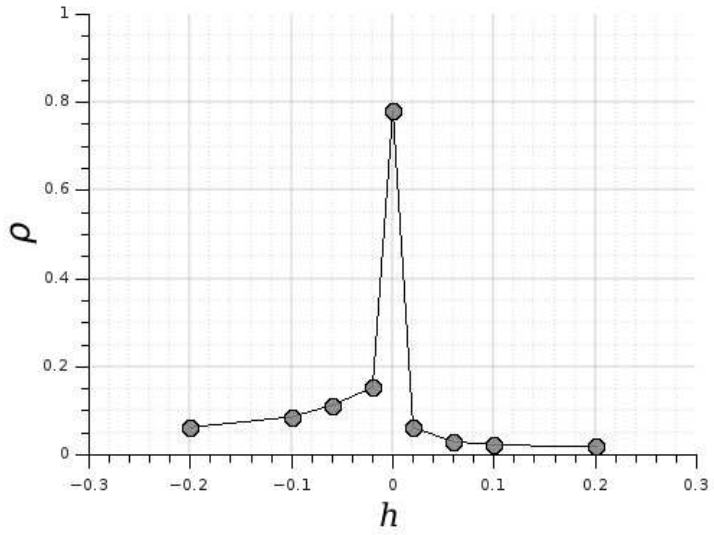


Figure 8: Average correlation estimation obtained comparing the sampling distributions of the average degree with the ordinary degree distributions. It is evident how at criticality ($h = 0$) the correlation is much higher then in the off-critical cases ($h \neq 0$). This shows how neglecting the correlation among the nodes make the master equation approach fail at criticality.

tical to the average degree case since they are proportional, so sharing the same statistical properties as shown in Fig.9.

In order to capture the efficiency of the network in the linking creation let us call *connectance* ℓ the number of actual edges, expressed as a proportion of the total number of potential edges:

$$\ell = \frac{L}{N_0 N_1} \tag{26}$$

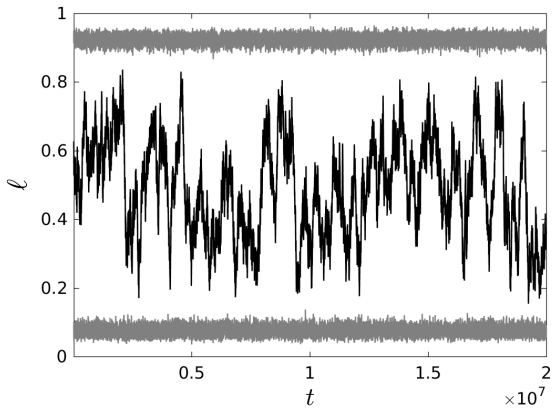
$$= \frac{\langle k \rangle_1}{N_0} \tag{27}$$

i.e. the links density as the ratio between the actual number of links created for that particular populations sizes and the total possible links available (bipartite all to all case). Anyway we treat connectance as a temporal variable, and becasue of the ergodic property of the process, we use time averages in the place of the ensamble averages for the total number of links L and the connectance ℓ.

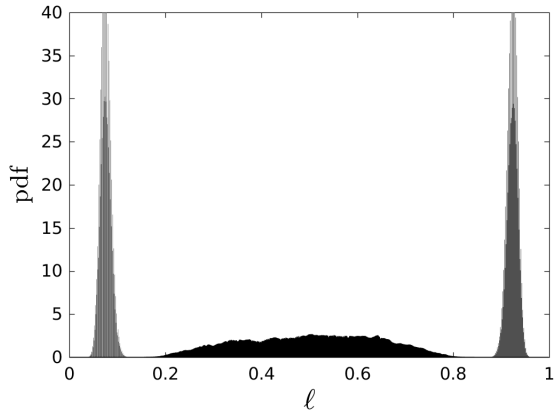
3 A more general model: intermittent CMED and connection velocity

As an extension of the previous model, we provide an more plausible agent-driven network based on the XIE model where the unilateral initiatives are smoothed by the possibility that the unit does not manage to act according to its attitude of preferred degree.

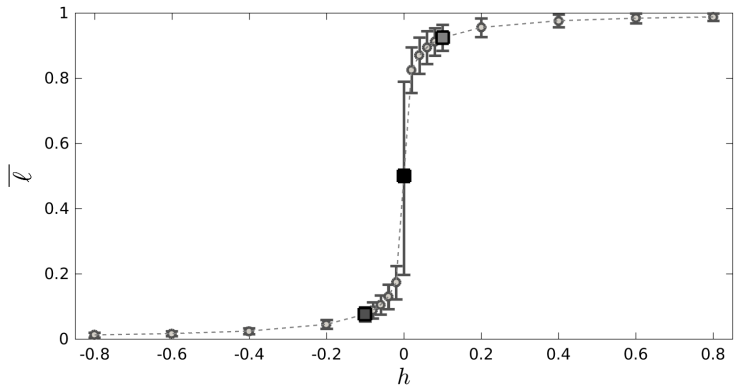
Specifically, let us take in account the chance that the action of the agent to generate (or destroy) a link is not accomplished. So that, for several reasons, the connection (or deconncetion) attempt fails, for example because of the opposition of the counterpart agent. So whenever a unit is drawn, it tries to act according his preference, adding or cutting a link, but



(a) Time series of the linkage (i.e. link density) for three different cases: top gray series for $N_1 = 55, N_0 = 45$ that is the over-connected network. The down gray series is the case $N_1 = 45, N_0 = 45$ the minimal-connected network, the black time series represent the critical point $N_1 = N_0$.



(b) Three examples of link distribution strictly relate to the previously discussed average degree distribution. The center flatter histogram corresponds to the critical case.



(c) Time series of the connectance (i.e. link density) for three different cases: top gray series for $N_1 = 55, N_0 = 45$ that is the over-connected network. The down gray series is the case $N_1 = 45, N_0 = 45$ the minimal-connected network, the black time series represent the critical point $N_1 = N_0$.

Figure 9: We plot the connectance $\ell = L/(N_1 N_0)$, namely the link density of the network. From the side of the toal links we get the same statistical behavior of the average degree of the groups. Usually the link dynamics can be chosen as time series. In (a) we plot three different typical temporal dynamics in the subcritical regime, in the critical point and in the supercritical phase. The same three regimes are studied from the point the distribution perspective, observing the exaggerate "flat" behavior at criticality. The entire figure of the phase transition is given in (c) where the emergent property is evident in the errorplots.

here we consider the chance that this attempt fails with a rate λ (target-failure probability). It is possible to generalize the CMED approach just rewriting the master equation as in eq.(15) but taking in account two modes of the degree motion, the add-cut process phase and the resting phase that justifies the presence of "idle" states due to the missed action of the node tendency. Consequently, we consider the population density at rest $\rho_1^{(i)}(k,t)$ in addition to the population

density in motion $\rho_1(k,t)$, and we can rewrite the CMED equation as:

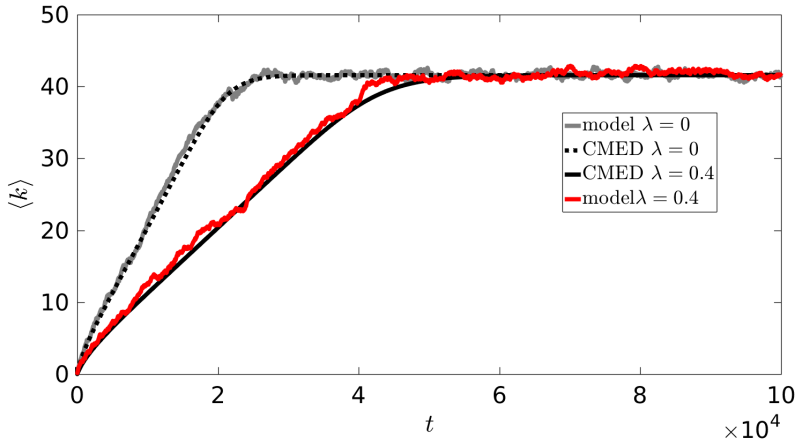
$$\begin{cases} \rho_1(k,t+1) - \rho_1(k,t) = \ \Gamma^+[k-1]\,\rho_1(k-1,t) - \Gamma^+[k]\,\rho_1(k,t) + \\ \qquad\qquad\qquad\qquad - \Gamma^-[k,t]\,\rho_1(k,t) + \Gamma^-[k+1,t]\,\rho_1(k+1,t) + \\ \qquad\qquad\qquad\qquad - \lambda\rho_1(k,t) + \lambda\rho_1^{(i)}(k,t) \\ \rho_1^{(i)}(k,t+1) - \rho_1^{(i)}(k,t) = \lambda\rho_1(k,t) - \lambda\rho_1^{(i)}(k,t) \end{cases} \tag{28}$$

with the usual transition rates eqs.(16),(17) and boundary conditions as in eqs (18),(19). This new master equation described the intermittent add-cut process where idle states introduce pauses in the process due to the probability of each nodes to fail in their intention. Here we considered that the failure rates are the same for all the units in the network. The existence of a stationary solution is guaranteed by the closed domain and the intermittent phases do not contribute to the steady state degree distribution[2]. Finally we have the same truncated poisson stationary distribution as in the classical case of XIE. In principle we can also consider different failure rates for the two groups λ_1 and λ_0, then the degree distributions differ from the standard case of XIE. Finally as shown in Fig.10, the generalized network model shows the same degree distribution of the original CMED but the process is slowed down by the connection failures, so that the trajectories of the degree are time series where "pauses" arise in correspondence of connection failures. This in practice introduces a delay in reaching the equilibrium state as well as a slower speed in the linking mechanism.

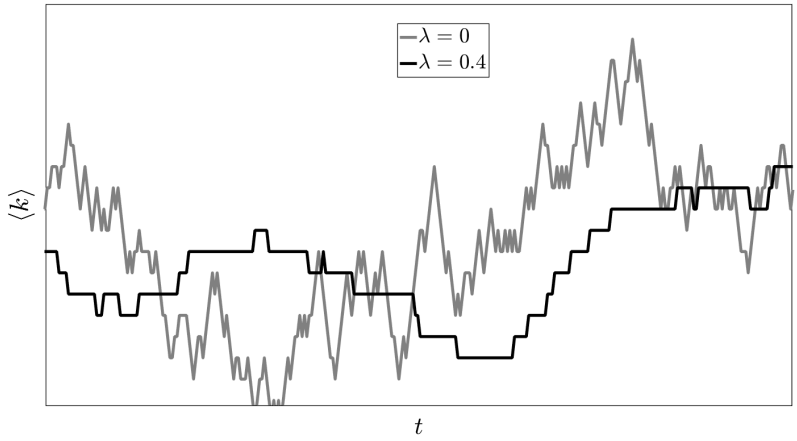
4 Conclusions

In the present paper we described the time evolution of the agent-driven model of XIE via a dynamics of coupled master equations. Our modeling predicts correctly both the transient and the equilibrium distribution of degree outside the transition, and our analysis supports the existence of an extreme Thouless effect, i.e. a discontinuous jump and large dynamic fluctuations, at the critical point where the sizes of extroverts and introverts populations coincide. This type of model, due to its minimal dynamical rules, may shed lights on a broad range of phenomena in complex systems with agents displaying opposed behaviors. In particular, the mixed transition at criticality may help explaining abrupt changes in financial behavior during market crisis. The Thouless effect predicts an explosive, non-linear, collective behavior at a precise critical point, suggesting that controlling for - exogenous or endogenous - changes in agents's behavior could keep the system far from abrupt transitions. In particular, this model of network formation, or its similar generalizations, may help understanding sudden freezes in financial markets, when agents collectively stop trading a specific instrument between each other after a small change in market conditions. This phenomenon will be the subject of further studies. Moreover the introduction of intermittency in the network can give a more realistic approach on networks dynamics where there are times in which system jumps from idle states to active ones. and vice-

[2]the stationary degree distribution can be also recovered just using the ordinary CMED approach as in eq.(15) and using an event-based representation of temporal networks [12]. We can here modify the transition rates defining : $\tilde{\Gamma}^+[k-1] = \Gamma^+[k-1](1-\lambda)$ and $\tilde{\Gamma}^-[k,t] = \Gamma^-[k,t](1-\lambda)$. The problem with this approach is that we ignore the idle states and that the process has pauses in its trajectories. But, at the end, it reaches, with different transient times, the same degree distribution as using the intermittent CMED.



(a) Transition to equilibrium obtained using the model and the CMED in two cases: standard model with no failure rate, and the intermittent case where the failur target probability is $\lambda = 0.4$. CMED approach approximates well both cases, and the introduction of a λ makes the system slower to reach the stationary solution for the average degree.



(b) An inset of the trajectories for large timee, where we notice the presence of pauses in the case where failure rate is present. This slows down the system but at large times it gives the same degree distribution of the standard model ($\lambda = 0$).

Figure 10: trajectories of the average degree for the agent-driven network with target failures and the relative intermittent CMED description. In (a) we notice that the two model and the numerical solution coincides, reflecting that the introduction of failure probability λ slows down the transient state and the reaching of the equilibrium state. The speed factor λ does not affect the final degree distribution but , as shown in (b), introduces pauses in the trajectories of the average degree over time so making the system slower

versa, giving to the entire system a notion of connection velocity. Finally, numerical simulations show that degree fluctuations in time are connected to a peak in degree correlation at the transition, suggesting the need to develop a theory beyond mean-field that will incorporate degree correlations to gain a better description at the transition.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 640772 - DOLFINS.

Appendix A Bounded Poisson degree distribution

Let us take in account that we know the correction factor of the cut rate ϵ in eq.(17) so we can define:

$$\chi = \frac{\langle k \rangle_0}{\epsilon} \tag{29}$$

In the stationary regime the master equation eq.(15) can be written:

$$\left[\frac{k+1}{\chi}\rho_1^{st}(k+1) - \rho_1^{st}(k)\right] - \left[\frac{k}{\chi}\rho_1^{st}(k) - \rho_1^{st}(k-1)\right] = 0 \tag{30}$$

with the boundary conditions as $\rho_1^{st}(k=-1) = 0$ and $\rho_1^{st}(k = N_0 + 1) = 0$.

As proved by the authors in [3], in such closed system the stationary distribution corresponds to the steady state where the detailed balance relation is valid and fulfills the following recurrence equation:

$$\rho_1^{st}(k) = \frac{\chi}{k}\rho_1^{st}(k-1) \tag{31}$$

and by applying the iteration successively we obtain the relation:

$$\rho_1^{st}(k) = \rho_1^{st}(0) \prod_{m=1}^{k} \frac{\chi}{m}. \tag{32}$$

Taking into account the normalization condition:

$$\rho_1^{st}(0) + \sum_{k=1}^{N_0} \rho_1^{st}(k) = 1, \tag{33}$$

the stationary probability distribution in finite network becomes:

$$\rho_1^{st}(k) = \begin{cases} \dfrac{\prod_{m=1}^{k} \frac{\chi}{m}}{1 + \sum_{j=1}^{N_0} \prod_{m=1}^{j} \frac{\chi}{m}} & \text{for } k = 1, \ldots, N_0 \\ \dfrac{1}{1 + \sum_{j=1}^{N_0} \prod_{m=1}^{j} \frac{\chi}{m}} & \text{for } k = 0. \end{cases} \tag{34}$$

Since $\prod_{m=1}^{j} \frac{a}{m} = \frac{a^j}{j!}$, we get the the truncated poissonian distribution :

$$\rho_1^{st}(k) = \frac{1}{e^{\chi}\frac{\Gamma(N_0+1,\chi)}{\Gamma(N0+1)} - 1} \frac{\chi^k}{k!} \qquad , \; k = 0 \ldots N_0 \tag{35}$$

where $\Gamma(N_0+1,\chi)$ is the incomplete gamma function[3]. We want to notice that is the same solutions the authors in [3] found the same numerical solution where the term χ had been evaluated using a fixed-point algorithm, which gives the exact same distributions at equilibrium of our approach of interacting master equations as shown in fig.11 and the correspondent tables 1 and 2 where we compute also a distance between distributions in terms of Kullback-Leibler divergence (KL). This is a measure of how estimated probability distribution diverges from a true expected probability distribution obtained, in this case, from the network model. The results are that the CMED approach and the fixed-point algorithm provide the same solution in determining the stationary degree distributions. Despite the fixed-point approach converge faster to the solution, the CMED approach allows a complete description of the degree motion capturing also the transient states as well as the it can be easy modify to obtain more general agent-driven models. We can observe that the information lost when we use the CMED and fixed-point distribution to approximate the real distribution is not zero even if very small. This is in line with the fact that we are neglecting the although smal correlations among the nodes in the network. Actually at criticality, the KL-divergence is much larger (typically two order of magnitude) highlighting the presence of ignored strong correlations. In order to have a full exploration of all the values of h we ned a consistent measure for the distance of approximated distribution from the true one. We used the Hellinger distance since is a metric satisfying Cauchy-Schwarz inequality. In Fig.12, plotting the Hellinger distance versus the population difference we notice again that the approximation is pretty good for all the cases except for the critical point $h=0$ where the distance is 10 times larger.

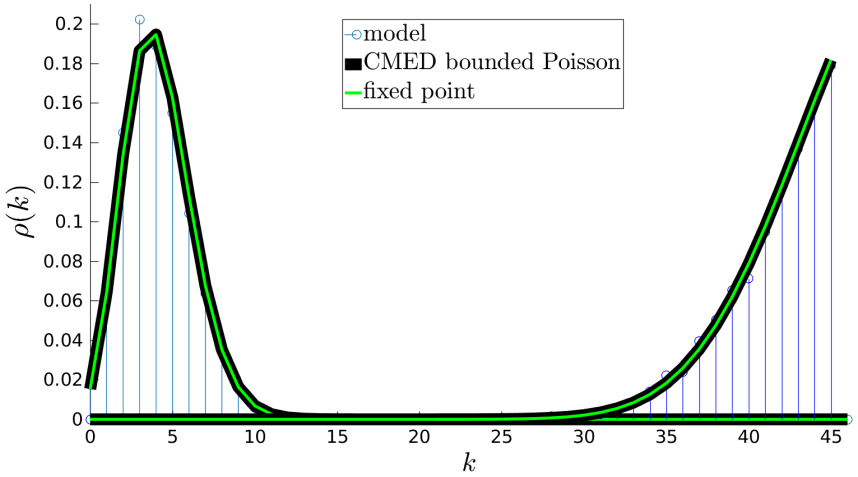


Figure 11: Comparison among the three distributions obtained via montecarlo simulation of the model, using the CMED approach and then substituting the the bounded poisson formula as in eq.(35) and then runngin the numerical solution through the fixed point algorithm used in [3]. We explore the case of $N=10$, the left distributions is for the case $N_0=55, N_1=45$ and the right distribution for $N_0=45, N_1=55$.

[3] for $N_0\to\infty$ we can write $\sum_{j=1}^{N_0}\frac{a^j}{j!}\approx e^a-1$, getting the Zero-truncated Poisson distribution:

$$\rho_1^{st}(k)=\frac{\chi^k}{(e^\chi-1)}\frac{1}{k!} \qquad , k=0\dots N_0.$$

$N_0 = 45, N_1 = 55$	AverageDegree	Variance	KL divenrgence
model (*true distribution)*	41.525	10.224	0
CMED Bounded Poisson	41.582	10.001	0.0022275
fixed point	41.582	10.001	0.0022275

Table 1

$N_0 = 55, N_1 = 45$	AverageDegree	Variance	KL divenrgence
model (*true distribution*)	4.0878	4.1881	0
CMED Bounded Poisson	4.1772	4.1772	0.0023504
fixed point	4.1772	4.1772	0.0023504

Table 2

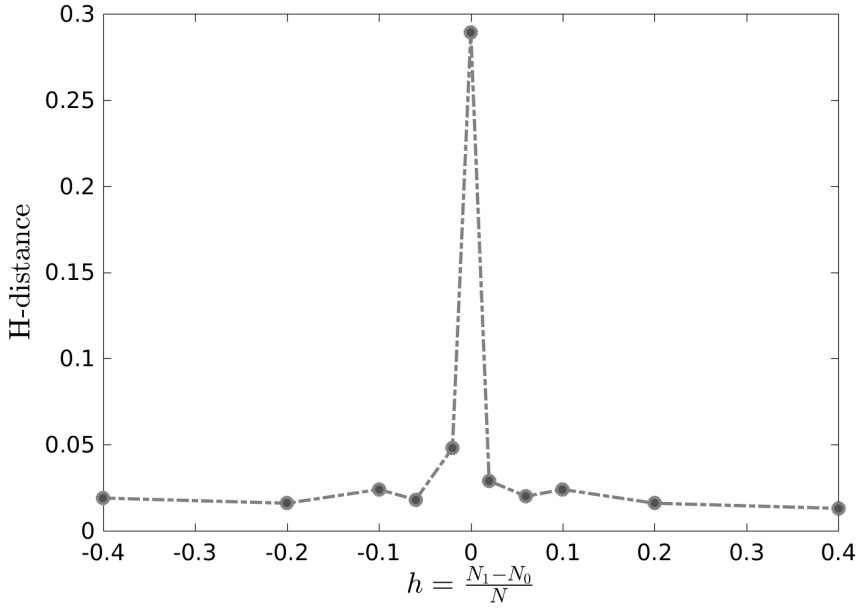


Figure 12: Hellinger distance as measure of divergence between the estimated degree distribution (via CMED) and the true probability obtained from the model. We notice a peak at the critical value, reveling again a much stronger discrepancy between the two distributions since in our mathematical description we consider the network to be uncorrelated.

References

[1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[2] Paolo Barucca and Fabrizio Lillo. The organization of the interbank network and how ecb unconventional measures affected the e-mid overnight market. *arXiv preprint arXiv:1511.08068*, 2015.

[3] KE Bassler, Wenjia Liu, B Schmittmann, and RKP Zia. Extreme thouless effect in a minimal model of dynamic social networks. *Physical Review E*, 91(4):042102, 2015.

[4] Stefano Battiston, J Doyne Farmer, Andreas Flache, Diego Garlaschelli, Andrew G Haldane, Hans Heesterbeek, Cars Hommes, Carlo Jaeger, Robert May, and Marten Scheffer. Complexity theory and financial regulation. *Science*, 351(6275):818–819, 2016.

[5] Guido Caldarelli, Andrea Capocci, Paolo De Los Rios, and Miguel A Munoz. Scale-free networks from varying vertex intrinsic fitness. *Physical review letters*, 89(25):258702, 2002.

[6] Daniel Fricke. Trading strategies in the overnight money market: Correlations and clustering on the e-mid trading platform. *Physica A: Statistical Mechanics and its Applications*, 391(24):6528–6542, 2012.

[7] Crispin Gardiner. *Stochastic Methods : A Handbook for the Natural and Social Sciences*. Springer Berlin, Berlin, 2010.

[8] Martin Grandjean. A social network analysis of twitter: Mapping the digital humanities community. *Cogent Arts & Humanities*, 3(1):1171458, 2016.

[9] Giulia Iori, Giulia De Masi, Ovidiu Vasile Precup, Giampaolo Gabbi, and Guido Caldarelli. A network analysis of the italian overnight money market. *Journal of Economic Dynamics and Control*, 32(1):259–278, 2008.

[10] Giulia Iori, Roberto Reno, Giulia De Masi, and Guido Caldarelli. Trading strategies in the italian interbank market. *Physica A: Statistical Mechanics and its Applications*, 376:467–479, 2007.

[11] Wenjia Liu, Beate Schmittmann, and RKP Zia. Extraordinary variability and sharp transitions in a maximally frustrated dynamic network. *EPL (Europhysics Letters)*, 100(6):66007, 2013.

[12] Naoki Masuda and Renaud Lambiotte. *A guide to temporal networks*, volume 4. World Scientific, 2016.