Temporal properties of network growth with node removal

Author: Miquel Puig Turón
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: Oriol Artime

Abstract: Reproduction of two papers regarding network growth with link/node removal theoretically and with simulations. Temporal properties regarding size of components and percolation transitions have been further studied.

I. INTRODUCTION

Hopelessly complicated systems surround us. Consider for example the society that requires cooperation between billions of individuals. This systems are collectively called *complex systems*. Networks have taken on a new practical role in recent years as a primary tool in the study of complex systems. Some examples are:

- Cellular Network: encoding interactions between genes, proteins and metabolites.
- Social Network: encoding the sum of all professional, friendship, and family ties.
- World Wide Web: encoding webpages and their relations via hyperlinks.

Network Science has an interdisciplinary nature, with an empirical scope, as it is driven by real world data. Based on the mathematics of graph theory, describing random systems with models borrowed from statistical physics and with the help of computer science, for data processing.

The impact that network science has had in society is obvious. From the most successful companies, such as Google, Facebook, Twitter, LinkedIn... to helping in drug design and metabolic engineering, to aiding in the study of epidemics and disease spreading.

The scientific impact is even bigger: Nature, Science, Cell, PNAS, have devoted reviews and editorials addressing the impact of networks on various topics. The 1998 paper by Watts and Strogatz in Nature [1] and the 1999 paper by Barabási and Albert[2] are in the top ten most cited papers in physical sciences during the decade after their publication.

Once we start studying networks, we see that the structure and the evolution of the networks behind each system is driven by a common set of fundamental laws and principles.

Networks in the stationary phase of the asymptotic limit of large times have been extensively studied, and the processes of evolution that can describe them have been modeled. The most famous are the random model, for its mathematical ease, and capacity to reach closed formulae that describe the system, and preferential attachment due to the reproducibility of real world graphs, and the scale-free propierties that arise within them. The mathematical models of network evolution that can be found in this paper are studied with a master equation, which describes how the network evolves in a given state, and timestep.

Vulnerability has also been extensively studied, and how removing certain nodes, a process known as percolation, affects the overall structure. This is interesting in relation to cascading failures, that have been observed in complex systems. The 2009-2011 financial meltdown is an example, the US credit crisis paralyzing the economy of the globe, leaving behind scores of failed banks, corporations, and even bankrupt states. Artificially induced failures are also of interest, for example, cancer researchers aim to induce cascading failures to kill cancer cells.

There has not been extensive study in the combination of both, studying network evolution together with vulnerability. The aim of this TFG is to do so, by studying network growth with node removal. It is divided in two clear parts.

The first part, based on Oriol Artime's paper [3], which consists on the replication of the stochastic resetting in a random network. To do so, theoretical derivations are compared with computer simulations of the model, and compared with networks of different size, and different parameters. More work has been done in deriving additional formulae for sizes of the smaller component, and have been checked computationally.

The second part consists on the replication of the addition and deletion of nodes in evolving networks[4]. This article studies the asymptotic form in the limit of large times, so computer simulations, and numerical resolution are used to see the network evolution, and give insight of the transient phase.

All the code used for the simulations and numerical solutions, is hosted at github. [5]

II. STOCHASTIC RESETTING IN A RANDOM NETWORK

Oriol's paper is based on random networks, which serve as a reference to explore the properties of real networks. Random networks are useful, as it is possible to derive some of their properties analitically, and then compare them to the properties seen in real networks, and see if they could be originated from randomness, or are specific

to the processes that have lead their grow.

A. Model definition

In this article, a set of N interconnected nodes, caracterized by their degree k, forming an undirected graph is considered, with two processes competing in their formation:

- A link between two non connected random nodes is added, at rate $\frac{\alpha N}{2}$.
- A random node with all its links is removed, leaving a node with degree 0, at rate rN

This model is an extension of the Erdos-Rényi model with parameter $p = \frac{\alpha t}{N}$, together with a resetting process.

B. Time dependent degree distribution

The temporal behaviour of the model can be studied from the degree distribution, $p_k(t)$. It can be done with the master equation, which describes the fraction of nodes with degree k, at time t:

$$\frac{dp_k}{dt} = \alpha p_{k-1} - \alpha p_k - r p_k + r(k+1)p_{k+1} - rkp_k + \delta_{k,0} r \quad (1)$$

The first two terms refer to the random addition of a link, achieving degree k from a node that had degree k-1, or losing degree k by reaching k+1. The remaining four are due to the removal of a node. They consider direct removal, and loss of a link due to removal of a direct neighbor. The last term indicates that a the node that has all his links removed achieves degree 0.

It is initially considered a network with no links. $p_k = \delta k, 0$

To solve the master equation, the time-dependent degree generating function is introduced. $g(z,t) = \sum_{j=1}^{k} z^k p_k(t)$

After some algebra, the desired solution can be obtained.

$$g(z,t) = \frac{1}{1-z}\left[e^{\frac{\alpha z}{r}}\mathcal{G}((1-z)e^{-rt} + \frac{r}{\alpha})\right] \quad (2)$$

where $\mathcal{G}((1-z)e^{-rt}) = (x - \frac{r}{\alpha})e^{\frac{\alpha}{r}(x-1)}$.

The generating function can now be rewritten as a power series of the auxiliary variable z, by expanding each of its terms, and the degree distribution is obtained.

$$p_k(t) = \frac{r}{\alpha}\left[1 - \mathcal{Q}(k+1, c(t))\right] + e^{-c(t)-rt}\frac{c(t)^k}{k!} \quad (3)$$

Where $\mathcal{Q}(a,b) \equiv \frac{\gamma(a,b)}{\Gamma(a)}$, and $c(t) = \frac{\alpha}{r}(1 - e^{-rt})$ have been introduced to ease the notation.

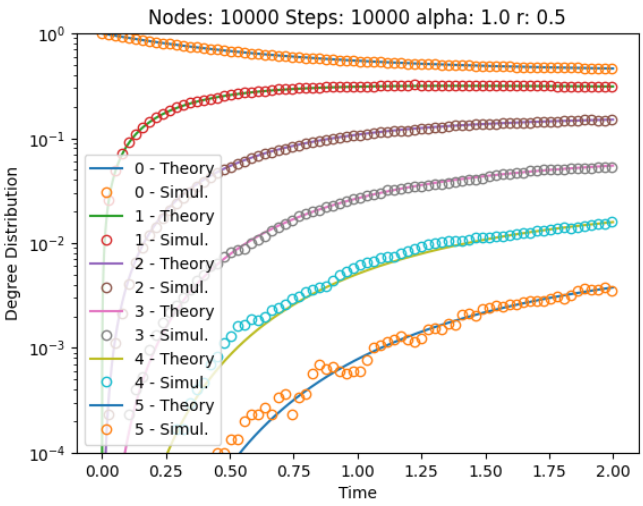

FIG. 1: Comparison of the degree distribution of the simulated network with the theoretical derivation, for degree up to 5.

C. Percolation transition

It is interesting for network robustness to study the size of the giant component. In graphs drawn randomly from a probability distribution over arbitrarily large graphs, a giant component is a connected component whose fraction of the overall number of vertices is bounded away from zero.

A closed expression can be found with the following reasoning. u(t) is the probability that, at time t, a node is not in the giant component via one of its links.

$$u(t) = \frac{1}{\langle k \rangle (t)} \sum_{k=0}^{\infty} k p_k(t)(1 - u(t)^k) \quad (4)$$

If that node has k connections, the probability to belong to the giant component is then $1 - u(t)^k$. Averaging over the degree distribution, the size of the giant component is obtained.

$$S(t) = \sum_{k=0}^{\infty} p_k(t) u(t)^{k-1} \quad (5)$$

Setting u(t) = 1 always results to a solution, leading to S = 0. The conditions for a second solution, and of the percolation phase appearing can be found through Eq. (4). Thus the condition for criticallity is obtained.

$$1 = \frac{2\alpha}{3r}\frac{1 - 2e^{-rt} + e^{rt}}{1 + e^{rt}} \quad (6)$$

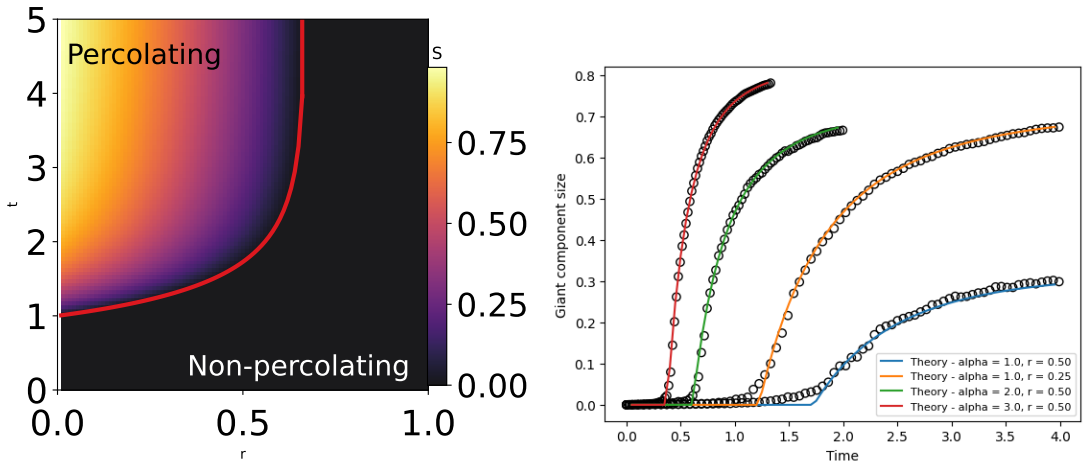

FIG. 2: In the left, size of the giant component for different times and r, for alpha = 1. The critical line is separating the percolating and non-percolating phases. In the right, temporal evolution of the size of the giant component for different parameters. Markers come from simulations, and lines from the theoretical derivation.

D. Average size of small component

This section is new to this TFG, and can't be found in the original paper, yet the study of the small components helps in the visualization and understanding of the percolation as a second-order transition.

A theoretical derivation can be obtained, by considering the second order generating function.n_s is defined as the number of components of size s.

$$R = \frac{\sum_s s n_s}{ns} = \frac{2}{2 - \langle k \rangle * \frac{u^2}{1-S}} \tag{7}$$

The equation is similar to those which can be found in statistical physics describing second order transitions. Notice that the point of divergence is not the same as in the percolating transition, which happens at S = 0, and u = 1, and gives a value of $\frac{2}{2-k}$ which is usually perfectly fine for this equation.

The derivation also assumes an infinite number of nodes, which broadens the peak in simulations due to the size of the graph being bounded.

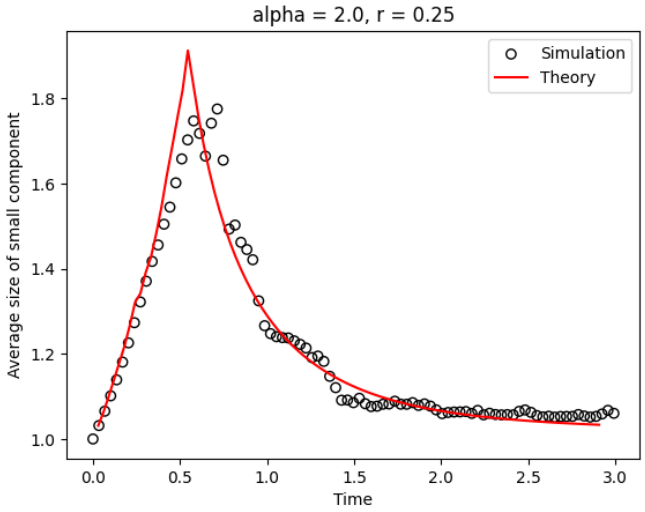

FIG. 3: Average size of small components vs time. Markers come from simulation and the line from the analytical expression. Due to criticallity, it has been hard to obtain a well defined plot around the transition.

III. MODELS OF EVOLVING NETWORKS WITH ADDITION AND DELETION OF NODES

This second paper takes a more general approach on considering network evolution. It allows a variable number of nodes, and an attachment kernel π_k, which decides which growing model is used. In this TFG, like the original paper, uniform attachment and preferential attachment are considered. Another difference is that nodes are removed and not only their links.

A. Model definition

Time has been normalised to alpha, without loss of generality,which leaves this model with 3 different parameters.

1. π_k is the attachment kernel with the probabilities of the new links attaching to the existing nodes for each degree.
2. c new links are added to a new node from existing different nodes
3. r nodes are removed

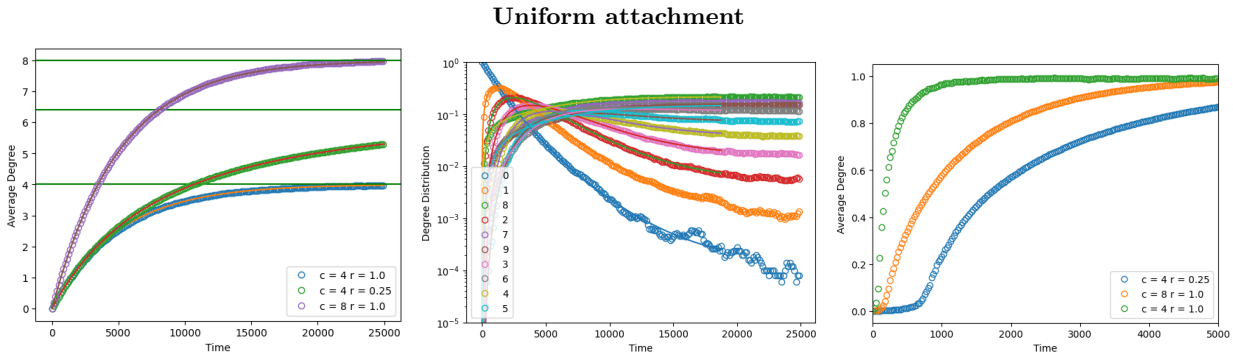

FIG. 4: All three plots consider the uniform attachment model. In the leftmost plot, average degree vs timesteps. The horizontal line is the theoretical value derived in the paper. C and r values indicated. In the middle plot, the degree distribution for degree up to 9 vs timesteps for c = 8 and r = 1. In the rightmost plot, the size evolution of the giant component vs timesteps. C and r values indicated. In the first two plots, markers are from simulations, and lines inside from the numerical solution. In the last one, markers are from simulations. Simulations have been run with 10 000 starting nodes and no links.

1. *Uniform attachment*

When new links choose nodes independently of their degree.

$$\pi_k = 1 \tag{8}$$

This model is similar to the previous one studied.

2. *Preferential attachment*

When links choose nodes according to their degree.

$$\pi_k = \frac{1}{2}\frac{1+r}{c}k \tag{9}$$

The constant is added, to normalize the expression: $\sum_{k=0}^{\infty} \pi_k p_k = 1$

Both models are studied and discussed simultaniously.

B. Time dependent degree distribution

In a similar manner, the temporal behaviour is driven by the degree distribution, $p_k(t)$ which evolves following the master equation.

$$\begin{aligned}(n-1+r)p_k' = np_k + \delta_{kc} + c\pi_{k-1}p_{k-1} - c\pi_k p_k + \\ r(k+1)p_{k+1} - rkp_k - rp_k\end{aligned} \tag{10}$$

The master equation is written as presented in the paper considering discrete unit time steps. In the following timestep, there are $n+1-r$ nodes. The first term refers to the previous number of nodes, the second one, describes a new node being added with degree c, the next two terms, reaching or losing degree k due to gaining a link, the next two terms, reaching or losing degree k due to losing a link, and the last term refers to direct removal of the node.

The master equation has been used rewritten as a partial derivative of time, for a continuous variable t, when it had to be solved numerically.

The model is solved for the asymptotic limit of large times, yet this TFG is interested in studying how the network evolves. No analytical expression has been found, so the model has been studied by solving the master equation numerically, and by simulating it.

In the paper, there is a theoretical derivation for the asymptotic value of the average degree, but no theoretical formula has been found for the transient phase.

When uniform attachment is considered, the behaviour is very similar to the previous paper studied, and doesn't give much insight.

For preferential attachment, as the start is with a clique with very few nodes, the asymptotic theoretical value is reached in very few steps. It is better seen in the degree distribution, where an asympotic value is reached almost instantly.

C. Giant component

In the previous paper, percolation was found through the study of the giant component. No theoretical formula for the giant component has been found for this model.

When uniform attachment is considered, again the behaviour is almost identical to that seen in the first paper, with the difference that the graph eventually is a connected component, being the giant component. This is due to the model, that considers the removal of nodes, and not only of links.

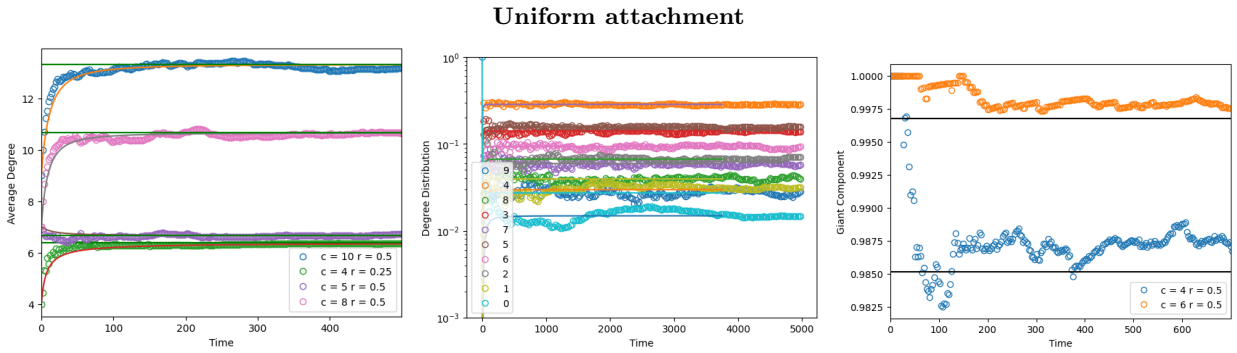

FIG. 5: All three plots consider the preferential attachment model. In the leftmost plot, average degree vs timesteps. The horizontal line is the theoretical value derived in the paper. C and r values indicated. In the middle plot, the degree distribution for degree up to 9 vs timesteps for c = 4 and r = 0.5. In the rightmost plot, the size evolution of the giant component vs timesteps. C and r values indicated. Horizontal lines correspond to the portion of nodes with degree 0. In the first two plots, markers are from simulations, and lines inside from the numerical solution. In the last one, markers are from simulations.

For preferential attachment, an asymptotic value is reached again in a few steps. This makes the model less interesting than expected, as there's no percolation to be studied. What's interesting, is seeing that the giant component isn't the whole graph, and that there is a fraction of nodes that don't belong to the giant component. This fraction is comprised mostly by nodes that end up having degree 0, and are waiting to be removed, as they have no chance to being connected to, but it is not limited to them.

IV. CONCLUSIONS

- Solutions for the degree distribution, and size of giant component have been found for both models. Sizes of small components are also studied in the first model.

  To do so, probabilistic models have been theoretically analyzed through the master equation, and solved for the first paper by introducing the generating function, and in the asymptotic time limit for the second paper. Numerical solutions have been computed and compared with the theoretical derivations, and simulations have been created and then tested with both the theoretical and numerical solutions.

- A transition due to percolation has been found for the networks with addition and removal of links/nodes that consider uniform attachment. This effects are similar to those found in statistical physics, with second order transitions. Not greater insight has been gained in this area from the study of the preferential attachment model.

- As a follow up, it would be interesting to see for the preferential model, networks that shrink. Starting from a known degree distribution, running the model with r¿1. And then compare those results, to see if simmetry is found from the process studied in this TFG.

Acknowledgments

Thank you Oriol for the help and guidance, and to my parents for the support.

[1] Watts, D., Strogatz, S. Collective dynamics of 'small-world' networks. Nature 393, 440–442 (1998).
[2] Barabasi, A. and Albert, R. (1999) Emergence of Scaling in Random Networks. Science, 286, 509-512.
[3] Stochastic resetting in a networked multiparticle system with correlated transitions, d'Oriol Artime, *Oriol Artime 2022 J. Phys. A: Math. Theor. 55 484004*
[4] Exact solutions for models of evolving networks with addition and deletion of nodes, *arXiv:cond-mat/0604069 [cond-mat.stat-mech]*
[5] www.github.com/jamipouchi/TFG
[6] Barabási, A.-L., Pósfai, M. (2016). Network science. Cambridge: Cambridge University Press. ISBN: 9781107076266 1107076269
[7] Newman, M. E. J. (2010). Networks: an introduction. Oxford; New York: Oxford University Press. ISBN: 9780199206650 0199206651.