Distance and height statistics in growing networks

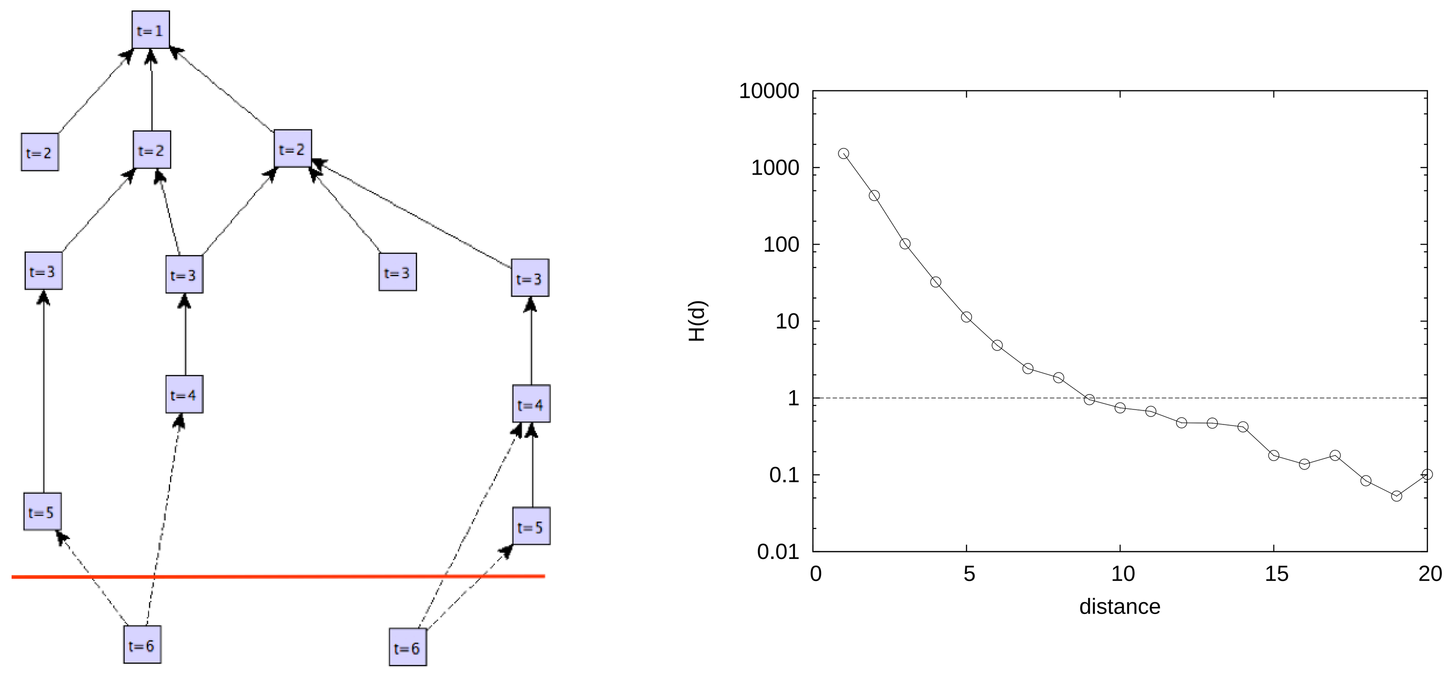
R. Lambiotte[1] and P. Krapivsky[2]

[1] *Université de Liège, Sart-Tilman B5, 4000 Liège, Belgium (renaud.lambiotte@ulg.ac.be; http://www.lambiotte.be)*
[2] *Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA*

Role of the localization in the network, models with copying processes

Motivation

To begin, let us take an example, and consider the case of an active researcher in *Special Relativity*. The probability that he begins a collaboration with e.g. A.-L. Barabási or cites one of his papers is very small, whatever the number of collaborators or number of articles of the latter scientist. Such examples are easily found in friendship networks, citation networks, Internet networks... Briefly speaking, in all systems where *communities* exist and play a strong role.



We first build the citation networks for all N_O articles published in *Phys. Rev. lett.* in the time period [1996 : 2001]. At each new issue t of the journal, we first measure the distances between all $N_O(N_O-1)/2$ pairs of old nodes, and consider independently all new articles. New articles citing (strictly) less than 2 old articles are neglected. Otherwise, say the new article cites k old articles, the distance between the $k(k-1)/2$ old nodes are measured. The distribution of distances between cited nodes $h_{d;t}$ is measured, as well as the distribution of distances between any old nodes $h^O_{d;t}$. Then, the new articles are considered as old ones, the old network is updated and the measurements are repeated. We plot the quantity $H(d)=\frac{1}{T}\sum_{t=T_0+1}^{T_0+T}\frac{h_{d;t}}{h^O_{d;t}}$, for $T=10$. By construction, $H(d)$ should be approximately constant $H(d)\sim 1$ if the distance between cited nodes plays no role.

This implies that preferential attachment models (where nodes connect to randomly chosen nodes in the whole network) are not sufficient in order to reproduce the structuring of many networks. A way to overcome this difficulty is to give attributes to the nodes (taste, speciality...). Unfortunately, it is very difficult to characterise human beings by a small number of variables or tags [2]. In this work, we chose another alternative, and assume that the localization of the network is the relevant information that accounts for the future behaviour of the node.

Model

The undirected network grows by adding nodes one at a time. A newly-introduced node randomly selects a target node and links to it, and it also attempts to link to each neighbor of the target node — it connects either to *all* neighbors of the target node (probability p_1), or to none of them (probability $1-p_1$). At this level, the model accounts for interactions between nodes at distance $d=1$. Generalizations are obtained by also copying links to nodes at a distance d of the target node, with probabilities p_d. This general model (defined by the string $\{p_1,p_2,\ldots\}$ of probabilities) has various specifications. In the following, we focus on a 'closest neighbour' model, $p_g=p\delta_{g,1}$, but other possibilities should be considered, e.g. a 'democratic' model, $p_g=p$ for all $g\geq 1$, a 'geometric' model, $p_g=p^g$...

This copying mechanism is particularly natural if we interpret the system as a growing social network. A new member entering the society establishes a connection (friendship) with someone, and then (with probability p_1) with every friend of the first friend, and then with probability p_2 with friends of new friends, etc.

For the 'closest neighbour model', it is straightforward to show that the total number of links shows a transition, i.e. from a sparse phase to a densifying phase:

$$L(N)=\begin{cases}(1-2p)^{-1}N & \text{for} \quad p<1/2,\\ N\,\ln N & \text{for} \quad p=1/2,\\ A(p)\,N^{2p} & \text{for} \quad 1/2<p\leq 1,\end{cases} \tag{1}$$

Let us stress that such transition in the number of links is also found in other models with copying mechanisms [3], that the number of triangles is very high for the model ($T(N)\sim N$ for $p<1/3$), and that the degree distribution is found to be a power-law in the sparse regime $p<1/2$.

Interestingly, it is possible to write an exact equation for the average number of pairs of nodes separated by distance d, $P_d(N)$:

$$P(d,N+1)=\left[1+\frac{2p}{N}\right]P(d,N)+\frac{2(1-p)}{N}P(d-1,N) \tag{2}$$

It is possible to show that the first moment $D(N)=\sum_{d\geq 1}dP(d,N)$ behaves like:

$$D(N)=(1-p)N(N+1)H_N-\left(2-\frac{5p}{2}\right)N^2-\frac{p}{2}N \tag{3}$$

where H_N denotes the generalized harmonic number $H_N=\sum_{1\leq n\leq N}n^{-1}$. Asymptotically, the diameter $\frac{D}{(N(N-1)/2}\sim 2(1-p)\ln(N)$. The fluctuations around this mean value are found to be Gaussian, and asymptotically vanish in the long time limit $N\to\infty$.

Directed networks with preferential attachment

Many links with PA

In undirected networks, the diameter, or average distance between nodes, is a primordial quantity that is related to mechanisms such as spreading of viruses, information transfers... It is well-known that many classes of random networks have a very small diameter, which scales as $d\sim\ln N$, where N is the number of sites (see left column). When preferential attachment mechanisms are involved, the grow may become ultra-small and behave like $\ln(\ln(N))$ [4]. The equivalent quantity for directed networks would be the average height, the height of a node being defined to be the minimum number of links to the seed node. In this work, we try to answer the following question: How does the height evolves in networks that grow with Preferential Attachment, and where entering nodes have $i>1$ outgoing links, e.g. $i=2$?

When a new node enters, it connects to i nodes chosen with preferential attachment, i.e. with a probability proportional to $1+\mu k$, where k is the indegree of the node. By definition, after t steps, starting from 1 initial node (the seed), there are $t+1$ nodes and (it) links, so that the normalisation factor is $Z=\sum(1+\mu k)=1+(\mu i+1)t$.

When each entering node has only one outgoing link, it is straightforward to show that $m_{1;t}\sim\frac{1}{1+\mu}\ln t$. Let us stress that the total number of links of the nodes at height $g-1$ is fixed $L_{g-1;t}=N_{g;t}$ when $i=1$. When $i>1$, this equality ceases to hold, and coupled equations for N and L have to be considered:

$$\begin{aligned}N_{g;t+1}&=N_{g;t}+\frac{N_{g-1;t}+\mu L_{g-1;t}}{1+(2\mu+1)t}\left(2-\frac{N_{g-1;t}+\mu L_{g-1;t}}{1+(2\mu+1)t}-2\sum_{h=0}^{g-2}\frac{N_{h;t}+\mu L_{h;t}}{1+(2\mu+1)t}\right)\\ L_{g;t+1}&=L_{g;t}+2\frac{N_{g;t}+\mu L_{g;t}}{1+(2\mu+1)t}.\end{aligned} \tag{4}$$

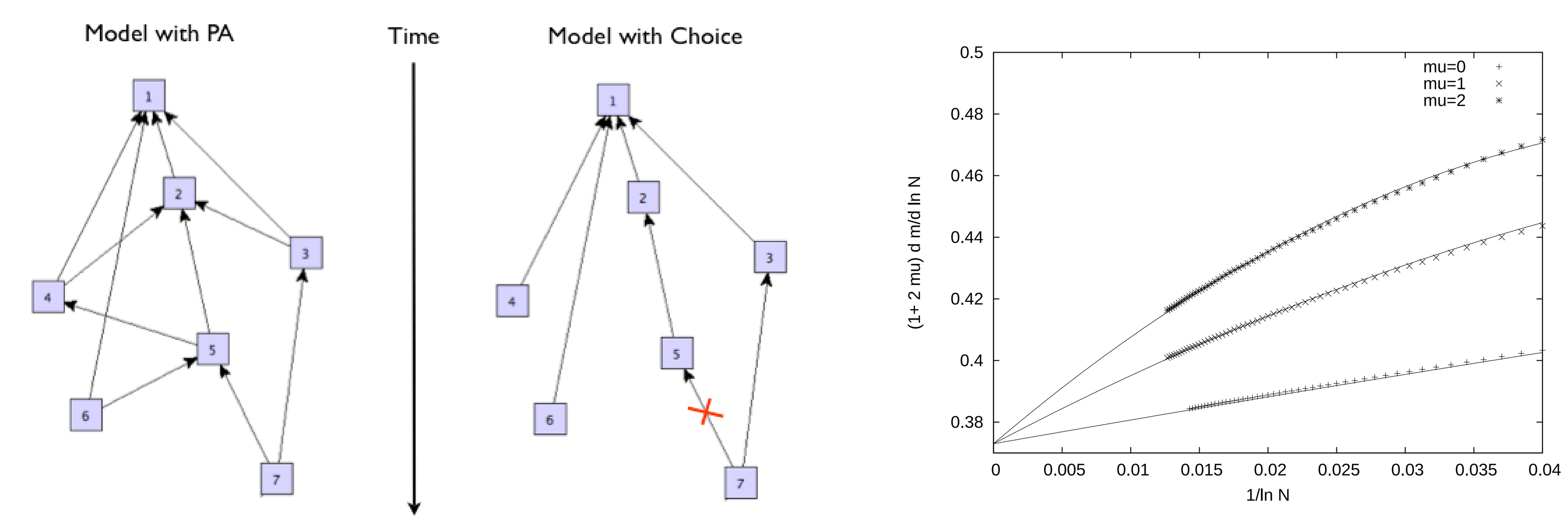
that simplifies after introducing the cumulative probabilities $A_k=1-\frac{\sum_{j=0}^{k}N_k}{1+t}$ and on $Z_g=1-\frac{\sum_{j=0}^{g}L_k}{2t}$, and changing the time scale $d\tau=(1+t)^{-1}dt$:

$$\begin{aligned}\partial_\tau A_g&=\frac{1}{(2\mu+1)^2}(A_{g-1}+2\mu Z_{g-1})^2-A_g\\ \partial_\tau Z_g&=\frac{1}{(2\mu+1)}(A_g+2\mu Z_g)-Z_g.\end{aligned} \tag{5}$$

When $\mu\to 0$ (no PA), the equation for A_g decouples, and becomes very similar to an equation obtained by Ben-Naim et al. [5] for the minimal heights of randomly generated binary trees. It is non-local and non-linear, and simulations show that the asymptotic solution is a propagating wave. Long calculations lead to the dispersion relation:

$$v=\frac{1}{\lambda(2\mu+1)}(1-2e^{-\lambda}). \tag{6}$$

Any such velocity (and corresponding λ) is in principle allowed, but velocity selection mechanism says that the maximal value of v is chosen by the front.



Condensation

A variant of the above model leads to an interesting property. In this Model With Choice (MWC), each time a node enters the system, it randomly selects with PA 2 old nodes and connects only to **one** node, chosen to be the closest to the seed. Similar calculations lead to the following dispersion relation:

$$\lambda v=1-\frac{2}{(\mu+1)}(e^{-\lambda}+\mu), \tag{7}$$

whose maximal velocity converges to zero for $\mu\to 1$, thereby showing a transition from a regime $v\sim\ln N$ ($\mu<1$) to a regime where the system is ultra-small $v/\ln N\to 0$ ($\mu>1$). It is interesting to note that this transition is associated to a condensation of the links around the seed, i.e. when $\mu>1$, the proportion of links connecting to the seed $n_1\equiv N_1/t$ converges to the non-vanishing value $n_1=\frac{\mu^2-1}{\mu^2}$, while $n_1\to 0$ otherwise.

Acknowledgments R.L. has been supported by European Commission Project CREEN FP6-2003-NEST-Path-012864.

References

[1] A.-L. Barabási and R. Albert , *Science* **286** (1999), 509
[2] S. Golder and B. A. Huberman, cs.DL/0508082
[3] P. L. Krapivsky and S. Redner, *Phys. Rev. E* **71** (2005), 036118
[4] R. Cohen and S. Havlin, *Phys. Rev. Lett.* **90**, 058701 (2003)
[5] E. Ben-Naim, P. L. Krapivsky, S. N. Majumdar, *Phys. Rev. E* 64, R035101 (2001)