Todo list

Degree distribution in GCC

Benoît Richard Guiyuan Shi

August 21, 2018

Abstract

Write abstract

1 Introduction

Studying the fundamental properties of networks require to be able to abstract from the particular examples found in nature. This is usually done [?] by using a random model for the network generation and averaging the properties of interest over the set of possible networks. One common model is the configuration model [?] that allows to uniformly sample the space of all networks with a given degree distribution [?]. However, many real examples of networks are connected, as for example the World Wide Web or railroad networks, but no model known to us allows to uniformly sample the space of all connected networks of a given degree distribution.

Missing ref

Missing ref

Missing ref

Find other examples

A way to still study connected networks is to only consider the Giant Connected Components (GCC) of networks generated using the configuration model [?]. We study here how this method implies bias on the degree distribution of the GCC and propose an algorithm based on this knowledge to generate connected networks of arbitrary degree distribution.

Missing ref

2 Degree distribution in GCC

Per Bayes theorem we have for two random events A and B

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(B|A)\frac{P(A)}{P(B)}. \tag{1}$$

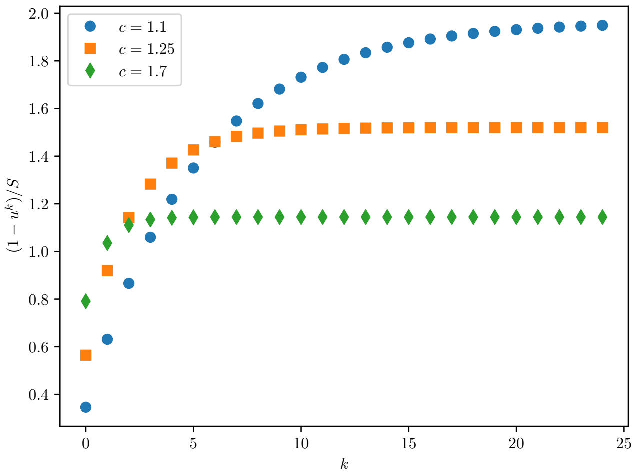
{Bayes theorem}

We can apply it to compute the probability r_k that a vertex in the GCC has degree k

$$\begin{aligned} r_k &= P\left(deg(v) = k | v \in GCC\right) && (2)\\ &= P(v \in GCC | deg(v) = k)\frac{P(deg(v) = k)}{P(v \in GCC)} && (3)\\ &= (1 - P(v \notin GCC | deg(v) = k))\frac{p_k}{S} && (4)\\ &= \frac{p_k}{S}(1 - u^k)., && (5) \end{aligned}$$

{Degree distribution in GCC}

where S is the probability that a random node is part of the GCC, p_k is the probability that a node has degree k and u is the probability that a node reached by following an edge of the network is not part of the GCC.



{Figure: low degree saturation}

Figure 1: Bias factor $(1 - u^k)/S$ for Poisson random graph of various mean degree c. The closer to the critical point at $c = 1$, the larger the effect.

Justification of the equality $P(v \notin GCC|deg(v) = k) = u^k$ can be found inf Ref. [1].

In the context of the configuration model we choose the probabilities p_k which determine u and S through the following equations [1]

$$u = \frac{\sum_{k=1}^{\infty} k p_k u^{k-1}}{\sum_{k=1}^{\infty} k p_k} \tag{6}$$

$$S = 1 - \sum_{k=1}^{\infty} p_k u^k, \tag{7}$$

{Expression for u}

thus eliminating all unknown in eq. (5).

As seen in eq. (5), considering a vertex in the GCC biases the probability that it has degree k by a factor $(1 - u^k)/S$ as compared to choosing a vertex uniformly in the network. The bias factor is depicted in fig. 1. Since both u and S are smaller than 1, the net effect is to lower the proportion of low degree vertices in the GCC and thus to increase the proportion of high degree vertices.

3 Generating connected networks

The knowledge of the degree distribution in the GCC can be used generate a connected component of a given degree distribution r_k as follow: we first determine a degree distribution p_k fulfilling eq. (5) for some target degree distribution r_k. Then we generate a network with degree distribution p_k using the configuration model. Finally we take its GCC as our connected network. By construction the vertices in the GCC will have degree distribution r_k. Determining the factors p_k is not immediate however since u is an unknown which is itself a function of p_k. We propose an algorithm to determine it numerically.

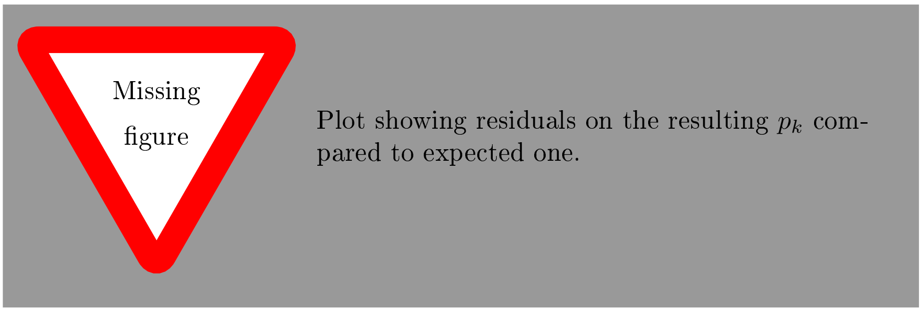


Figure 2: Do the caption as well

First we isolate p_k from eq. (5) to get

$$p_k = S\pi_k(u), \quad \text{with} \quad \pi_k(z) = \frac{r_k}{1 - z^k} \tag{8}$$

Inserting this in the expression(7) for u, we get

$$u = \frac{\sum_{k=1}^{\infty} k\pi_k(u)u^{k-1}}{\sum_{k=1}^{\infty} k\pi_k(u)}. \tag{9}$$

Therefore u is a fixpoint of the function

$$\mu(z) = \frac{\sum_{k=1}^{\infty} k\pi_k(z)z^{k-1}}{\sum_{k=1}^{\infty} k\pi_k(z)}, \tag{10}$$

{Defition of mu}

which is fully determined by the GCC degree distribution r_k. By virtue of the fixpoint theorem, the sequence u_k defined by $u_{k+1} = \mu(u_k)$ with $u_0 = 0$ will converge towards u for k going to infinity. See Appendix A for a more detailed proof of this claim. Since we can not deal numerically with infinite sums, we need to choose a cutoff index K such that

$$\sum_{k=K+1}^{\infty} k\pi_k(z) \ll 1. \tag{11}$$

Once u is approximated, we can compute the first K probabilities p_k, which is sufficient to sample random numbers between 1 and K with relative probability p_k. If K is chosen such that $r_k << 1$ for $k > K$, the degree distribution in the GCC closely approximate the distribution r_k.

Would be interesting to have a use case where the difference in distribution has a meaningful impact

This may not be possible for scale-free network due to fat tail, check ?

4 Discussion

Write discussion

A Analysis of the $\mu(z)$ function

{Appendix: Fixpoint convergence}

To prove that $\mu(z)$ is an increasing function, we compute its derivative with respect to z, which yields

$$\mu'(z) = \left[\sum_{k=1}^{\infty} k\pi_k(z)\right]^{-2} (s_1(z) + s_2(z)) \tag{12}$$

$$s_1(z) = \sum_{j,k} kj\pi'_k(z)\pi_j(z)\left(z^{k-1} - z^{j-1}\right) \tag{13}$$

$$s_2(z) = \sum_{j,k} k(k-1)j\pi_k(z)\pi_j(z)z^{k-2}. \tag{14}$$

The sum $s_1(z)$ can be rewritten as

$$\begin{aligned} s_1(z) &= \sum_{j>k} kj\left(\pi'_k(z)\pi_j(z) - \pi'_j(z)\pi_k(z)\right)\left(z^{k-1} - z^{j-1}\right) && (15)\\ &= \sum_{j>k} \frac{kr_k}{1-z^k}\frac{jr_j}{1-z^j}\frac{z^k - z^j}{z^2}\left(\frac{k}{z^{-k}-1} - \frac{j}{z^{-j}-1}\right) && (16)\\ &= \sum_{j>k} kj\pi_k(z)\pi_j(z)\frac{z^k - z^j}{z^2}\left(\frac{k}{z^{-k}-1} - \frac{j}{z^{-j}-1}\right). && (17)\end{aligned}$$

Using the fact that the function

$$f_z(\lambda) = \frac{\lambda}{z^{-\lambda} - 1} \tag{18}$$

is a decreasing function of λ we can see that for $z \in [0, 1)$ and $j > k$ we have

$$z^k - z^j \geq 0 \tag{19}$$

$$\frac{k}{z^{-k}-1} - \frac{j}{z^{-j}-1} \geq 0, \tag{20}$$

and thus $s_1(z) \geq 0$. Moreover each terms in we have $s_2(z) > 0$, so we can conclude that $\mu'(z) > 0$ and thus that $\mu(z)$ is a strictly increasing function of z.

In order to prove that $\mu(z) > z$, we first rewrite the function $\mu(z)$ as

$$\mu(z) = \frac{1}{z}\left[1 - \frac{\sum_{k=1}^{\infty} kr_k}{\sum_{k=1}^{\infty} \frac{kr_k}{1-z^k}}\right]. \tag{21}$$

Noting that since $0 \leq z < 1$ we have $1 - z^2 \leq 1 - z^k$ for $k \geq 2$, we find

Verify the following

$$\mu(z) \leq \frac{1}{z}\left[1 - (1 - z^2)\frac{\sum_{k=1}^{\infty} kr_k}{(1+z)r_1 + \sum_{k=2}^{\infty} kr_k}\right]. \tag{22}$$

References

[1] M. Newman. *Networks: an introduction*. 2010.