Power-Law Tail of the Degree Distribution in the Connected Component of the Duplication Graph

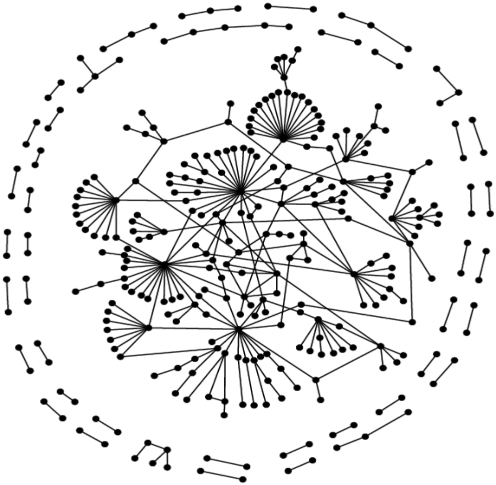
Krzysztof Turowski

Theoretical Computer Science Department, Jagiellonian University

Joint work with P. Jacquet, W. Szpankowski

AoFA 2020

Dynamic networks



Duplication-divergence model

Model definition: starting from a certain graph G on t_0 vertices we add vertices one by one in the following way:

1. pick any vertex v uniformly at random from all t vertices of a current graph,
2. add a new vertex u to G,
3. attach u to any vertex connected to v – independently, with probability p $(0 \leq p \leq 1)$.

We call this model duplication-divergence and denote by $\mathtt{DD(t, p)}$.

This model is supposed to be well-suited to many types of biological networks, e.g. protein-protein networks.

Example

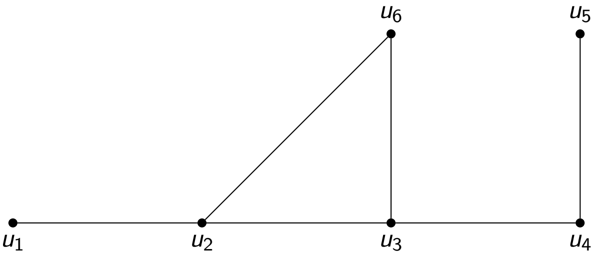


FIGURE: Example graph growth for $p = 0.8$.

Example

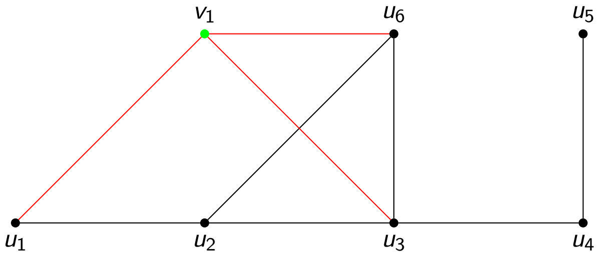


FIGURE: Example graph growth for $p = 0.8$.

Example

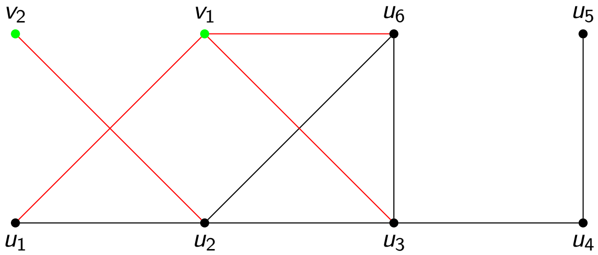


FIGURE: Example graph growth for $p = 0.8$.

Example

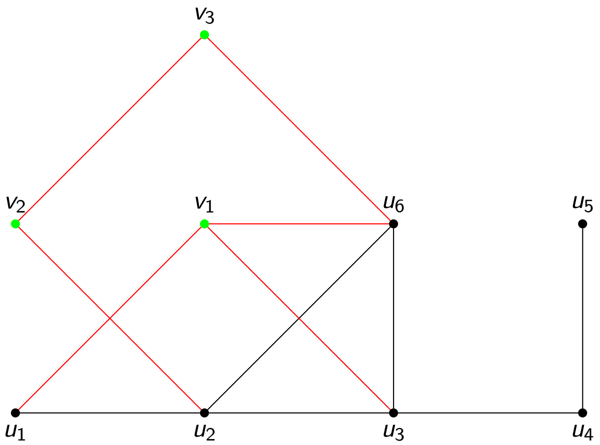


FIGURE: Example graph growth for $p = 0.8$.

Previous work

Let

$$f(k) = \lim_{n\to\infty} f_n(k) = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}|\{v \in V(G_n) : \deg_n(v) = k\}|.$$

1. Hermann, Pfaffelhuber, 2014: for $\mathtt{DD}(t, p)$ with $p < 1$ we have $f(k) = 0$ for all $k \neq 0$. Moreover, when $p < 0.567\ldots$ it holds that $f(0) = 1$, otherwise $f(0) = c \in (0, 1)$,
2. Li et al., 2013: for $\mathtt{DD}(t, p)$ with $0 < p < \frac{1}{2}$ it holds that $f_n(1) = \Omega(\ln t/t)$.

Other work includes studying average degree (Bebek et al., 2006), triangles (Hermann, Pfaffelhuber, 2014), open triangles (Sreedharan et al., 2020), maximum degree (Frieze et al., 2020), automorphisms (Turowski et al., 2019) in this model.

Jordan's result

Let us focus on the connected component of the graph:

$$a_n(k) = \frac{f_k(n)}{\sum_{i=1}^{\infty} f_n(i)} = \frac{f_n(k)}{1 - f_n(0)}.$$

Theorem (Jordan 2018, Theorem 2.1(3))

Assume $0 < p < \frac{1}{e}$. Let $\beta(p) > 2$ be the solution of $p^{\beta-2} + \beta - 3 = 0$. Then

$$\lim_{k\to\infty} \frac{a(k)}{k^q} = \begin{cases} 0 & \text{for } q < \beta(p), \\ \infty & \text{for } q > \beta(p). \end{cases}$$

This result established <u>almost</u> power-law behavior. We strengthen this theorem by proving the exact limit of $\frac{a(k)}{k^q}$.

Jordan's approach

Jordan constructed the generator Q of the continuous-time Markov chain $(\deg(V_t))_{t\geq 0}$, over the state space $\mathbb{N}_0$:

$$\begin{aligned} q_{j,k} &= \binom{j}{k} p^k (1-p)^{j-k} && \text{for } 0 \leq k \leq j-1, \\ q_{j,j} &= -jp - \left(1 - p^j\right), \\ q_{j,j+1} &= jp. \end{aligned}$$

The quasi-stationary distribution $(a(k))_{k=1}^{\infty}$ is the left eigenvector of a subset of Q so it holds that:

$$\sum_{j=k}^{\infty} a(j) \binom{j}{k} p^k (1-p)^{j-k} = -(k-1)pa(k-1) - (\lambda - kp - 1)\, a(k)$$

for $k = 1, 2, 3, \ldots$.

Jordan's approach

For GF $A(z) = \sum_{k=0}^{\infty} a(k) z^k$ we have the equation

$$A(pz + 1 - p) = (1 - \lambda)A(z) + pz(1 - z)A'(z) + A(1 - p).$$

Therefore the equation above implies:

- $A(0) = 0$,
- if $A'(1)$ is finite, then $A(1 - p) = \lambda A(1)$,
- if $A'(1)$ is non-zero and finite, then $\lambda = 1 - 2p$.

Jordan found that for $0 < p < e^{-1}$ the quasi-stationary distribution $a(k)$ does not have q-th moment for $p^{q-2} + q - 3 < 0$ – which implied his result.

Our result

Theorem

If $0 < p < e^{-1}$, then the stationary distribution $(a(k))_{k=0}^{\infty}$ of the pure duplication model has asymptotic value of the coefficient for $\frac{a(k)}{k^{\beta(p)}}$ as $k \to \infty$:

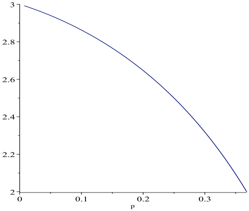
$$\frac{1}{E(1) - E(\infty)} \cdot \frac{p^{-\frac{1}{2}(\beta(p)-\frac{3}{2})^2}\Gamma(\beta(p)-2)}{D(\beta(p)-2)(p^{-\beta(p)+2}+\ln(p))\Gamma(1-\beta(p))}\left(1+O(k^{-1})\right)$$

where $\beta(p) > 2$ is the non-trivial solution of $p^{\beta-2}+\beta-3=0$, $\Gamma(s)$ is the Euler gamma function and

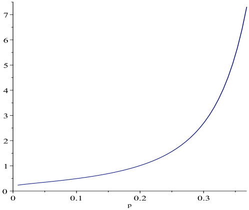
$$D(s) = \prod_{i=0}^{\infty}\left(1+p^{1+i-s}(s-i-2)\right),$$

$$E(1) - E(\infty) = \frac{1}{2\pi i}\int_{\mathrm{Re}(s)=c} p^{-\frac{1}{2}(s-\frac{1}{2})^2}\frac{\Gamma(s)}{D(s)}\,\mathrm{d}s, \quad \textit{for } c \in (0,1).$$

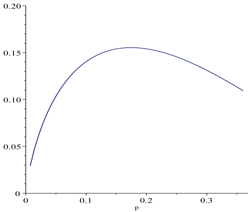
Numerical values of constants



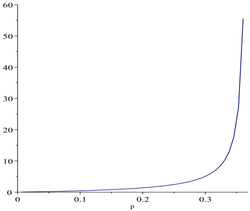
(A) $\beta(p)$



(B) $E(1) - E(\infty)$



(C) $D(\beta(p) - 2)$



(D) $\frac{p^{-\frac{1}{2}(\beta(p)-\frac{3}{2})^2}\Gamma(\beta(p)-2)}{\left(p^{-\beta(p)+2}+\ln(p)\right)\Gamma(1-\beta(p))}$

FIGURE: Numerical values of different parts of the equation for $0 < p < e^{-1}$.

Our proof

We want to solve the Jordan equation

$$A(pz+1-p) = (1-\lambda)A(z) + pz(1-z)A'(z) + A(1-p),$$
$$C(w/p) = 2pC(w) + p(w-1)C'(w) + A(1-p).$$

with boundary conditions $C(1) = A(0) = 0$ and $\lim_{w\to\infty} C(w) = A(1)$.

We are doing this via solving $E(w/p) = 2pE(w) + p(w-1)E'(w) + K$ for some constant K for which claim that the Mellin transform

$$E^*(s) = \int_0^\infty w^{s-1}E(w)\,\mathrm{d}w$$

exists in some fundamental strip and then by using the relation

$$C(w) = A(1)\frac{E(w)-E(1)}{E(\infty)-E(1)}.$$

Our proof

We first guess

$$E^*(s) = p^{-\frac{1}{2}(s-\frac{1}{2})^2} \frac{\Gamma(s)}{D(s)}$$

for $D(s) = \prod_{j=0}^{\infty} \left(1 + p^{1+j-s}(s-j-2)\right)$ already used in the theorem statement.

For $0 < p < e^{-1}$ we have $D(s) = 0$ only when $s = j+1$ and $s = j+1+s^*$, where s^* is the non-trivial (i.e. other than $s = 0$) real solution of $p^s + s - 1 = 0$.

Therefore, $E^*(s)$ has only simple, isolated poles of three types:

- for $s = 0, -1, -2, \ldots$, introduced by $\Gamma(s)$,
- for $s = 1, 2, 3, \ldots$, introduced by $\frac{1}{D(s)}$,
- for $s = s^* + 1, s^* + 2, s^* + 3, \ldots$, introduced by $\frac{1}{D(s)}$.

Our proof

Lemma

For $\mathrm{Re}(s) \in (-1, 0)$ *and* $0 < p < e^{-1}$ *it holds that* $\frac{1}{|D(s)|}$ *is absolutely convergent.*

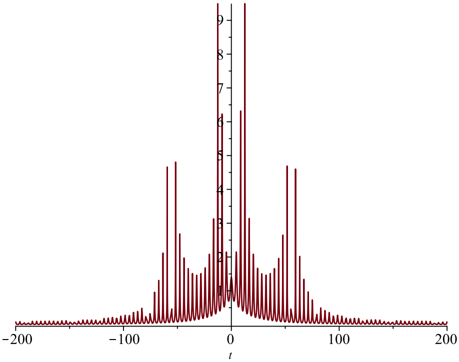


FIGURE: Example numerical values of $\frac{1}{|D(c+it)|}$ for $p = 0.2$ and $c = -0.5$.

Our proof

We may show that

$$E^*(s) = \frac{p(s-1)}{p^s + ps - 2p} E^*(s-1).$$

Moreover, for any given $c \in (-1, 0)$ we introduce

$$E(w) = \frac{1}{2\pi i} \int_{\mathrm{Re}(s)=c} E^*(s) w^{-s} \, \mathrm{d}s = \frac{1}{2\pi i} \int_{\mathrm{Re}(s)=c} p^{-\frac{1}{2}(s-\frac{1}{2})^2} \frac{\Gamma(s)}{D(s)} w^{-s} \, \mathrm{d}s,$$

such that it has function $E^*(s)$ as its Mellin transform with its fundamental strip being $\{s : \mathrm{Re}(s) \in (-1, 0)\}$.

Our proof

We may show that both

$$E(\infty) = \lim_{w\to\infty} E(w) = -\lim_{w\to\infty} \operatorname{Res}\left[E^*(s)w^{-s}, s = 0\right] = -\frac{p^{-\frac{1}{8}}}{D(0)}$$

and

$$E(\infty) - E(1) = -\operatorname{Res}[E(s), s = 0] - \frac{1}{2\pi i}\int_{\operatorname{Re}(s)=c} E^*(s)\,\mathrm{d}s$$
$$= -\frac{1}{2\pi i}\int_{\operatorname{Re}(s)=c'} E^*(s)\,\mathrm{d}s,$$

respectively for $c \in (-1, 0)$ and $c' \in (0, 1)$, are finite.

Integration area

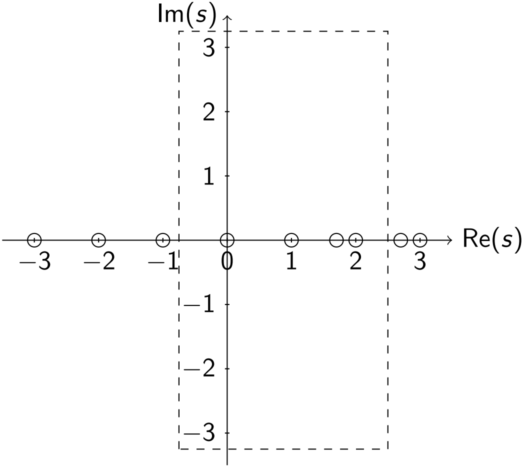


FIGURE: Example integration area for $E^*(s)$ and $E(w)$ with $s^* = 0.7$ and $M = 2.5$.

Our proof

For any $c \in (-1, 0)$ and $M \in (2, 2 + s^*)$ we have

$$C(w) = \frac{1}{E(\infty) - E(1)} \frac{1}{2\pi i} \int_{\mathrm{Re}(s)=c} E^*(s) w^{-s} \, \mathrm{d}s - \frac{E(1)}{E(\infty) - E(1)}$$
$$= -\frac{1}{E(\infty) - E(1)} \left(E(1) + \mathrm{Res}[E^*(s), s = 0]\right)$$
$$- \frac{1}{E(\infty) - E(1)} \left(\mathrm{Res}\left[E^*(s) w^{-s}, s = 1\right] + \mathrm{Res}\left[E^*(s) w^{-s}, s = 2\right]\right)$$
$$- \frac{1}{E(\infty) - E(1)} \mathrm{Res}\left[E^*(s) w^{-s}, s = s^* + 1\right]$$
$$+ \frac{1}{E(\infty) - E(1)} \frac{1}{2\pi i} \int_{\mathrm{Re}(s)=M} E^*(s) w^{-s} \, \mathrm{d}s.$$

Our proof

We may prove that

$$\frac{1}{2\pi i}\int_{\mathrm{Re}(s)=M} E^*(s)w^{-s}\,\mathrm{d}s = O(w^{-M}),$$

$$\mathrm{Res}\big[E^*(s)w^{-s}, s=0\big] = \left[p^{-\frac{1}{2}(s-\frac{1}{2})^2}\frac{w^{-s}}{D(s)}\right]_{s=0} = \frac{p^{-\frac{1}{8}}}{D(0)} = -E(\infty),$$

$$\mathrm{Res}\big[E^*(s)w^{-s}, s=1\big] = \left[\frac{p^{-\frac{1}{2}(s-\frac{1}{2})^2}\Gamma(s)}{p^{1-s}-(s-2)p^{1-s}\ln(p)}\frac{w^{-s}}{D(s-1)}\right]_{s=1}$$

$$= \frac{p^{-\frac{1}{8}}}{1+\ln(p)}\frac{w^{-1}}{D(0)},$$

$$\mathrm{Res}\big[E^*(s)w^{-s}, s=s^*+1\big] = \left[\frac{p^{-\frac{1}{2}(s-\frac{1}{2})^2}\Gamma(s)}{p^{1-s}-(s-2)p^{1-s}\ln(p)}\frac{w^{-s}}{D(s-1)}\right]_{s=s^*+1}$$

$$= \frac{p^{-\frac{1}{2}(s^*+\frac{1}{2})^2}\Gamma(s^*)}{p^{-s^*}+\ln(p)}\frac{w^{-s^*-1}}{D(s^*)}.$$

Our proof

Finally, we go back from $C(w)$ to $A(z)$ and from $w = (1-z)^{-1}$ to z and use Flajolet-Odlyzko transfer theorem:

- $(1-z)^\alpha$ for $\alpha \in \mathbb{N}$ is a polynomial and does not contribute to the asymptotics of $[z^k]A(z)$,
- for $\alpha \in \mathbb{R}_+ \setminus \mathbb{N}$ it holds that

$$[z^k](1-z)^\alpha = \frac{k^{-\alpha-1}}{\Gamma(-\alpha)}\left(1 + O\left(\frac{1}{k}\right)\right),$$
$$[z^k]o(1-z)^\alpha = o(k^{-\alpha-1}).$$

Putting all this together gives us the final result.

Future work

The case $p \geq e^{-1}$ remains open.

We conjecture that

$$f_n(k) = O(n^{-\alpha(p)} k^{-\beta(p)})$$

for some $0 < \alpha(p) < 1$ and $\beta(p) > 2$ asymptotically.