Preferential attachment with choice

Jonathan Jordan (joint with John Haslegrave)

School of Mathematics and Statistics, University of Sheffield

LMS Symposium, Durham, July 2017

Motivation: Achlioptas processes

Achlioptas processes: generalisation of Erdős-Rényi graph process.

Motivation: Achlioptas processes

Achlioptas processes: generalisation of Erdős-Rényi graph process.

- A sample of possible edges are selected randomly.

Motivation: Achlioptas processes

Achlioptas processes: generalisation of Erdős-Rényi graph process.

- A sample of possible edges are selected randomly.
- One edge is chosen to be added to the graph, using some criterion.

Motivation: Achlioptas processes

Achlioptas processes: generalisation of Erdős-Rényi graph process.

- A sample of possible edges are selected randomly.
- One edge is chosen to be added to the graph, using some criterion.

Some interest (papers by Malyshkin & Paquette and Krapivsky & Redner) in similar modifications to preferential attachment process.

Definition of process

Fix integers r, s with $r \geq s \geq 1$.

Definition of process

Fix integers r, s with $r \geq s \geq 1$.

We grow a tree, starting from the two-vertex tree at time 1. At each time step we select an ordered r-tuple of vertices (with replacement, so that the same vertex may appear more than once), where each choice is independent and vertices are selected with probability proportional to their degree.

Definition of process

Fix integers r, s with $r \geq s \geq 1$.

We grow a tree, starting from the two-vertex tree at time 1. At each time step we select an ordered r-tuple of vertices (with replacement, so that the same vertex may appear more than once), where each choice is independent and vertices are selected with probability proportional to their degree.

We then add a new vertex attached to the vertex with rank s (by degree) among the r vertices chosen,

Definition of process

Fix integers r, s with $r \geq s \geq 1$.

We grow a tree, starting from the two-vertex tree at time 1. At each time step we select an ordered r-tuple of vertices (with replacement, so that the same vertex may appear more than once), where each choice is independent and vertices are selected with probability proportional to their degree.

We then add a new vertex attached to the vertex with rank s (by degree) among the r vertices chosen, breaking ties uniformly at random.

Definition of process

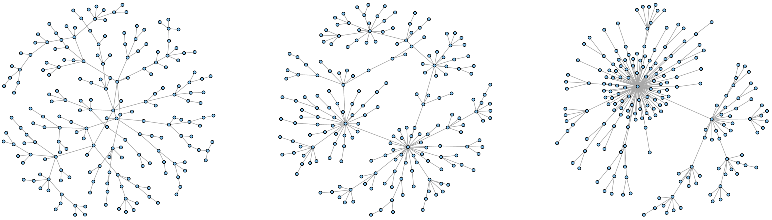
Fix integers r, s with $r \geq s \geq 1$.

We grow a tree, starting from the two-vertex tree at time 1. At each time step we select an ordered r-tuple of vertices (with replacement, so that the same vertex may appear more than once), where each choice is independent and vertices are selected with probability proportional to their degree.

We then add a new vertex attached to the vertex with rank s (by degree) among the r vertices chosen, breaking ties uniformly at random.

We will generally think of r being at least 2, since the case $r = s = 1$ is standard preferential attachment.

Pictures



200-vertex simulations, produced using igraph in R. Left to right: $r = 2, s = 2$, standard preferential attachment, $r = 2, s = 1$. The maximum degrees in these simulations are are 6, 30 and 90 respectively.

Krapivsky & Redner

Krapivsky & Redner investigated the degree distribution for this and some similar models non-rigorously. They observed three possibilities:

Krapivsky & Redner

Krapivsky & Redner investigated the degree distribution for this and some similar models non-rigorously. They observed three possibilities:

- A non-degenerate limit distribution with a heavy tail (power law or similar), similar to standard preferential attachment.

Krapivsky & Redner

Krapivsky & Redner investigated the degree distribution for this and some similar models non-rigorously. They observed three possibilities:

- A non-degenerate limit distribution with a heavy tail (power law or similar), similar to standard preferential attachment.
- A dominant vertex, with degree of the same order as the size of the graph.

Krapivsky & Redner

Krapivsky & Redner investigated the degree distribution for this and some similar models non-rigorously. They observed three possibilities:

- A non-degenerate limit distribution with a heavy tail (power law or similar), similar to standard preferential attachment.
- A dominant vertex, with degree of the same order as the size of the graph.
- A non-degenerate limit distribution with a doubly exponential tail.

Krapivsky & Redner continued

They suggested:

- When $s = 1$ ("greedy choice") a dominant vertex occurs if $r \geq 3$, and a degree distribution with tail decay $(n \log n)^{-2}$ if $r = 2$;

Krapivsky & Redner continued

They suggested:

- When $s = 1$ ("greedy choice") a dominant vertex occurs if $r \geq 3$, and a degree distribution with tail decay $(n \log n)^{-2}$ if $r = 2$;
- When $s > 1$ ("meek choice") doubly exponential decay happens whatever the values of r and s.

Krapivsky & Redner continued

They suggested:

- When $s = 1$ ("greedy choice") a dominant vertex occurs if $r \geq 3$, and a degree distribution with tail decay $(n \log n)^{-2}$ if $r = 2$;
- When $s > 1$ ("meek choice") doubly exponential decay happens whatever the values of r and s. (Even for r large and $s = 2$.)

Malyshkin & Paquette: $r = 2$

Malyshkin & Paquette rigorously investigated the cases $r = 2, s = 2$ ("min choice") and $s = 1$ ("max choice").

Malyshkin & Paquette: $r = 2$

Malyshkin & Paquette rigorously investigated the cases $r = 2, s = 2$ ("min choice") and $s = 1$ ("max choice").

Their results match Krapivsky & Redner's in these cases: doubly exponential decay for $r = 2, s = 2$, $(n \log n)^{-2}$ decay for $r = 2, s = 1$, and a dominant vertex for $r > 2, s = 1$.

Our results

We show:

- The proportion of vertices with degree at most k converges to a limit, p_k, almost surely.

Our results

We show:

- The proportion of vertices with degree at most k converges to a limit, p_k, almost surely.
- For a given s, there exists $r(s)$ such that $\lim_{k\to\infty} p_k = 1$ if and only if $r < r(s)$.

Our results

We show:

- The proportion of vertices with degree at most k converges to a limit, p_k, almost surely.
- For a given s, there exists $r(s)$ such that $\lim_{k\to\infty} p_k = 1$ if and only if $r < r(s)$. Note that a non-degenerate limiting degree distribution requires $\lim_{k\to\infty} p_k = 1$.
- For $s = 2$, $r(s) = 7$.

Our results

We show:

- The proportion of vertices with degree at most k converges to a limit, p_k, almost surely.
- For a given s, there exists $r(s)$ such that $\lim_{k\to\infty} p_k = 1$ if and only if $r < r(s)$. Note that a non-degenerate limiting degree distribution requires $\lim_{k\to\infty} p_k = 1$.
- For $s = 2$, $r(s) = 7$. In particular, a non-degenerate limiting degree distribution does not exist if $r \geq 7, s = 2$.

Convergence of degree proportions: notation

Define $B_{r,s}(p)$ to be the probability that a $Bin(r,p)$ random variable takes a value greater than $r - s$.

Convergence of degree proportions: notation

Define $B_{r,s}(p)$ to be the probability that a $Bin(r,p)$ random variable takes a value greater than $r-s$.

Write $F_m(k)$ for the sum of degrees of vertices with degree at most k at time m.

Convergence of degree proportions: notation

Define $B_{r,s}(p)$ to be the probability that a $Bin(r, p)$ random variable takes a value greater than $r - s$.

Write $F_m(k)$ for the sum of degrees of vertices with degree at most k at time m.

Then $F_m(k)/2m$ is the probability of selecting a vertex of degree at most k with a single preferential choice.

Evolution of $F_m(k)$

Write

$$f_k(x,p) = (k+1)B_{r,s}(p) - kB_{r,s}(x) - 2x + 1$$

for $x, p \in [0,1]$ and $k \geq 0$.

Evolution of $F_m(k)$

Write

$$f_k(x,p) = (k+1)B_{r,s}(p) - kB_{r,s}(x) - 2x + 1$$

for $x, p \in [0,1]$ and $k \geq 0$.

By considering the degree of the vertex selected as neighbour to the new vertex, we find $\mathbb{E}\left(\frac{F_{m+1}(k)}{2(m+1)} - \frac{F_m(k)}{2m} | \mathcal{F}_m\right)$ is

$$\frac{1}{2(m+1)} f_k\left(\frac{F_m(k)}{2m}, \frac{F_m(k-1)}{2m}\right).$$

The p_k

This suggests that if $\frac{F_m(k)}{2m} \to p_k$ a.s. as $m \to \infty$, we expect $f_k(p_k, p_{k-1}) = 0$: ideas similar to *stochastic approximation* processes.

For any $p \in [0, 1)$ there is a unique $x \in (0, 1)$ with $f_k(x, p) = 0$.

The p_k

This suggests that if $\frac{F_m(k)}{2m} \to p_k$ a.s. as $m \to \infty$, we expect $f_k(p_k, p_{k-1}) = 0$: ideas similar to *stochastic approximation* processes.

For any $p \in [0, 1)$ there is a unique $x \in (0, 1)$ with $f_k(x, p) = 0$.

Define $(p_k)_{k \geq 0}$ by setting $p_0 = 0$ and for each $k \geq 0$ letting p_k be the unique value in $(0, 1)$ such that $f_k(p_k, p_{k-1}) = 0$.

The p_k

This suggests that if $\frac{F_m(k)}{2m} \to p_k$ a.s. as $m \to \infty$, we expect $f_k(p_k, p_{k-1}) = 0$: ideas similar to *stochastic approximation* processes.

For any $p \in [0, 1)$ there is a unique $x \in (0, 1)$ with $f_k(x, p) = 0$.

Define $(p_k)_{k \geq 0}$ by setting $p_0 = 0$ and for each $k \geq 0$ letting p_k be the unique value in $(0, 1)$ such that $f_k(p_k, p_{k-1}) = 0$.

Stochastic approximation intuition now suggests $\frac{F_m(k)}{2m} \to p_k$ a.s. as $m \to \infty$. More precise results in paper.

The limit of the p_k

Theorem

- *The sequence p_k is increasing with limit $p_* \leq 1$, where p_* is the smallest positive root of* $f(p) = B_{r,s}(p) - 2p + 1 = 0$.

The limit of the p_k

Theorem

- *The sequence* p_k *is increasing with limit* $p_* \leq 1$, *where* p_* *is the smallest positive root of* $f(p) = B_{r,s}(p) - 2p + 1 = 0$.
- *There exists a function* $r(s)$ *such that* $p_* = 1$ *if and only if* $r < r(s)$, *which satisfies* $r(s) = 2s + o(s)$ *but also* $r(s) = 2s + \omega(\sqrt{s})$.

The limit of the p_k

Theorem

- *The sequence p_k is increasing with limit $p_* \leq 1$, where p_* is the smallest positive root of* $f(p) = B_{r,s}(p) - 2p + 1 = 0$.
- *There exists a function $r(s)$ such that $p_* = 1$ if and only if $r < r(s)$, which satisfies $r(s) = 2s + o(s)$ but also* $r(s) = 2s + \omega(\sqrt{s})$.
- *Provided $s \geq 2$, if $p_* = 1$ then* $-\log(1 - p_k) = \Omega(s^k)$.

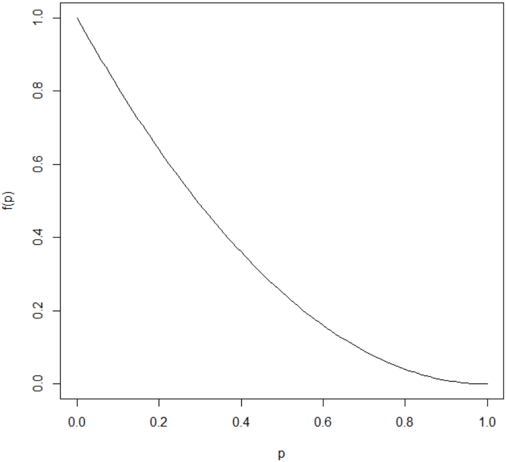
The limit of the p_k

Theorem

- *The sequence* p_k *is increasing with limit* $p_* \leq 1$, *where* p_* *is the smallest positive root of* $f(p) = B_{r,s}(p) - 2p + 1 = 0$.
- *There exists a function* $r(s)$ *such that* $p_* = 1$ *if and only if* $r < r(s)$, *which satisfies* $r(s) = 2s + o(s)$ *but also* $r(s) = 2s + \omega(\sqrt{s})$.
- *Provided* $s \geq 2$, *if* $p_* = 1$ *then* $-\log(1 - p_k) = \Omega(s^k)$.
- *The only other case with* $r > 1$ *where* $p_* = 1$ *is* $r = 2, s = 1$, *and then* $1 - p_k = (2 + o(1))/\log k$.

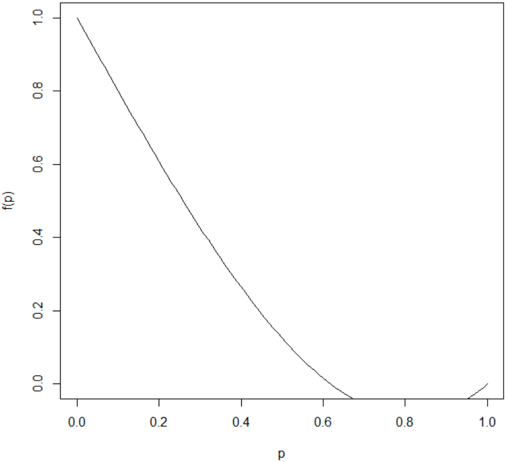
Plots of $f(p)$

$r = 2, s = 1$



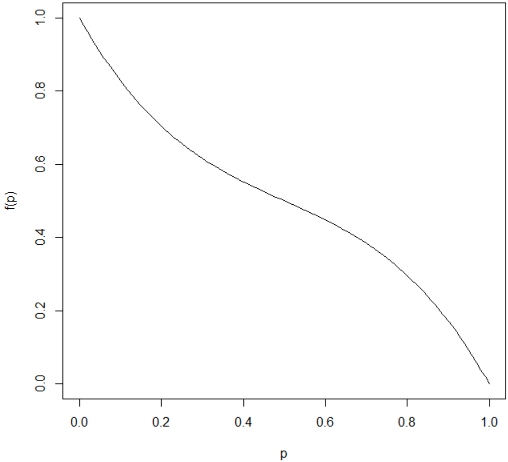
Plots of $f(p)$

$r = 3, s = 1$



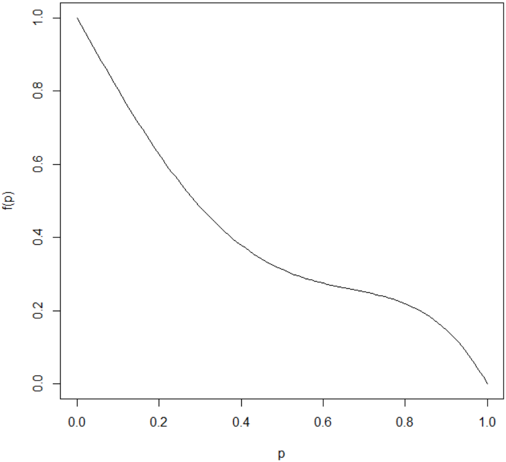
Plots of $f(p)$

$r = 3, s = 2$



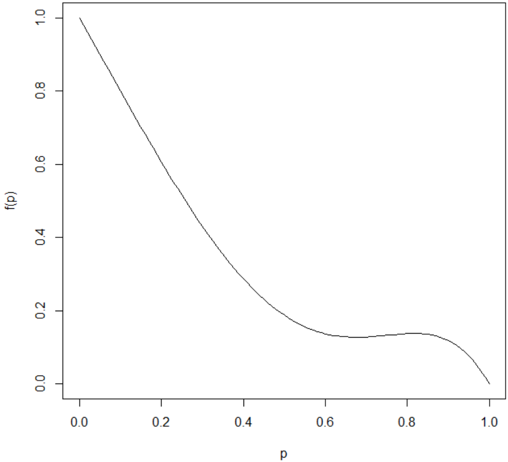
Plots of $f(p)$

$r = 4, s = 2$



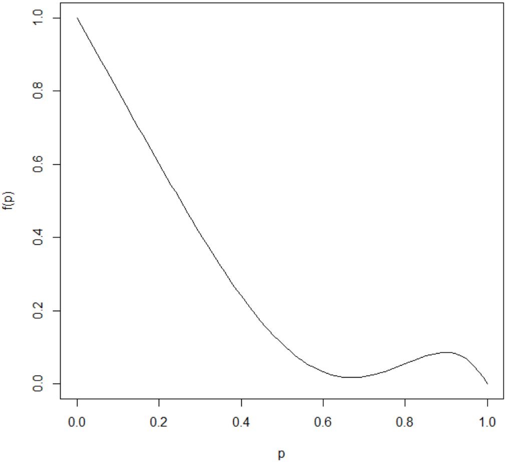
Plots of $f(p)$

$r = 5, s = 2$



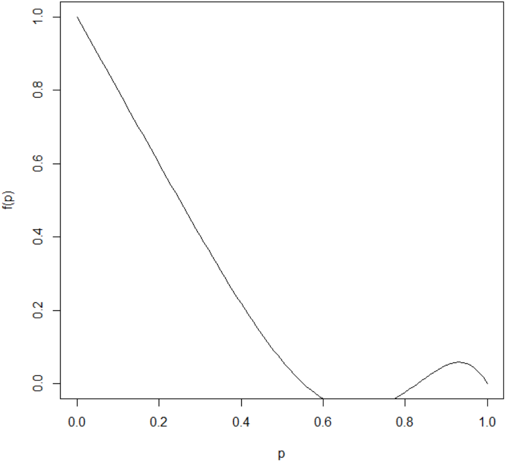
Plots of $f(p)$

$r = 6, s = 2$



Plots of $f(p)$

$r = 7, s = 2$



Non-degenerate limit or not

Pictures indicate $r(2) = 7$: no non-degenerate limiting distribution for second largest choice if $r \geq 7$.

Non-degenerate limit or not

Pictures indicate $r(2) = 7$: no non-degenerate limiting distribution for second largest choice if $r \geq 7$.

Similarly $r(3) = 10, r(4) = 13, r(5) = 16, r(6) = 19$

Non-degenerate limit or not

Pictures indicate $r(2) = 7$: no non-degenerate limiting distribution for second largest choice if $r \geq 7$.

Similarly $r(3) = 10, r(4) = 13, r(5) = 16, r(6) = 19$ but $r(7) = 21\ldots$.

Non-degenerate limit or not

Pictures indicate $r(2) = 7$: no non-degenerate limiting distribution for second largest choice if $r \geq 7$.

Similarly $r(3) = 10, r(4) = 13, r(5) = 16, r(6) = 19$ but $r(7) = 21\ldots$.

In fact we can show $r(s) > 2s$ and $r(s)/s \to 2$ but $s^{-1/2}(r(s) - 2s) \to \infty$, by further analysis of the function f.

Doubly exponential decay

When $s > 1$ and $p_k \to 1$, we show doubly exponential decay of $q_k = 1 - p_k$.

Doubly exponential decay

When $s > 1$ and $p_k \to 1$, we show doubly exponential decay of $q_k = 1 - p_k$.

Step 1: If $p_k \to 1$, then for k sufficiently large q_k satisfies

$$q_k < \frac{k+1}{2} \binom{r}{s} q_{k-1}^s \,.$$

Doubly exponential decay

When $s > 1$ and $p_k \to 1$, we show doubly exponential decay of $q_k = 1 - p_k$.

Step 1: If $p_k \to 1$, then for k sufficiently large q_k satisfies

$$q_k < \frac{k+1}{2}\binom{r}{s} q_{k-1}^s \,.$$

Step 2: Step 1 implies doubly-exponential decay provided we can find some k_0 with

$$q_{k_0} < \left(\frac{2}{\binom{r}{s}(k_0+3)}\right)^{1/(s-1)} . \tag{1}$$

Doubly exponential decay continued

Step 3: $q_k \to 0$ implies $q_k = O(1/k^2)$ (more than enough for (1)).

Doubly exponential decay continued

Step 3: $q_k \to 0$ implies $q_k = O(1/k^2)$ (more than enough for (1)).

For $r = 2, s = 2$, (1) satisfied for $k_0 = 4$,

Doubly exponential decay continued

Step 3: $q_k \to 0$ implies $q_k = O(1/k^2)$ (more than enough for (1)).

For $r = 2, s = 2$, (1) satisfied for $k_0 = 4$, for $r = 3, s = 2$ for $k_0 = 18$, for $r = 4, s = 2$ for $k_0 = 98$, for $r = 5, s = 2$ for $k_0 = 2416$. For $r = 6, s = 2$ $k_0 > e^{23}$.

Doubly exponential decay continued

Step 3: $q_k \to 0$ implies $q_k = O(1/k^2)$ (more than enough for (1)).

For $r = 2, s = 2$, (1) satisfied for $k_0 = 4$, for $r = 3, s = 2$ for $k_0 = 18$, for $r = 4, s = 2$ for $k_0 = 98$, for $r = 5, s = 2$ for $k_0 = 2416$. For $r = 6, s = 2$ $k_0 > e^{23}$.

Simulations don't easily distinguish between doubly exponential decay above very large threshold and a dominant vertex.

References

P. L. Krapivsky and S. Redner, Choice-driven phase transition in complex networks, *Journal of Statistical Mechanics: Theory and Experiment*, 2014.

Yu. Malyshkin and E. Paquette, The power of 2 choices over preferential attachment.

Yu. Malyshkin and E. Paquette, The power of choice combined with preferential attachment, *Electronic Communications in Probability*, 2014.

J. Haslegrave and J. Jordan, Preferential attachment with choice, *Random Structures and Algorithms*, 2015/16.