Branching processes and random bipartite networks

Michael Drmota, Bernhard Gittenberger, Tyll Krüger, RS^2
TU Wien, Univ. Bielefeld, Berlin

October 22, 2009

1 Some branching process basics

Branching processes are a basic stochastic model, they have applications in many fields, from population dynamics to nuclear physics. The simplest branching model is called Galton-Watson process.

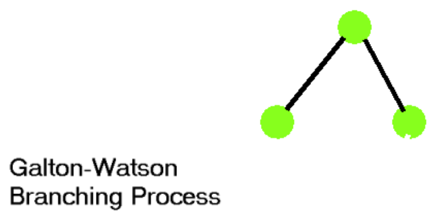
Two assumptions are made:

- we have particles, which individually produce offspring particles during oen step of time, where the probability for a particle to have k descendants is $p_k, k = 0, 1, 2, \ldots$ and

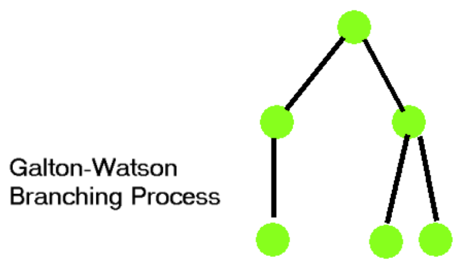
- the offspring generation of different particles is statistically independent

So, starting with one particle...

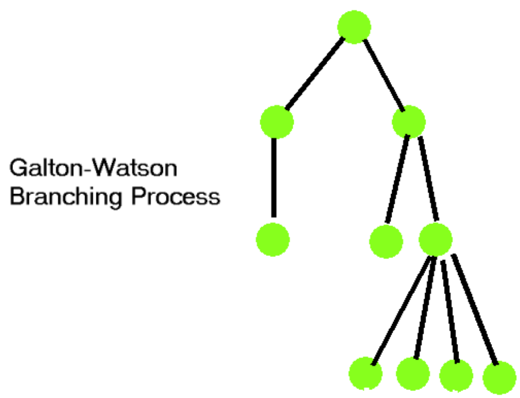
... we get a random number of daughters....



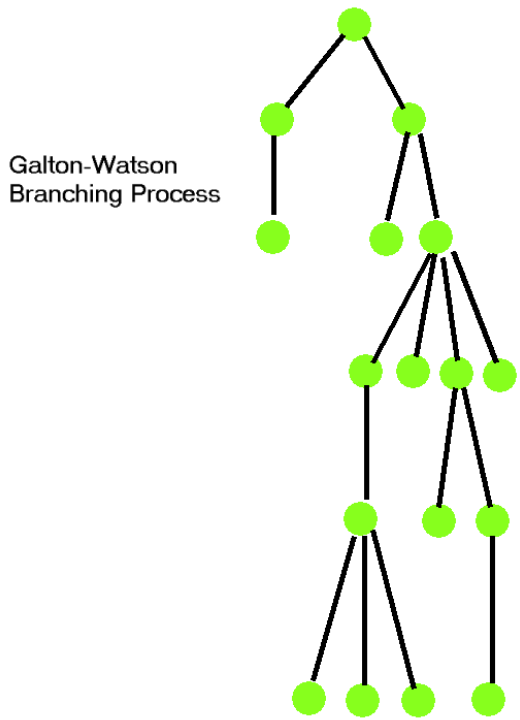
... these independently produce grand-daughters....



...then particles of third generation...



...and on.



As it turns out, the principal behaviour of that simple kind of model depends only on one parameter: the expected number of offspring:

$$M = \sum_{k=0}^{\infty} k p_k.$$

This M is also called the Malthus parameter of the branching model. The behaviour is:

- if $M \leq 1$, then the population goes extinct with certainty, and
- if $M > 1$, the population survives forever with a positive probability, and in case of non-extinction it explodes exponentially fast.

Hence, there is no stability in that simple model, not even in the case $M = 1$, where each particle produces one descendant in the average (the population dies out eventually).

Of course, real population dynamics cannot be described in such a simple scheme.

We have developed several simulation programs (BRANLANG, BRAP, SPREAD) in order to study more sophisticated branching schemes with a kind of 'birth control' implemented to control the population size. These programs also allow to trace genealogical lines or spatial spread.

But the scheme produces random trees , and is very useful as a building block in several random graph models.. It defines important features of these models, for instance the appearance of a giant component.

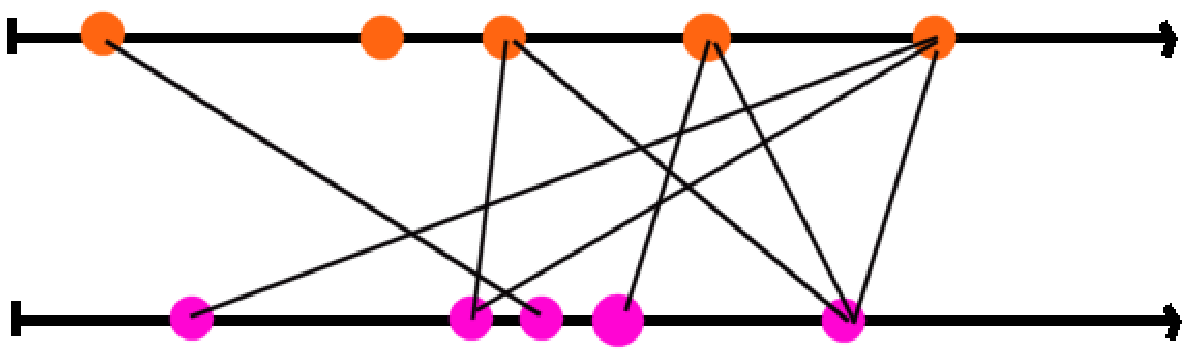
2 The bipartite random graph model

We intend to investigate a stochastic model for bipartite networks, e.g.

- projects-organizations
- articles-authors
- actors-movies
- ...

So we consider undirected graphs $G = (V_{\mathsf{X}} \cup V_{\mathsf{Y}}, E)$ with $V_{\mathsf{X}} \cap V_{\mathsf{Y}} = \varnothing$ and edges only between V_{X} and V_{Y}: $E \subset V_{\mathsf{X}} \times V_{\mathsf{Y}}$.

We have X-individuals and Y-individuals and links between them:



This is the general picture, and as the drawing indicates, individuals of both types have different 'attractivities', again denoted by x and y. We assume $x, y \in \mathbb{R}_+$.

This is a special case of the general asymptotic random graph model with independent edges considered by Bollobás, Janson and Riordan [BJR 2005]

This general model (which was initiated by Söderberg and Turova) works as follows:

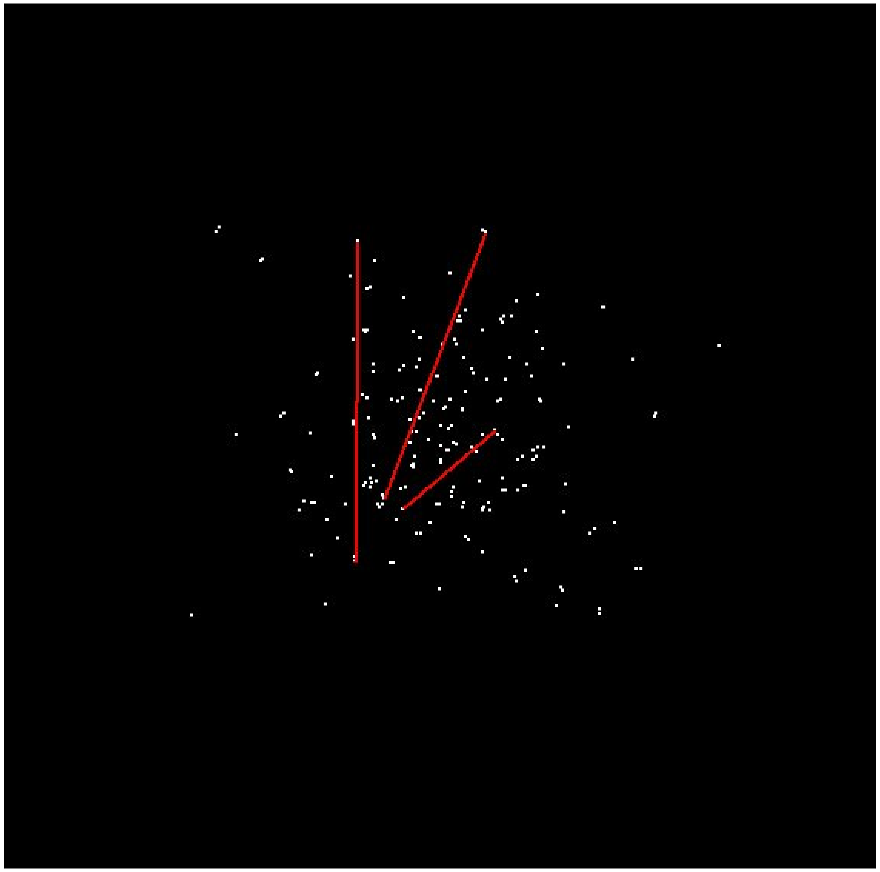
- Take a complete separabe metric space S equipped with a Borel probability measure μ, and consider a kernel $\kappa \in L_1(\mu \times \mu)$. Roughly, the measure μ is used to throw points into S, and the kernel is used to construct edges between them.

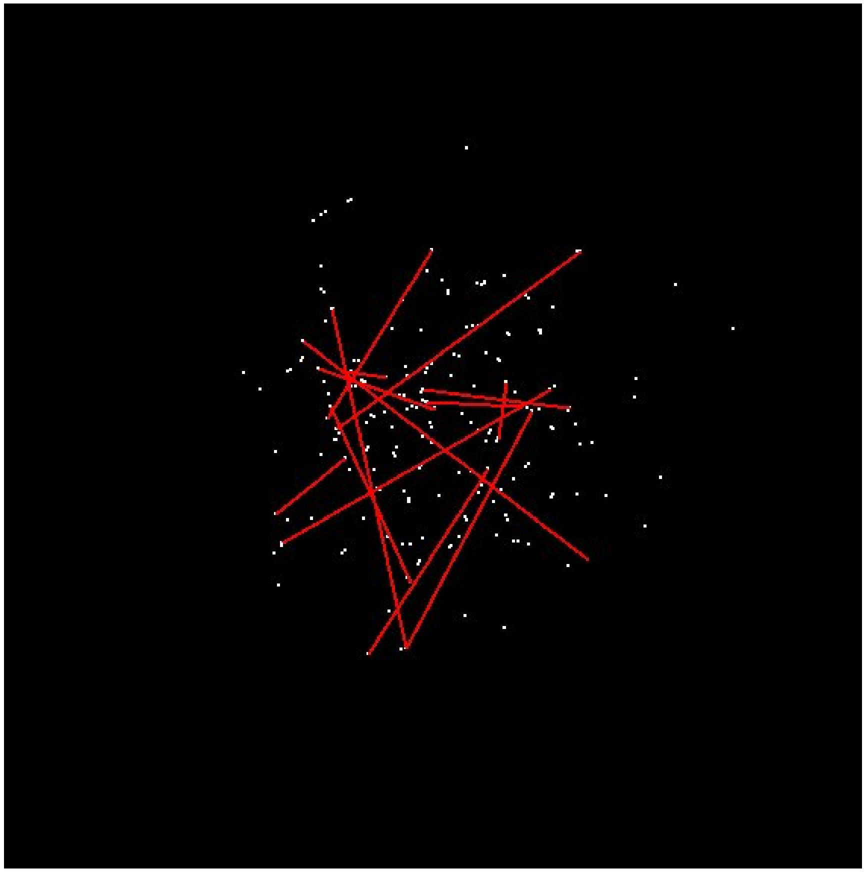
- The model is asymptotic in the sense that we consider graphs $\mathcal{G}$ obtained by first choosing a large number n of vertices $V^{(n)} = \{x_1, x_2, ..., x_n\}$ such that $\frac{1}{n}\sum_i \delta_{x_i} \sim \mu$ (in an appropriate sense referring to the weak topology in the set of probability measures).

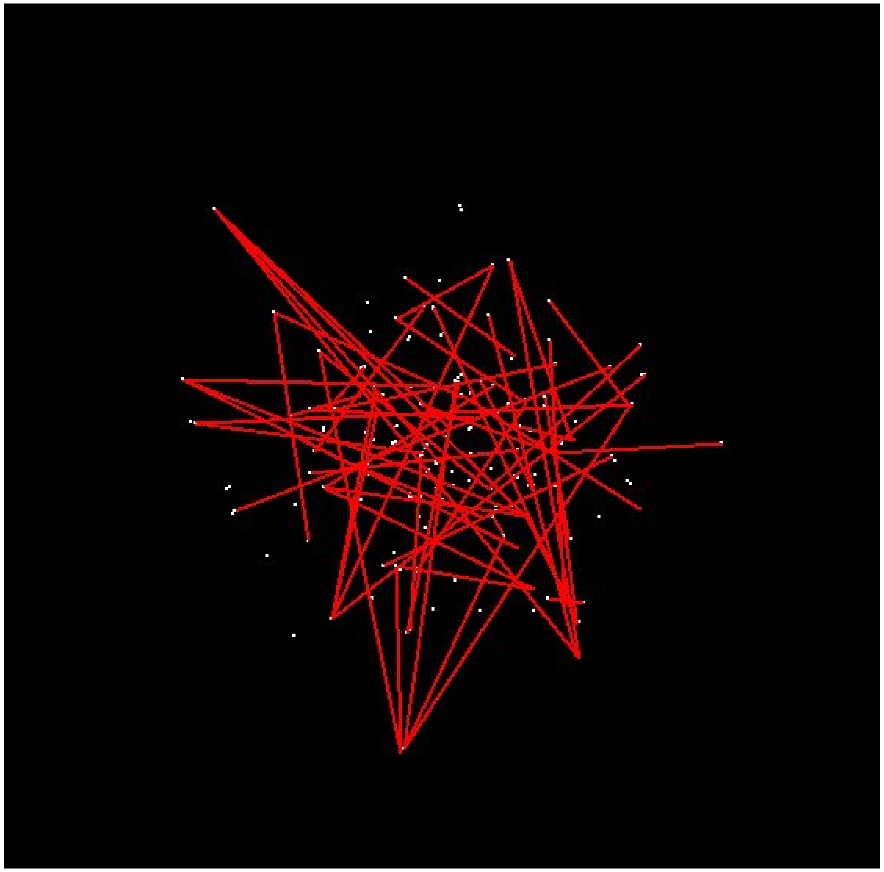
- Given $V^{(n)}$, we connect any two vertices x_i, x_j by an edge with the infinitesimal probability $\frac{1}{n}\kappa(x_i, x_j)$, and we do this independently for all possible couples.

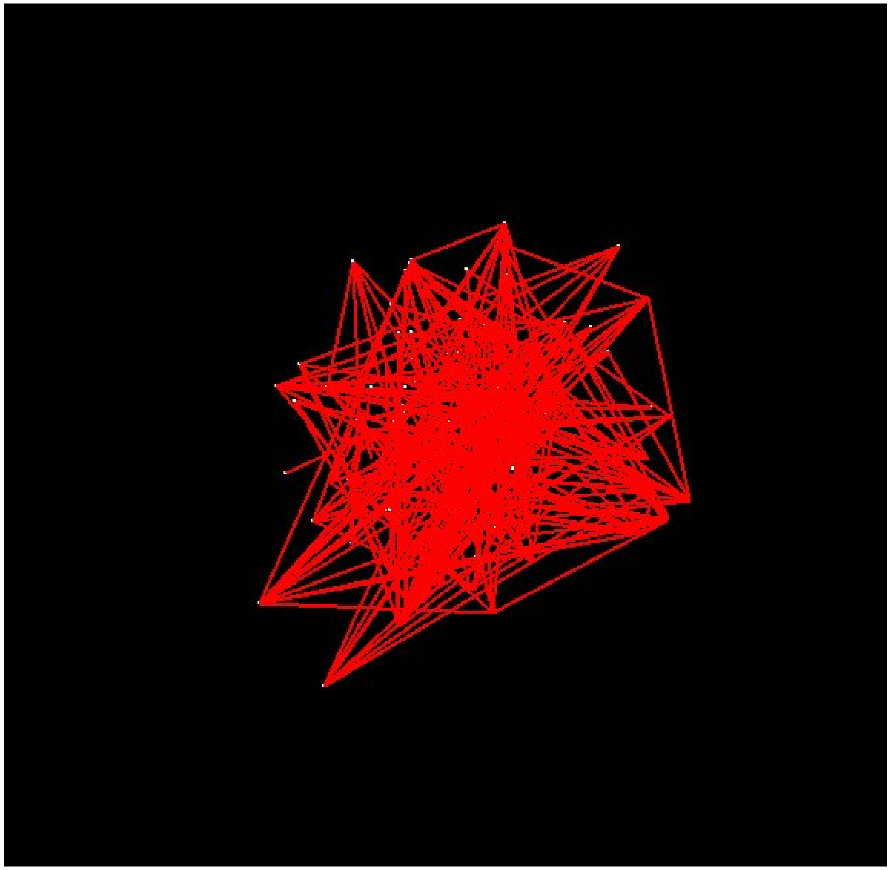
This is a generalization of the classical Erdős-Rényi model, which we get for S being a one-point set.

Under some regularity conditions, many interesting results were obtained in the BJR paper, such like the existence of a phase transition: If we consider kernels $\kappa = \lambda\kappa_0$, there exists a unique **giant component** of size $O(n)$ for $\lambda > \lambda_c$, while for $\lambda < \lambda_c$ the largest component is of size $O(\log n)$.









We consider a bipartite version of this model. This means the set of possible vertices S splits up into two disjoint sets $S = S_{\mathsf{X}} \cup S_{\mathsf{Y}}$, and there are only links connecting X-individuals with Y-individuals, but not inside X or Y. Moreover, we choose $S_{\mathsf{X}}, S_{\mathsf{Y}}$ as disjoint copies of $\mathbb{R}_+$ and interpret the position of a node in $\mathbb{R}_+$ as its attractivity.

So, let there be given

- Two probability measures $\mu_{\mathsf{X}}, \mu_{\mathsf{Y}}$ on $\mathbb{R}_+$: distributions of attractivity

- A kernel $\kappa(x, y) \geq 0$, essentially $\kappa(x, y)$ is the infinitesimal probability of an edge between individuals with attractivities x, y, assume e.g. $\kappa(x, y) = c(x + y)$ or $\kappa(x, y) = cxy$ or $\kappa(x, y) = c \min(x, y)$, where c is a suitable constant. Which kernel is chosen depends on the specific way the 'attractivities' interact.

- Two positive numbers with $m_{\mathsf{X}} + m_{\mathsf{Y}} = 1$: gives the relation of the total numbers of X- and Y-individuals.

Next, construct a sequence of random bipartite graphs $\mathcal{G}(n) = \mathcal{G}^{(m_{\mathsf{X}}\mu_{\mathsf{X}}, m_{\mathsf{Y}}\mu_{\mathsf{Y}}, \kappa)}(n)$, with $n \to \infty$ being the total number of X+Y-individuals:

- Take $m_{\mathsf{X}} \cdot n$ individuals and throw each of them by chance to the X-axis, according to the probability measure μ_{X}. So, if n is large, the empirical distribution of attractivity will be close to μ_{X}, by the law of large numbers. Take the remaining $m_{\mathsf{Y}} \cdot n$ individuals and throw them to the Y-axis according to μ_{Y}. So we get two random point configurations, the two kinds of nodes $V_{\mathsf{X}}^{(n)}, V_{\mathsf{Y}}^{(n)}$ of our bipartite graph.

Remark: It is in fact not so important, that the nodes were obtained by a random procedure, the essential requirement is the asymptotics of the empirical distribution of attractivities (μ_X and μ_Y). But in the construction of the links/edges we need randomization, because we want to use techniques from probability theory:

- For any pair (x, y) of nodes of $V_{\mathsf{X}}^{(n)}, V_{\mathsf{Y}}^{(n)}$, connect them by an edge with probability $\frac{1}{n}\kappa(x, y)$ (to be completely precise, with probability $\min(\frac{1}{n}\kappa(x, y), 1)$), do this completely independently for different pairs.

So far, we are in a special situation of the Bollobás et al. paper. In particular, $\mu = m_{\mathsf{X}}\mu_{\mathsf{X}} + m_{\mathsf{Y}}\mu_{\mathsf{Y}}$, with μ_{X} and μ_{Y} 'sitting' on disjoint copies of $\mathbb{R}_+$. Just as in BJR, we impose two mild regularity conditions:

i) κ is continuous

(or at least $\mu_{\mathsf{X}} \times \mu_{\mathsf{Y}}$ a.e. continuous)

ii) $\displaystyle\int\int \kappa(x, y)\mu_{\mathsf{X}}(dx)\mu_{\mathsf{Y}}(dy) < \infty$

The kernel κ together with $m_{\mathsf{X}}\mu_{\mathsf{X}}, m_{\mathsf{Y}}\mu_{\mathsf{Y}}$ specifies the degree distributions of the X- and Y-individuals: By the Poisson limit law an X-individual at $x \in \mathbb{R}_+$ has asymptotically a Poisson distributed number of Y-neighbours with expectation $m(x) := m_{\mathsf{Y}} \int \kappa(x,y)\mu_{\mathsf{Y}}(dy)$, which by ii) is almost surely finite. Hence the degree distribution $Z_x^{(n)}$ fulfils

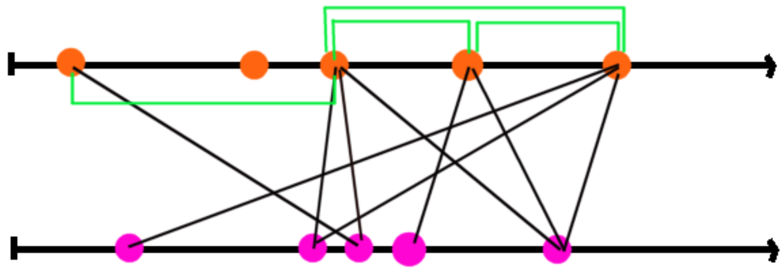
$$Z_x^{(n)} \xrightarrow[\text{in law}]{} \Pi_{m(x)}.$$

So the total degree distribution of X-individuals is asymptotically mixed Poisson:

$$Z^{(n)} \xrightarrow[\text{in law}]{} \int \Pi_{m(x)}\mu_{\mathsf{X}}(dx).$$

(The corresponding results hold for Y-individuals.)

The general theory may be applied to derive a corresponding threshold for the appearance of a giant component , local tree-like structure of components, the size of the second-largest component and so on.



3 Projected graph

In our investigations, we are especially interested in the projected graph :

- Projection: connect any two X-individuals by an edge, which have a common Y-neighbour

As can be seen, the projected graph contains cliques : All X-individuals, which are connected to the same Y-individual, form a complete sub-graph. So, unlike the original bipartite graph, the projection graph's edge structure is no longer independent, and hence does not immediately fit into the Bollobás, Janson & Riordan scheme.

This observation may lead to the idea, that by means of procedures similar to our 'projection', the class of random graphs which are covered and can be analyzed (asymptotically) by this BJR scheme, is much larger than those with independent edges.

The random graph structure generated on X is obtained from an BJR-structured graph with a second 'hidden' set of nodes Y, in a similar way as hidden Markov processes are derived from Markov processes and are more general.

So, if we have a real-world graph with a structure which is not compatible with the idea of independent edges, in the sense of model fitting it still may fit well with the assumption it is the projection of a BJR-graph, and one might try to find/interpret the hidden component.

For the projected graph, we are interested in the asymptotic degree distribution.

In applications, the degree distribution often shows some **power-law** shape (sometimes called 'heavy tail'). More specifically,

Definition: A probability distribution P (on $\mathbb{R}_+$) is said to have a **regularly varying tail of index** α iff

$$P([x, +\infty)) = x^{-\alpha} L(x),$$

where L is slowly varying, i.e. $L(tx)/L(x) \underset{x\to\infty}{\longrightarrow} 1$ for each $t > 0$.

We also consider the case $\alpha = \infty$:

Definition: A probability distribution P (on $\mathbb{R}_+$) is said to have a tail of index ∞ (**thin tail**) iff

$$P([x, +\infty)) = x^{-R(x)},$$

with $R(x) \to \infty$ as $x \to \infty$.

Consider now the case, where the X-individuals' distribution μ_{X} has a **regularly varying tail** of index α (or a thin tail $\alpha = \infty$), and μ_{Y} has a **regularly varying tail** of index β (or a thin tail $\beta = \infty$). Then we have

Theorem: If $\kappa(x, y) = cxy$ or $\kappa(x, y) = c(x + y)$, and $\alpha, \beta > 1$, then the asymptotic degree distribution of the projected graph has a tail of index $\min(\alpha, \beta - 1)$, supposed the following is fulfilled:

(**ugly** condition): It is **not** $\alpha = \beta - 1 \in \mathbb{N}$.

This result is based on the following assertion about a two-step branching process:

Theorem (branching composition tail): Let P, Q be two probability distributions on $\mathbb{Z}_+$ with regularly varying tails of index α and β, respectively. Then the distribution $P \circ Q := \sum_n Q^{*n} P(\{n\})$ of the second generation of a two-step branching process, starting with P and taking Q for the second step, has a regularly varying tail of index $\min(\alpha, \beta\,, \alpha\beta)$, supposed

$$\text{not } \alpha = \beta \in \mathbb{N}$$

is fulfilled.

This can be shown, using the well known relation between the tail behaviour of a regularly varying tail probability distribution on $[0, +\infty)$, and the remainder term of the Taylor expansion of its generating function around 1, where the latter result breaks down, unfortunately, for integer values of the parameters α, β. Some nice results of Stam (1973) cover the remaining cases, with the exception of the **ugly** case, that both are the same integer.

Remark1: Both Theorems are probably valid without this condition, but as far as we know, the Theorem on the branching composition tail is still a **conjecture** for $\alpha = \beta \in \mathbb{N}$. It is a little bit surprising that it is the case of integer parameters, which is difficult. There is a standard classical monograph by Bingham, Goldie and Teugels on 'Regular variation', but even those gave up in the integer case in an article about the tail index of the distribution of $\frac{1}{m^n} Z_n$, where Z_n is the n-th generation of a super-critical Galton-Watson process with expectation m. What can be shown without this 'ugly' condition is the following:

Consider a weaker version of the definition of a heavy tail distribution:

Definition: A probability distribution P (on $\mathbb{R}_+$) is said to have a **tail of index** α iff

$$P([x, +\infty)) = x^{-\alpha + r(x)},$$

with $r(x) \to 0$ as $x \to \infty$.

Then we have

Theorem (branching composition tail,1): Let P, Q be two probability distributions on $\mathbb{Z}_+$ with **regularly varying tails** of index α and β, respectively. Then the distribution $P \circ Q$ of the second generation of a two-step branching process, starting with P and taking Q for the second step, has a **tail** of index $\min(\alpha, \beta\,, \alpha\beta)$.

As you see, it is not completely clear, whether the class of regularly varying tail distributions is closed under branching composition, but a possible leakage might only occur at the integers and will not leave the class of distributions with an indexed tail.

Remark2: The term $\alpha\beta$, which is relevant only if $\alpha, \beta < 1$, cannot occur for the additive and multiplicative kernels, since the basic condition ii) of the BJR model implies $\alpha, \beta \geq 1$. But for the kernel $\kappa(x,y) = c\,\mathsf{min}(x,y)$ this interesting behaviour, that the degree distribution of the projected graph has a heavier tail, can be expected for $\alpha, \beta < 1$.

Some words about the proof of the Theorem about the degree distribution.

- First note that the next neighbours in the projected graph are the next-next neighbours in the original bipartite graph.

- Starting from a node at $x \in V_{\mathsf{X}}$, the distribution of its (random) degree is asymptotically (in n) Poisson with expectation

$$m(x) := x m_{\mathsf{Y}} \int y \; \mu_{\mathsf{Y}}(dy) = x m_{\mathsf{Y}} \mathbb{E}\mu_{\mathsf{Y}}$$

by the 'law of small numbers, as mentioned above (if $\kappa(x,y)$ is the **multiplicative kernel**). This is its asymptotical degree distribution. If we choose x randomly by μ_{X} we get a mixed Poisson limit law $P_1 := \int \Pi_{m(x)} \mu_{\mathsf{X}}(dx)$ for the number of neighbours. It is not hard to see that this has the same tail index as μ_{X}, which is α.

- A randomly chosen Y-neighbour of x is **not** distributed according to μ_{Y}, but according to $\widetilde{\mu}_{\mathsf{Y}} := \frac{1}{\mathbb{E}\mu_{\mathsf{Y}}} \int_{(\cdot)} y \;\mu_{\mathsf{Y}}(dy)$, since neighbours with high attractivity y are **preferred**. So in the second step, not μ_{Y} having tail index β is relevant, but $\widetilde{\mu}_{\mathsf{Y}}$, which has tail index $\beta - 1$.

- In the considered case of the multiplicative kernel xy, it is now rather easy to see, that the limit law of the number of next-next neighbours is simply $P_1 \circ P_2$, where $P_2 := \frac{1}{Z} \int \Pi_{m(x)} y \mu_{\mathsf{Y}}(dy)$. Observe that in this

multiplicative case the overall limit distribution is a Galton-Watson type composition, though we have spatially inhomogeneous branching. Now mainly the Theorem on the branching composition tail applies. One has to be careful with the interchange of limit procedures.

The case of the **additive kernel** is a bit more involved, since the second neighbours random number is not simply the result of two Galton-Watson branching steps, but it can be treated similarly.