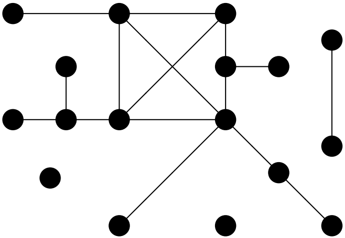
The greedy independent set in a random graph with given degrees

Svante Janson
(joint work with Graham Brightwell and Malwina Luczak)

In honour of Gösta Mittag-Leffler
Stockholm, 19 March 2016

Graphs

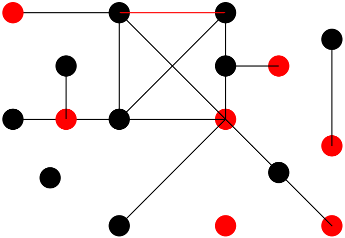
A graph is a set of *nodes* (or *vertices*) together with *edges* (or *links*), where each edge connects two nodes.



Independent set

A set of nodes in a graph is *independent* if there is no edge connecting two nodes in the set.

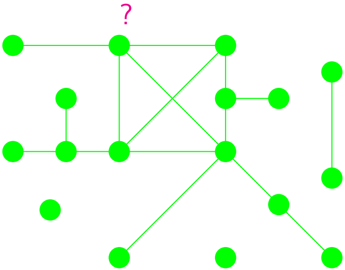
Example:

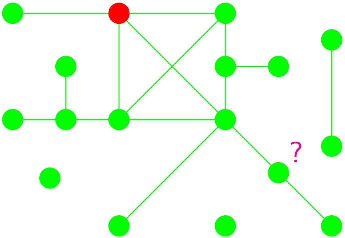


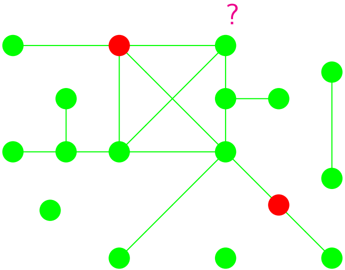
The red independent set has 7 nodes.

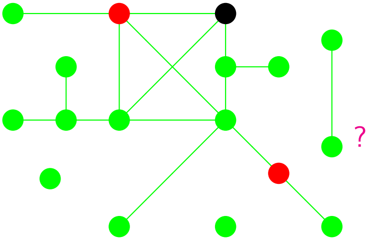
Greedy independent set

We construct a large (but typically not maximum size) independent set by inspecting the vertices in random order, and selecting every vertex that is not joined to a vertex already selected.

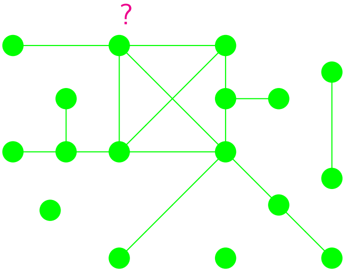


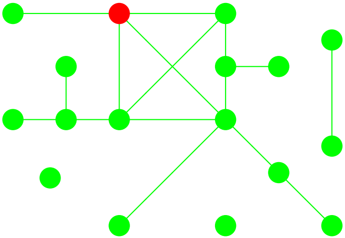


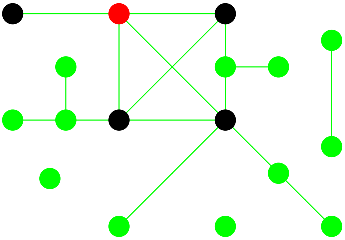


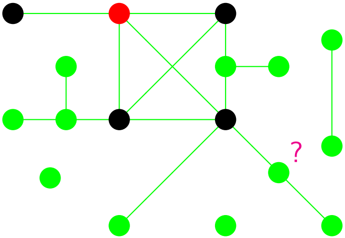


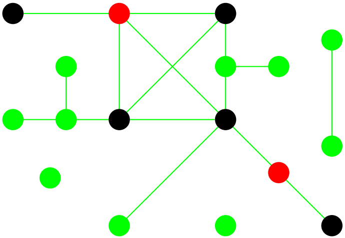
We modify a little, by marking a vertex as *blocked* if it is a neighbour of a *selected* vertex.

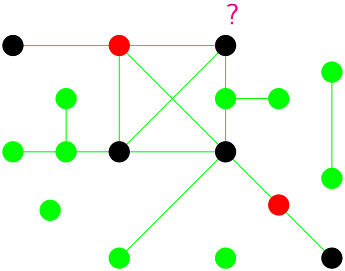


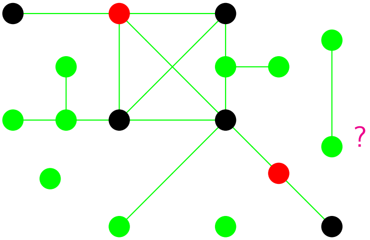












Random graphs with given degrees

Fix the node degrees as a given sequence $d_1, \ldots, d_n$, and take a random graph, uniformly among all possible graphs with these degrees.

(We assume that there are any such graphs. In particular, $\sum d_i$ must be even.)

The configuration model

We are interested in a sparse case (with average degree bounded), and then the standard method is the "*Configuration model*" by Bollobás (1980,1985) (see also Bender and Canfield (1978)).

- Assign d_i "half-edges" to node i, $i = 1, \ldots, n$.
- Pair the half-edges uniformly at random.
- Merge each pair of half-edges into an edge.

The configuration model

We are interested in a sparse case (with average degree bounded), and then the standard method is the "*Configuration model*" by Bollobás (1980,1985) (see also Bender and Canfield (1978)).

- Assign d_i "half-edges" to node i, $i = 1, \ldots, n$.
- Pair the half-edges uniformly at random.
- Merge each pair of half-edges into an edge.

This may construct multiple edges and loops, but if we condition on this not happening, we obtain a uniformly random graph with the given degrees. In the sparse case, the conditioning is no problem, and we may work with the random multigraph generated by the configuration model.

The configuration model

We are interested in a sparse case (with average degree bounded), and then the standard method is the "*Configuration model*" by Bollobás (1980,1985) (see also Bender and Canfield (1978)).

- Assign d_i "half-edges" to node i, $i = 1, \ldots, n$.
- Pair the half-edges uniformly at random.
- Merge each pair of half-edges into an edge.

This may construct multiple edges and loops, but if we condition on this not happening, we obtain a uniformly random graph with the given degrees. In the sparse case, the conditioning is no problem, and we may work with the random multigraph generated by the configuration model.

In the pairing step, we may process the half-edges one by one, in any order, if we like. This gives a useful flexibility.

Assumptions

We study asymptotics as $n \to \infty$. d_i generally depends on n.
Let

$$n_k = n_k(n) = \#\{i : d_i = k\}, \qquad k \in \{0, 1, 2, \dots\},$$

the number of vertices of degree k.
Let $(p_k)_0^\infty$ be a probability distribution, and assume that

$$n_k(n)/n \to p_k$$

for each k. Assume further that

$$\sum_k k^2 n_k(n) = O(n).$$

(Equivalently, the second moment of the degree distribution of a random vertex is uniformly bounded.) Then the distribution $(p_k)_0^\infty$ has a finite and mean $\lambda = \sum_k kp_k$ and the average vertex degree $\sum_k kn_k/n$ converges to λ (since the distribution of the degree of a random vertex is uniformly integrable). Assume also $\lambda > 0$.

Theorem

Let $S_\infty^{(n)}$ denote the size of a random greedy independent set. Let τ_∞ be the unique value in $(0,\infty]$ such that

$$\lambda \int_0^{\tau_\infty} \frac{e^{-2\sigma}}{\sum_k k p_k e^{-k\sigma}}\, d\sigma = 1.$$

Then

$$\frac{S_\infty^{(n)}}{n} \to \lambda \int_0^{\tau_\infty} e^{-2\sigma} \frac{\sum_k p_k e^{-k\sigma}}{\sum_k k p_k e^{-k\sigma}}\, d\sigma \quad \text{in probability.}$$

This limit is called the *jamming constant*.
This was proved, without an explicit formula for the jamming constant (and with somewhat stronger assumptions), by Bermolen, Jonckheere and Moyal (2013). The version above is by Brightwell, Janson and Luczak (2015).

Example

A random regular graph, where all nodes have a fixed degree $d \geq 2$. Then $p_d = 1$ and $\lambda = d$, and thus

$$1 = \int_{\sigma=0}^{\tau_\infty} e^{(d-2)\sigma}\, d\sigma = \frac{1}{d-2}(e^{(d-2)\tau_\infty} - 1),$$

for $d \geq 3$, and so $\tau_\infty = \frac{\log(d-1)}{d-2}$. For $d = 2$ we obtain $\tau_\infty = 1$. By the theorem, the jamming constant is

$$\int_0^{\tau_\infty} e^{-2\sigma}\, d\sigma = \frac{1}{2}(1 - e^{-2\tau_\infty}) = \frac{1}{2}\left(1 - \frac{1}{(d-1)^{2/(d-2)}}\right),$$

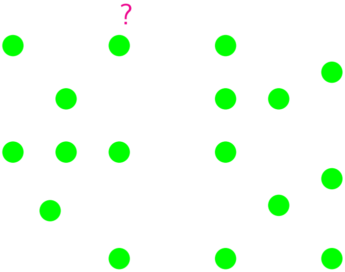
for $d \geq 3$ [Wormald, 1995], and $\frac{1}{2}(1 - e^{-2})$ for $d = 2$ [Flory, 1939].

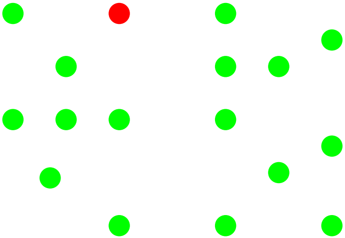
Proof

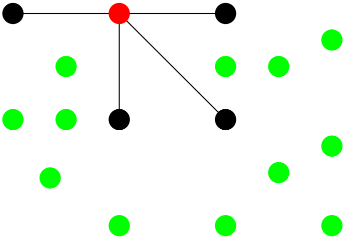
We use a standard method and reveal the edges as we go.

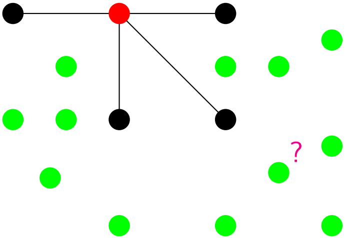
- Each vertex is given an independent exponential clock (rate 1).
- Initially, every vertex is *empty*.
- When the clock at an empty vertex rings, the vertex becomes *selected*, its edges are revealed, and its neighbours become *blocked*.
- When the clock at a blocked vertex rings, do nothing.

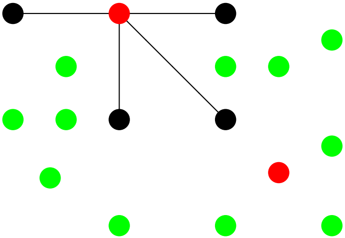
When revealing the edges at a vertex that is selected, we choose a partner of each half-edge at the vertex, uniformly at random among all free half-edges.

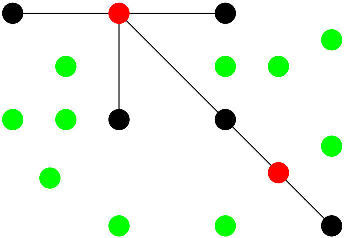


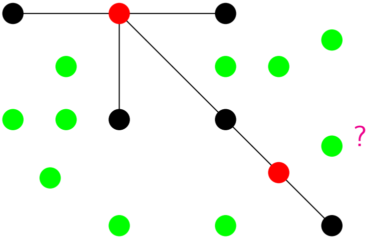












The variables that we track, for $0 \leq t \leq \infty$, are

- $E_t(k)$, the number of empty vertices of degree k at time t, for $k \geq 0$.
- U_t, the number of unpaired half-edges
- S_t, the number of vertices selected so far.

The vector $(U_t, E_t(0), E_t(1), \ldots, S_t)$ is a Markov process.

S_t has drift

$$\sum_{k=0}^{\infty} E_t(k).$$

In other words,

$$S_t - \int_0^t \sum_{k=0}^{\infty} E_s(k) ds.$$

is a martingale.

This is immediate, since S_t increases by 1 each time the clock at an empty vertex goes off, and they all go off with rate 1.

Similarly,
U_t has drift

$$-\sum_{k=1}^{\infty} kE_t(k)\left(2-\frac{k-1}{U_t-1}\right).$$

$E_t(k)$ has drift

$$-E_t(k)-\sum_{j=1}^{\infty} p_{jk}(U_t)E_t(j)\big(E_t(k)-\delta_{jk}\big),$$

where p_{jk} is the probability that in a configuration model with u half-edges, two vertices v and w of degrees j and k, respectively, are connected by at least one edge.

Thus, for example,

$$\frac{U_t}{n} = \frac{U_0}{n} - \int_0^t \sum_{k=1}^{\infty} \frac{kE_s(k)}{n} \Big(2 - \frac{k-1}{U_s - 1} \Big) \, ds + \frac{M_t}{n},$$

where M_t is a martingale.

Thus, for example,

$$\frac{U_t}{n} = \frac{U_0}{n} - \int_0^t \sum_{k=1}^{\infty} \frac{kE_s(k)}{n} \left(2 - \frac{k-1}{U_s - 1}\right) ds + \frac{M_t}{n},$$

where M_t is a martingale.

- An estimate of the quadratic variation shows, using Doob's inequality, that $\sup_t |M_t/n| \xrightarrow{\mathrm{p}} 0$.

Thus, for example,

$$\frac{U_t}{n} = \frac{U_0}{n} - \int_0^t \sum_{k=1}^{\infty} \frac{kE_s(k)}{n} \left(2 - \frac{k-1}{U_s - 1}\right) ds + \frac{M_t}{n},$$

where M_t is a martingale.

- An estimate of the quadratic variation shows, using Doob's inequality, that $\sup_t |M_t/n| \xrightarrow{\mathrm{p}} 0$.
- The integrand is uniformly bounded, so the integral is uniformly Lipschitz, and the Arzela–Ascoli theorem shows that the stochastic process defined by the integral is tight in $C[0,\infty)$. (I.e., in $C[0,t_0]$ for every t_0.)

Thus, for example,

$$\frac{U_t}{n} = \frac{U_0}{n} - \int_0^t \sum_{k=1}^{\infty} \frac{kE_s(k)}{n} \Big(2 - \frac{k-1}{U_s - 1}\Big)\, ds + \frac{M_t}{n},$$

where M_t is a martingale.

- An estimate of the quadratic variation shows, using Doob's inequality, that $\sup_t |M_t/n| \xrightarrow{\mathrm{p}} 0$.
- The integrand is uniformly bounded, so the integral is uniformly Lipschitz, and the Arzela–Ascoli theorem shows that the stochastic process defined by the integral is tight in $C[0,\infty)$. (I.e., in $C[0, t_0]$ for every t_0.)
- Hence, U_t/n is tight in $D[0,\infty)$.

Thus, for example,

$$\frac{U_t}{n} = \frac{U_0}{n} - \int_0^t \sum_{k=1}^{\infty} \frac{kE_s(k)}{n}\Big(2 - \frac{k-1}{U_s - 1}\Big)\, ds + \frac{M_t}{n},$$

where M_t is a martingale.

- An estimate of the quadratic variation shows, using Doob's inequality, that $\sup_t |M_t/n| \xrightarrow{\mathrm{p}} 0$.
- The integrand is uniformly bounded, so the integral is uniformly Lipschitz, and the Arzela–Ascoli theorem shows that the stochastic process defined by the integral is tight in $C[0,\infty)$. (I.e., in $C[0,t_0]$ for every t_0.)
- Hence, U_t/n is tight in $D[0,\infty)$.
- Similarly for $E_t(k)/n$.

Thus, at least along a subsequence, U_t/n and each $E_t(k)/n$ converge in distribution in $D[0,\infty)$ to some limits u_t and $e_t(k)$. By the Skorohod representation theorem, we may assume

$$U_t/n \to u_t, \qquad E_t(k)/n \to e_t(k) \qquad \text{a.s.}$$

A priori, u_t and $e_t(k)$ are random processes, and depend on the selected subsequence, but we shall prove that they are deterministic functions that do not depend on the subsequence. Thus the convergence holds for the full sequence.

Still considering a suitable subsequence, we use the a.s. convergence stated above and dominated convergence and obtain the infinite system of equations

$$u_t = \lambda - 2\int_0^t \sum_{k=1}^{\infty} k e_s(k)\, ds$$

$$e_t(k) = p_k - \int_0^t e_s(k)\, ds - \int_0^t k e_s(k) \frac{\sum_{j=1}^{\infty} j e_s(j)}{u_s}\, ds, \quad k \geq 0.$$

It follows first that u_t and $e_t(k)$ are continuous, and then that they are continuously differentiable.

The integral equations above thus yield the differential equations

$$\frac{du_t}{dt} = -2\sum_{k=1}^{\infty} k e_t(k)$$

$$\frac{de_t(k)}{dt} = -e_t(k) - k e_t(k)\frac{\sum_{j=1}^{\infty} j e_t(j)}{u_t}, \quad k \geq 0.$$

with the initial conditions $u_0 = \lambda$ and $e_0(k) = p_k$.

The integral equations above thus yield the differential equations

$$\frac{du_t}{dt} = -2\sum_{k=1}^{\infty} k e_t(k)$$

$$\frac{de_t(k)}{dt} = -e_t(k) - k e_t(k)\frac{\sum_{j=1}^{\infty} j e_t(j)}{u_t}, \quad k \geq 0.$$

with the initial conditions $u_0 = \lambda$ and $e_0(k) = p_k$.

Note that the system is infinite, and it is not a priori obvious that it has a solution, or that the solution is unique. The system is not obviously Lipschitz with respect to any of the usual norms on sequence spaces.

The integral equations above thus yield the differential equations

$$\frac{du_t}{dt} = -2\sum_{k=1}^{\infty} k e_t(k)$$

$$\frac{de_t(k)}{dt} = -e_t(k) - k e_t(k)\frac{\sum_{j=1}^{\infty} j e_t(j)}{u_t}, \quad k \geq 0.$$

with the initial conditions $u_0 = \lambda$ and $e_0(k) = p_k$.

Note that the system is infinite, and it is not a priori obvious that it has a solution, or that the solution is unique. The system is not obviously Lipschitz with respect to any of the usual norms on sequence spaces.

Fortunately, it is possible to decouple the system via a change of variables and a time-change.

We make the change of variables

$$h_t(j) = e^t e_t(j),$$

for each j, and rescale time by introducing a new time variable $\tau = \tau_t$ such that

$$\frac{d\tau}{dt} = \frac{\sum_j j e_t(j)}{u_t},$$

with $\tau_0 = 0$.

We make the change of variables

$$h_t(j) = e^t e_t(j),$$

for each j, and rescale time by introducing a new time variable $\tau = \tau_t$ such that

$$\frac{d\tau}{dt} = \frac{\sum_j j e_t(j)}{u_t},$$

with $\tau_0 = 0$.

This yields the system

$$\frac{du_\tau}{d\tau} = \frac{du}{dt}\frac{dt}{d\tau} = -2u_\tau,$$
$$\frac{dh_\tau(j)}{d\tau} = \frac{dh(j)}{dt}\frac{dt}{d\tau} = -jh_\tau(j), \quad j \geq 0,$$

with the initial conditions $u_0 = \lambda$ and $h_0(j) = p_j$.

The system has the obvious unique solution

$$u_\tau = \lambda e^{-2\tau}$$
$$h_\tau(j) = p_j e^{-j\tau}, \quad j \geq 0.$$

This yields

$$\frac{d\tau}{dt} = e^{-t}\frac{\sum_k kp_k e^{-k\tau}}{\lambda e^{-2\tau}},$$

and separating the variables gives

$$1 - e^{-t} = \int_0^t e^{-s}\, ds = \int_0^{\tau_t} \frac{\lambda e^{-2\sigma}}{\sum_k kp_k e^{-k\sigma}}\, d\sigma.$$

This determines τ_t uniquely for every $t \in [0, \infty)$, and thus u_t, $h_t(j)$ and $e_t(j)$ are determined by u_τ and $h_\tau(j)$.
Consequently, the differential equation have a unique solution, $e_t(k)$ and u_t.
This completes the proof that $U_t/n \xrightarrow{\mathrm{p}} u_t$ and $E_t(k)/n \xrightarrow{\mathrm{p}} e_t(k)$ in $D[0, \infty)$.

Furthermore, letting $t \to \infty$,

$$\int_0^{\tau_\infty} \frac{\lambda e^{-2\sigma}}{\sum_k k p_k e^{-k\sigma}} \, d\sigma = 1,$$

as claimed in the theorem. (This determines τ_∞ uniquely.)

Similarly,

$$\frac{S_t}{n} \to s_t := \int_0^t \sum_{k=0}^{\infty} e_s(k)\, ds$$

for any $t < \infty$, and this can be extended to $t = \infty$.
This proves the theorem, with the jamming constant

$$s_\infty = \lim_{t\to\infty} s_t = \int_0^\infty \sum_{k=0}^{\infty} e_t(k)\, dt.$$

The equations above yield

$$\begin{aligned} s_\infty &= \int_0^\infty \sum_k e_t(k)\, dt \\ &= \int_0^{\tau_\infty} \frac{u_\tau \sum_k h_\tau(k)}{\sum_k k h_\tau(k)}\, d\tau \\ &= \lambda \int_0^{\tau_\infty} e^{-2\tau} \frac{\sum_k p_k e^{-k\tau}}{\sum_k k p_k e^{-k\tau}}\, d\tau. \end{aligned}$$