Maximum independent sets of random graphs with given degrees

Matthieu Jonckheere (UBA - Argentina)
and Manuel Sáenz (ICTP - Italy)

Youth in high dimensions - ICTP
July 2020

Maximum independent sets

Definition

An *independent set* is a subset of vertices $I \subseteq V$ s.t., for every $u, v \in I$, $(u, v) \notin E$. It is said to be *maximal* if there is not a larger independent set $I' \subseteq V$ s.t. $I \subseteq I'$. And *maximum* if there is no larger one.

The *independence number* ($\alpha(G)$) is the size of the MIS.

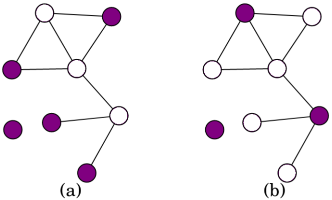

Figure: *(a) In purple, example of a maximum independent set. (b) This independent set is maximal but not maximum, as the previous one is larger.*

(!) Finding MIS or IN are NP-hard tasks[FriezeMcDiarmid'97]

Random graphs

Sparse Erdös-Rényi graphs

A graph $G = (V, E)$ of size n is formed by establishing an edge between each pair of vertices independently and with a fixed probability λ/n. Where $\lambda > 0$ is the mean degree.

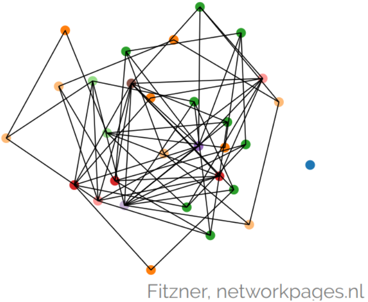

Figure: *Image generated using applet from* `www.networkpages.nl`

Random graphs

Configuration Model

- assign to each vertex $v \in V$ a number d_v of *half-edges*.
- sequentially match uniformly each half-edge to another.
- repeat until finished.

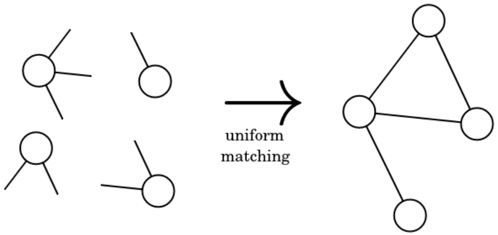

Figure: *Visualisation of the Configuration Model construction.*

Random graphs

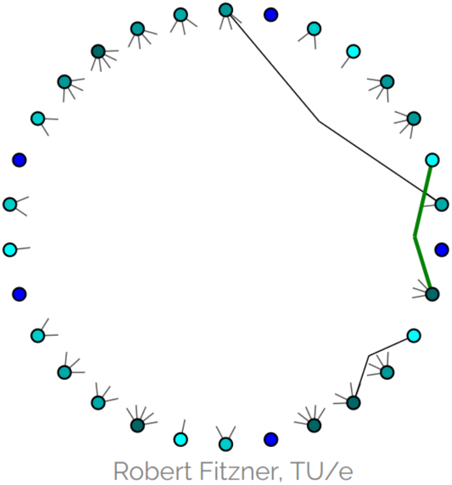

Figure: *Image generated using applet from* `www.networkpages.nl`

Random graphs

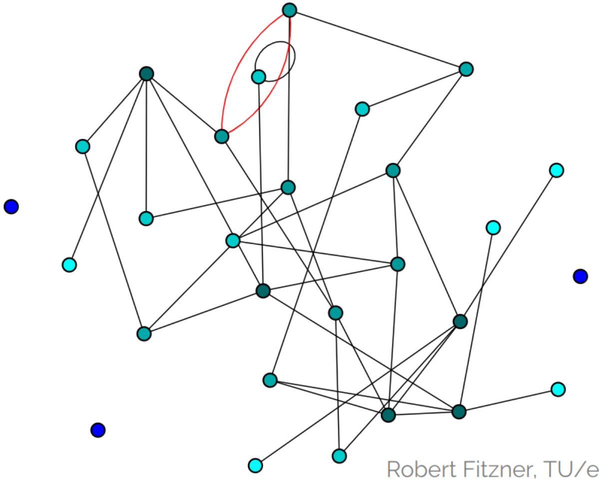

Figure: *Image generated using applet from* `www.networkpages.nl`

Random graphs

Let $D^{(n)}$ be the r.v. that gives the degree of a uniform vertex.

Convergence assumption (CA): we will always assume that, for some r.v. D, $D^{(n)} \xrightarrow{\mathbb{P}} D$ and $\mathbb{E}(D^{(n)2}) \xrightarrow{n\to\infty} \mathbb{E}(D^2) < \infty$.

Theorem (Probability of simplicity[Janson'09])

Under the (CA), the probability of $G \sim CM_n(\bar{d})$ being simple is asymptotically positive.

Then, properties of this model also hold for simple graphs.

Theorem (Giant component[MolloyReed'98])

Let $G \sim CM_n(\bar{d})$ and assume (CA). Then, there is w.h.p. a giant component iff $\nu > 1$.

Where ν is some parameter explicitly defined in terms of the asymptotic degree distribution.

Sequential algorithms

At each time $t \geq 0$ they break the vertex set into three sets: *active* $\mathcal{A}_t$, *blocked* $\mathcal{B}_t$ and *unexplored* $\mathcal{U}_t$ vertices.

In each step, they select a vertex from $\mathcal{U}_t$, declare it active and block their neighbours.

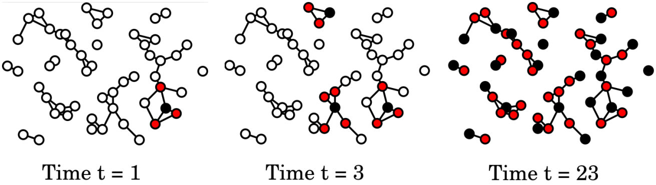

Figure: Example of sequential exploration.

In the *degree-greedy* only minimum degree vertices activate.

(!) At most n steps. Have polynomial time complexity.

Previous results

Very few existing characterisations of MIS of random graphs.

Erdös-Rényi

The work on maximum matchings in [KarpSipser'81][Aaronson et al'98] implies the optimality of DG for ER graphs of mean degree $\lambda < e$.

Configuration Model

Fluid limit for DG [Wormald'99] and asymptotic independence number [Ding et al'16] for regular graphs. Fluid limits for greedy algorithms [Bermolen et al'17] [Brightwell et al'17].

First optimality characterisation

A *selection sequence* is a finite sequence of vertices $(v_i)_{i=1}^k$ s.t. $\{v_i\}_{i=1}^k$ defines a maximal independent set.

Given a selection sequence, the *j-th remaining graph* G_j is the subgraph obtained by removing $v_1, \ldots, v_j$ and their neighbourhoods from G.

Proposition (Sufficient condition for asymptotic optimality)

Let $G \sim CM_n(\bar{d}^{(n)})$. If the asymptotic degree distribution has an exponentially thin tail and the DG defines w.h.p. a selection sequence that selects only vertices of degree 1 or 0 until the remaining graph is subcritical, then (for every $0 < \alpha < 1$) $\sigma_{DG}(G) = \alpha(G) + \mathcal{O}_{\mathbb{P}}(n^{\alpha})$; where $\sigma_{DG}(G)$ is the size of the IS found by the DGA.

(!) Objective: find conditions for this proposition to hold.

Main results

The map $M_1(\cdot)$ gives the distribution obtained after matching the original degree 1 vertices of the graph.

Two characterisations of the near optimality condition based on this:

Theorem (One application of the map)

Assume the degree distribution has mean $\lambda > 0$ and finite second moment. If $\tilde{\nu} := G''_D(Q)/\lambda < 1$ (where $Q := (1 - p_1/\lambda)$ and $G_D(z)$ is the generating function of the degree r.v. D), then, (for every $0 < \alpha < 1$) $\sigma_{DG}(G) = \alpha(G) + \mathcal{O}_{\mathbb{P}}(n^\alpha)$.

(!) Easy to verify but not general.

Main results

General criterion for asymptotic optimality: after a finite number of applications of $M_1(\cdot)$ the resulting distribution is subcritical.

Theorem (Further applications of the map)

Define (for every $i, j \geq 1$) $\eta_j(i) := (-1)^{j-i}\binom{j}{i}\mathbb{I}_{i\leq j}$ and $\tilde{Q} := \sum_{i\geq 2} iQ^i p_i / Q^2\lambda$ and $(a_j)_{j\in\mathbb{N}}$ the components of $(Q^k p_k \mathbb{I}_{\{k\geq 2\}})_{k\in\mathbb{N}}$ in the base $\{\eta_j(\cdot)\}_{j\in\mathbb{N}}$. The remaining graph after one application of the map has distribution

$$M_1^{(n)}\left(p_k^{(n)}\right)(i) \xrightarrow{\mathbb{P}} M_1\left(p_k^{(n)}\right)(i) := \sum_{j\geq i} a_j (-1)^{j-i}\tilde{Q}^j\binom{j}{i}, \text{ for } i \geq 1.$$

(!) By means of this theorem, the general criterion may be verified.

Applications

Proposition (Value of independence numbers)

Let $G \sim CM_n(\bar{d}^{(n)})$. Then, if the general criterion holds, we will have that

$$\alpha(G) = n\left(1 - \sum_{i=1}^{\infty} \frac{\mu^{(i)}(1)(1-Q_i)}{2} + \sum_{j=2}^{\infty}(1-Q_i^j)\mu^{(i)}(j)\right) + o_{\mathbb{P}}(n).$$

Proposition (Upper bound for independence numbers)

Let $G \sim CM_n(\bar{d}^{(n)})$. We can define parameters $\mu^{*(i)}(j)$ and Q_i^* s.t.

$$\alpha(G) \leq n\left(1 - \sum_{i=1}^{\infty} \frac{\mu^{*(i)}(1)(1-Q_i^*)}{2} + \sum_{j=2}^{\infty}(1-Q_i^{*j})\mu^{*(i)}(j)\right) + o_{\mathbb{P}}(n).$$

Applications

Proposition (e-phenomenon)

Let $G \sim ER_n(\lambda)$. If $\lambda < e$, then $\sigma_{DG}(G) = \sigma(G) + o_{\mathbb{P}}(n)$. Furthermore, in this case, $\alpha(G) = n\left(z(\lambda) + \frac{\lambda}{2} z(\lambda)^2\right) + o_{\mathbb{P}}(n)$; where $z(\lambda) := e^{-W(\lambda)}$ with $W(x)$ the Lambert function.

For this, we analyse the sequence of degree distributions (the distribution is always $p_j = A_i \frac{\mu_i^j}{j!} e^{-\mu_i} + B_i \delta_{j1}$)

$$
\begin{cases}
\mu_{i+1} = \left(e^{-\frac{A_i e^{-\mu_i} - A_{i-1} e^{-\mu_{i-1}}}{A_i - A_{i-1} e^{-\mu_{i-1}}} \mu_i} - e^{-\mu_i} \right) \frac{A_i \mu_i}{A_i - A_{i-1} e^{-\mu_{i-1}}}, \\
A_{i+1} = e^{-\frac{A_i e^{-\mu_i} - A_{i-1} e^{-\mu_{i-1}}}{A_i - A_{i-1} e^{-\mu_{i-1}}} \mu_i} A_i.
\end{cases}
$$

The solution is given in terms of the tetration operation. Result follows by convergence of tetration in $(e^{-1}, e^{1/e})$.

Appendix: scaling limits

They are deterministic limits for stochastic processes, analogous to the law of large numbers. Using Prohorov's approach we proved that:

Lemma

Let $(X_t^{(n)}(1), X_t^{(n)}(2), ...)$ *be a sequence of continuous time Markov jump processes. Then, under suitable conditions the processes converge in probability towards the solution of the following system of equations*

$$y_t(1) = y_0(1) + \int_0^t \delta_1(y_s(1), y_s(2), ...)ds$$

$$y_t(2) = y_0(2) + \int_0^t \delta_2(y_s(1), y_s(2), ...)ds$$

$$\dots$$

Here $\delta_i(\cdot)$ is a function called *drift* associated to the process $X_t^{(n)}(i)$ that gives *its expected evolution.*

Appendix: strategy of the proofs

Remember: the map $M_1(\cdot)$ gives the distribution obtained after matching the original degree 1 vertices of the graph.

The matching dynamics is broken into two stages:

- *Phase 1:* vertices of degree 1 are matched but the edges of their neighbours are kept as new vertices.
- *Phase 2:* the vertices corresponding to edges of blocked vertices are matched and the edges so formed removed.

First characterisation: concentration of the final values of unmatched half-edges ($U_t^{(n)}$) and unexplored degree $k \geq 1$ vertices ($\mu_t^{(n)}(k)$) at the end of phase 1 (T_1) + percolation argument.

$$U_{T_1}^{(n)}/n \xrightarrow{\mathbb{P}} Q^2\lambda$$

$$\left(\mu_{T_1}^{(n)}(2), ..., \mu_{T_1}^{(n)}(i), ...\right) \xrightarrow{\mathbb{P}} \left(Q^2 p_2, ..., Q^i p_i, ...\right)$$

Where we define $Q := 1 - p_1/\lambda$.

Appendix: strategy of the proofs

Second characterisation: scaling limit for the dynamics of phase 2.
Number of unmatched half-edges (u_t), blocked half-edges (b_t) and number of degree $k \geq 1$ vertices ($\mu_t(k)$):

$$\begin{cases} u_t = \lambda - \int_0^t 2u_s ds \\ b_t = b_0 - \int_0^t (b_s + u_s) ds \\ \mu_t(k) = Q^k p_k \mathbb{I}_{\{k \geq 2\}} + \int_0^t (k+1)\mu_s(k+1) - k\mu_s(k) ds \text{ (for } k \geq 1) \end{cases}$$

Then, $u_t = \lambda Q^2 e^{-2t}$ and $b_t = Q^2 \lambda e^{-2t} - e^{-t} \sum_{i \geq 2} i Q^i p_i$.
And the equations for the degree measure are decoupled if rewritten in the base

$$\eta_k(i) = (-1)^{k-i} \binom{k}{i} \mathbb{I}_{\{i \leq k\}}$$