Universitat
Pompeu Fabra
Barcelona

Seminar 5 · Networks, Crowds and Markets

Random graphs and communities

Configuration model and preferential attachment

In the first part we will experiment a bit with the two models.

Sampling the configuration model

We have N nodes and degree sequence $(k_1, \dots, k_N)$. Let each node i own k_i *stubs* (half–edges).

Sampling algorithm (uniform over all pairings):

1. Make a list of all $2L = \sum_i k_i$ stubs. Label each stub by its node owner.
2. While unpaired stubs remain:
   - 2.1 Select a stub uniformly at random.
   - 2.2 Choose its partner uniformly at random among the remaining stubs.
   - 2.3 Connect their owners with an edge (i, j).
   - 2.4 Remove both stubs.
   - 2.5 Repeat.
3. The resulting multigraph is one sample from the configuration model.

Note: loops (i, i) and parallel edges may appear. They are rare for large sparse networks.

Exercise: Correctness of the algorithm

Show that the algorithm described above generates multigraph with the right degree sequence and uniformly distributed. What is the probability of a particular pairing?

Exercise: Degree distribution in PA

Consider preferential attachment with $m = 1$.

Let d_t be the degree of the initial vertex in time t.

What is the distribution of d_5?

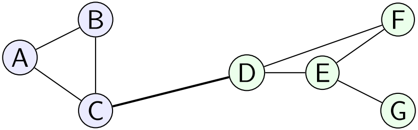
What are $\mathbb{P}(d_t = 1)$ and $\mathbb{P}(d_t = t - 1)$?

Find the exact formula for $\mathbb{E}[d_{t+1} - d_t | d_t = k]$.

Show that $\mathbb{E}[d_t] = \frac{(2t-3)!!}{2^{t-2}(t-2)!} = \frac{(2t-3)!!}{(2t-4)!!}$, where $n!! = n(n-2)(n-4)\cdots$.

Exercise: Detecting communities

Graph: two dense parts joined by a single bridge.



(a) Check if $\{A, B, C\}$ and $\{D, E, F, G\}$ form *strong* or *weak* communities.

(b) What minimal edges would make $\{D, E, F, G\}$ strong?

(c) Compute the *topological overlap similarity* and perform one step of average linkage: which clusters merge first?

(d) Compare the modularity for two partitions:

$$\mathcal{P}_1 = \{\{A, B, C\}, \{D, E, F, G\}\}, \quad \mathcal{P}_2 = \{\{A, B, C\}, \{D, E, F\}, \{G\}\}.$$

For simplicity, on the next slide we give the similarity matrix.

Similarity matrix for the toy network (corrected)

$B_A = \{A, B, C\}$, $B_B = \{A, B, C\}$, $B_C = \{A, B, C, D\}$,
$B_D = \{D, E, F, C\}$, $B_E = \{E, D, F, G\}$, $B_F = \{F, D, E\}$, $B_G = \{G, E\}$.

	A	B	C	D	E	F	G
A	1.00	1.00	1.00	0.33	0	0	0
B	1.00	1.00	1.00	0.33	0	0	0
C	1.00	1.00	1.00	0.50	0.25	0.33	0
D	0.33	0.33	0.50	1.00	0.75	1.00	0.50
E	0	0	0.25	0.75	1.00	1.00	1.00
F	0	0	0.33	1.00	1.00	1.00	0.50
G	0	0	0	0.50	1.00	0.50	1.00

Note: bridge $C - D$ yields $s_{CD} = 0.5$; triangle $D - E - F$ gives $s_{DF} = s_{EF} = 1$; leaf G on E gives $s_{EG} = 1$ and $s_{DG} = s_{FG} = 0.5$.