

## 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.

See the corresponding colab.

# 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

This is a reformulated version of this result.

Show that the algorithm described above generates **multigraph** with the right degree sequence. Show that the probability of each pairing is uniform.

## Solution: Why the pairing is uniform

**Pairing** of a set of  $2L$  elements: a collection of 2-elements subsets that are disjoint and whose union gives the whole set.

Let  $2L = \sum_i k_i$  be the number of stubs. There are  $(2L - 1)!! = (2L - 1)(2L - 3) \cdots 1$  pairings between these stubs.

### Proof of uniformity.

Record the sampled stubs as an ordered sequence  $(s_1, s_2; s_3, s_4; \dots; s_{2L-1}, s_{2L})$ . Any fixed ordered sequence has probability

$$\frac{1}{2L} \cdot \frac{1}{2L-1} \cdot \frac{1}{2L-2} \cdots \frac{1}{2} \cdot \frac{1}{1} = \frac{1}{(2L)!}.$$

A *single* unordered pairing can be realized by  $L!$  orders of the  $L$  pairs and 2 orders inside each pair:  $2^L L!$  sequences total. Hence

$$\Pr(\text{a given pairing}) = \frac{2^L L!}{(2L)!} = \frac{1}{(2L - 1)!!},$$

the same for every pairing.

## Pairings do not define the multigraph uniquely

As pointed out by Calixta in the first seminar group: pairings do not uniquely define multigraphs. For example, if we have two degree two vertices  $A, B$  with stubs  $A_1, A_2, B_1, B_2$  then pairings  $\{A_1, B_1\}, \{A_2, B_2\}$  and  $\{A_1, B_2\}, \{A_2, B_1\}$  both encode the double edge between  $A$  and  $B$ .

The distribution over multigraphs is in general not uniform.

Depending on a multigraph, there may be a different number of pairings leading to it. On the next slide we discuss an example.

## Counterexample

Take four nodes  $A, B, C, D$  with degrees  $(1, 2, 2, 1)$ . The stubs are  $A_1, B_1, B_2, C_1, C_2, D_1$  with 15 pairings.

The simple graph  $A - B - C - D$  corresponds to **four** pairings of the form  $\{A_1, B_i\}, \{B_{3-i}, C_j\}, \{C_{3-j}, D_1\}$  for  $i, j = 1, 2$ .

The simple graph  $A - C - B - D$  also corresponds to **four** pairings

Double edge between  $B, C$  and an edge between  $A, D$  is defined by **two** pairings  $\{B_i, C_j\}, \{B_{3-i}, C_{3-j}\}, \{A_1, D_1\}$  with  $i = j$  or  $i \neq j$ .

The graph with  $B, C$  having loops and an edge between  $A, D$  is given by a **single** pairing.

The graphs with  $B$  has a loop  $A - C - D$  is given by **two** pairings.

The graphs with  $C$  has a loop  $A - B - D$  is given by **two** pairings.

The probability of each multigraph =  $\frac{\# \text{pairings}}{15}$

## Exercise: Degree distribution in preferential attachment

Consider the **preferential attachment model** with  $m = 1$ .

Let  $d_t$  denote the degree of the initial vertex at time  $t$ .

- (a) What is the distribution of  $d_5$ ?
- (b) What are  $\Pr(d_t = 1)$  and  $\Pr(d_t = t - 1)$ ?
- (c) Find the exact expression for

$$\mathbb{E}[d_{t+1} - d_t \mid d_t = k].$$

- (d) Show that

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

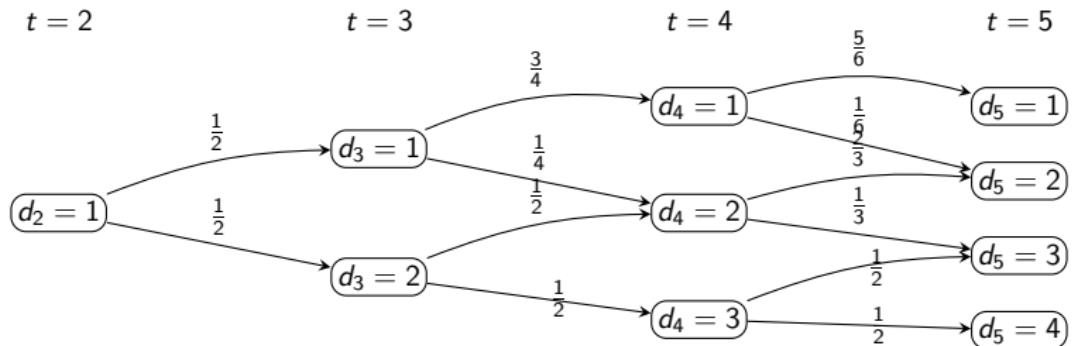
**Hint:** express the degree growth as a random multiplicative process and use the recurrence relation for  $\mathbb{E}[d_t]$ .

## Solution sketch (a) — corrected transitions

**Setup.** At time  $t$ , total degree is  $2(t - 1)$ . If  $d_t = k$ , then

$$\Pr(d_{t+1} = k+1 \mid d_t = k) = \frac{k}{2(t-1)}, \quad \Pr(d_{t+1} = k \mid d_t = k) = 1 - \frac{k}{2(t-1)}.$$

**Transition diagram (from  $t = 2$  to  $t = 5$ ).**



**Distribution at  $t = 5$ :**

$$\Pr(d_5 = 1) = \frac{5}{16}, \quad \Pr(d_5 = 2) = \frac{5}{16}, \quad \Pr(d_5 = 3) = \frac{1}{4}, \quad \Pr(d_5 = 4) = \frac{1}{8}.$$

**Endpoints (all  $t \geq 2$ ):**

$$\Pr(d_t = 1) = \prod_{s=2}^{t-1} \left(1 - \frac{1}{2(s-1)}\right) = \frac{(2t-4)!!}{(2t-3)!!},$$

$$\Pr(d_t = t-1) = \frac{(t-2)!}{(2t-4)!!} = \frac{1}{2^{t-2}}.$$

**Drift and mean.**

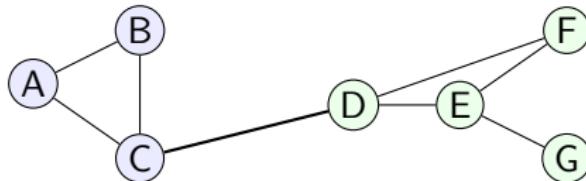
$\mathbb{E}[d_{t+1} - d_t \mid d_t = k] = \frac{k}{2(t-1)} \Rightarrow \mathbb{E}[d_{t+1}] = \left(1 + \frac{1}{2(t-1)}\right) \mathbb{E}[d_t]$ , with  $d_2 = 1$ , hence

$$\mathbb{E}[d_t] = \prod_{s=2}^{t-1} \left(1 + \frac{1}{2(s-1)}\right) = \frac{(2t-3)!!}{2^{t-2}(t-2)!} = \frac{(2t-3)!!}{(2t-4)!!}.$$

Check at  $t = 5$ :  $\mathbb{E}[d_5] = \frac{7!!}{6!!} = \frac{105}{48} = 35/16 \approx 2.1875$  (matches the diagram).

## 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) Compute the *topological overlap similarity* and perform one step of average linkage: which clusters merge first?
- (c) 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  $s_{DG} = s_{FG} = 0.5$ .