X. Preferential attachment.

The preferential attachment effect can be observed ~~globally~~ ~~over~~ directly over the global network, or ~~locally~~ indirectly on ~~each~~ latent classes.

~~Previous~~ ~~Work~~

X.1. Global preferential attachment

~~The models~~

Stochastic block-models lead to the ~~gen~~ following generative process for links*:

- For each node i, $1 \leq i \leq N$:
  - For each node j, $1 \leq j \leq N$; generate a link between i and j with probability $P(y_{ij} = 1 \mid \mathcal{M})$

where $\mathcal{M}$ is either $\mathcal{M}_c$ or $\mathcal{M}_g$.

This ~~above~~ process, for any given node i, considers all nodes in turn, from node 1 to node N. The indexing, i.e. the mapping between nodes and integers in $[1..N]$, is however arbitrary and the results that follow are stated for ~~each~~ possible indexings.

For a given node i, at step p of the above process, p nodes, from node 1 to node p, have been considered and links from these nodes to node i generated or not. We will denote by $d_i^{(p)}$ the degree of node i, i.e. the number of links of node i, at the p^{th} step of this process. By definition:

$$d_i^{(p)} = \sum_{j=1}^{p} y_{ij}$$

* For simplicity in the notation, we consider that nodes can be linked to themselves. Excluding such links does not raise particular problems.

The following definition states how a model $\mathcal{M}$ (as defined above) behaves wrt global preferential attachment.

Definition (global preferential attachment): Let $\mathcal{M}$ be a model as defined above. We say that $\mathcal{M}$ satisfies the global preferential attachment effect iff, for any indexing, for any node i, $1 \le i \le N$, for any p, $1 \le p < N$,

$p(d_i^{(N)} \geq n+1 \mid d_i^{(p)} = n; \mathcal{M})$ increases with n, $1 \le n < p$.

If $P(d_i^{(N)} \geq n+1 \mid d_i^{(p)} = n; \mathcal{M})$ is independent of n, the model is said to be neutral wrt the global preferential attachment effect.

[INCLUDE TEXT ON

This definition directly translates the fact that the more links a node i has at some point in the process, the more likely a new link will be created with that node. This is exactly the preferential attachment effect.

For both ILFM and IMMSB, in model $\mathcal{M}_e$, the generation of links are independent of each other. The fact that n links have been created after p steps has no impact on the future links to a given node. In $\mathcal{M}_g$, as one first needs to generate F and Φ prior to generate all the links, a similar behavior is likely to be observed. Intuitively thus, both ILFM and IMMSB are neutral wrt the global preferential attachment effect. The following property formalizes this intuition.

Property 1: Both SEFH and EMRSB, for both $\mathcal{M}_e$ and $\mathcal{M}_g$, are neutral wrt the global preferential attachment effect.

Proof: We first consider model $\mathcal{M}_e$. Fix any indexing, a node i, $1 \le i \le N$, and a step p, $1 \le p < N$. One has, $\forall n$, $1 \le n < p$:

$$P(d_i^{(N)} \geq n+1 \mid d_i^{(p)} = n ; \mathcal{M}_e) = 1 - P(d_i^{(N)} = n \mid d_i^{(p)} = n ; \mathcal{M}_e)$$

$$= 1 - P(y_{i,p+1} = 0, \dots, y_{iN} = 0 \mid \mathcal{M}_e)$$

$$= 1 - \prod_{j=p+1}^{N} P(y_{ij} = 0 \mid \mathcal{M}_e)$$

where the last equality comes from the fact that, in $\mathcal{M}_e$, links are independently generated.

Similarly:

$$P(d_i^{(N)} \geq n+2 \mid d_i^{(p)} = n+1 ; \mathcal{M}_e) = 1 - \prod_{j=p+1}^{N} P(y_{ij} = 0 \mid \mathcal{M}_e) = P(d_i^{(N)} \geq n+1 \mid d_i^{(p)} = n ; \mathcal{M}_e)$$

This completes the proof for $\mathcal{M}_e$.

For $\mathcal{M}_g$, it suffices to observe that the above result is true for any F, Φ and that:

$$P(d_i^{(N)} \geq n+1 \mid d_i^{(p)} = n ; \mathcal{M}_g) = \iint_{F,\Phi} P(F \mid \mathcal{M}_g)\, P(\Phi \mid \mathcal{M}_g)\, P(d_i^{(N)} \geq n+1 \mid d_i^{(p)} = n ; F, \Phi)\, dF\, d\Phi$$

As $P(d_i^{(N)} \geq n+1 \mid d_i^{(p)} = n ; F, \Phi) = P(d_i^{(N)} \geq n+2 \mid d_i^{(p)} = n+1 ; F, \Phi)$, one has:

$$P(d_i^{(N)} \geq n+2 \mid d_i^{(p)} = n+1 ; \mathcal{M}_g) = P(d_i^{(N)} \geq n+1 \mid d_i^{(p)} = n ; \mathcal{M}_g). \qquad \square$$

We now turn to local preferential attachment, as ~~studied~~ exemplified in eg → [REF TO LESKOVEC SHOULD BE ADDED].

X.2. Local preferential attachment.

X.2.1 ILFM

For ILFM, the situation wrt to local preferential attachment is very similar to the one for global preferential attachment. This is due to the fact that, ~~as Test~~ in M_e (i.e given F and ϕ), a local degree can be defined in the same way as the global degree above.

Considering the same generative process as before, for M_e wd ILFM, the local degree in class k, $1 \leq k \leq K$, for a node i such that $f_{ik} = 1$ is defined by:

$$d_{i,k}^{(p)} = \sum_{j=1,\, f_{jk}=1}^{p} y_{ij}$$

Note that if $f_{ik} = 0$, $d_{i,k}^{(p)} = 0$ for all p.

This then leads to the following definition of the local preferential attachment for ILFM.

Definition (ILFM - local preferential attachment): ~~Let M_e be a model (M_g or M_l) associated to ILFM.~~ We say that M_e ILFM in ~~...~~ satisfies the local preferential attachment effect iff for any indexing, for any node i $(1 \leq i \leq N)^{*}$, for any step p $(1 \leq p < N)$:

$$P(d_{i,k}^{(N)} \geq n+1 \mid d_{i,k}^{(p)} = n ; M_e) \text{ increases with } n, \ 1 \leq n < p.$$

$*$ such that $f_{ik} = 1$

If $P(d_{i,k}^{(N)} \geq n+1 \mid d_i^{(p)} = n ; M_e)$ is independent of n, ~~we say that~~ the model is said to be ~~...~~ neutral wrt to the local preferential attachment effect.

As before, we have the following property.

<u>Property 2.</u> ILFM, for $\mathcal{M}_e$, is neutral wrt to the local preferential attachment effect.

Proof: The proof is identical to the first part of the proof for Property 1.

The definition of the local degree we have used does not directly translate to $\mathcal{M}_g$ as it is based on the knowledge of ~~F~~. One can nevertheless define an expected local degree for $\mathcal{M}_g$ ~~as follows~~ as:

$$gd_{i,\ell}^{(p)} = \mathbb{E}_{F,\Phi}\left[d_{i,\ell}^{(p)}\right]$$

$$= \iint_{F,\Phi} P(F|\mathcal{M}_g)\, P(\Phi|\mathcal{M}_g)\, d_{i,\ell}^{(p)}\, dF\, d\Phi$$

A notion of local preferential attachment for $\mathcal{M}_g$ can then be defined on the basis of this expected local degree.

<u>Definition</u> (ILFM - local preferential attachment; $\mathcal{M}_g$) We say that ILFM in $\mathcal{M}_g$ satisfies the local preferential attachment effect iff for any indexing, for any node i $(1 \leq i \leq N)$, for any step p $(1 \leq p < N)$; $\exists\, \varepsilon \in]0,1[$ s.t. $P(gd_{i,\ell}^{(N)} \geq x + \varepsilon \mid gd_{i,\ell}^{(p)} \geq x; \mathcal{M}_g)$ increases with x, $x \in [0, N[$.

[INCLUDE A ~~NOTE~~ § ON LINK TO CONTINUOUS BURSTINESS]

~~As before,~~

~~By definition,~~ One has ~~$P(gd_{i,\ell}^{(N)} \geq x + \varepsilon$~~ $\forall\, \varepsilon \in]0,1[,\ x \in]0, N[$:

$$P(gd_{i,\ell}^{(N)} \geq x + \varepsilon \mid gd_{i,\ell}^{(p)} \geq x; \mathcal{M}_g) = 1 - P(gd_{i,\ell}^{(N)} = x \mid gd_{i,\ell}^{(p)} = x; \mathcal{M}_g)$$

$$= 1 - P(y_{i,p+1} = 0, \ldots, y_{i,N} = 0 \mid \mathcal{M}_g)$$

where the last equation derives from the definition of $gd_{i\ell}^{(p)}$ ($gd_{i\ell}^{(N)} = gd_{i\ell}^{(p)}$ iff $y_{ip+1} = \dots = y_{iN} = 0$). (6)

The last quantity is independent of x, leading to:

<u>Property 3.</u> ICFM with Π_g is neutral wrt to the local preferential attachment effect.

X.2.2. IMMSB

For IMMSB in M_e, ~~we~~ ~~do~~ the situation is similar to the one of ICFM in M_g, as we do not have ~~access~~ a direct access to classes, encoded in the z variables.

One can nevertheless define, ~~through expected~~ a local random variables $y_{ij,\ell}$ that are 1 if a link is generated between nodes i and j using class ℓ and 0 otherwise. One has:

$$P(y_{ij,\ell} = 1 \mid M_e) = P(y_{ij} = 1 \mid z_{i\to j} = \ell, z_{i\leftarrow j} = \ell, \Phi) \, P(z_{i\to j} = \ell \mid F) \, P(z_{j\leftarrow i} = \ell \mid F)$$

$$= f_{i\ell} \, \Phi_{\ell\ell} \, f_{j\ell}$$

The local degree $d_{i,\ell}^{(p)}$ can then be defined as the expectation over the nodes $1, \dots, p$ of $y_{ij,\ell}$:

$$d_{i\ell}^{(p)} = \sum_{j=1}^{p} P(y_{ij,\ell} = 1 \mid M_e) = \sum_{j=1}^{p} f_{i\ell} \, \Phi_{\ell\ell} \, f_{j\ell}$$

→ THEN DEFINITION OF LOCAL PREF. ATT., PROPERTY AND PROOF AS IN ~~ECML~~ DSAA

→ ~~ALL~~ DEVELOPMENT FOR Π_g TO BE DONE (SOON)