Supplementary Material for Properties of Latent Variable Network Models

RICCARDO RASTELLI, NIAL FRIEL
School of Mathematics and Statistics, University College Dublin, Dublin, Ireland
Insight: Centre for Data Analytics, Ireland
and ADRIAN E. RAFTERY
Department of Statistics, University of Washington, Seattle, USA
(*e-mail:* riccardo.rastelli@ucdconnect.ie; nial.friel@ucd.ie; raftery@u.washington.edu)

A Appendix: proofs

A.1 Theorem 1

Q1. This is straightforward since $\forall \mathbf{z}_s \in \mathbb{R}^d$:

$$\theta(\mathbf{z}_s) = Pr(y_{sj} = 1 | \mathbf{z}_s) = \int_{\mathscr{Z}} p(\mathbf{z}_j)\, r(\mathbf{z}_s, \mathbf{z}_j)\, d\mathbf{z}_j. \tag{A 1}$$

Q2.

$$\begin{aligned}
G(x) &= \sum_{k=0}^{n-1} x^k p_k = \sum_{k=0}^{n-1} x^k Pr(D_s = k) \\
&= \sum_{k=0}^{n-1} x^k \int_{\mathscr{Z}} \cdots \int_{\mathscr{Z}} p(\mathbf{z}_1) \cdots p(\mathbf{z}_n)\, Pr(D_s = k | \mathscr{M})\, d\mathbf{z}_1 \cdots d\mathbf{z}_n \\
&= \int_{\mathscr{Z}} \cdots \int_{\mathscr{Z}} \left[\prod_{j=1}^{n} p(\mathbf{z}_j) \right] \mathbb{E}\left[x^{D_s} \middle| \mathscr{M} \right] d\mathbf{z}_1 \cdots d\mathbf{z}_n \\
&= \int_{\mathscr{Z}} \cdots \int_{\mathscr{Z}} \left[\prod_{j=1}^{n} p(\mathbf{z}_j) \right] \left\{ \prod_{j=1}^{n} \mathbb{E}\left[x^{Y_{sj}} \middle| \mathscr{M} \right] \right\} d\mathbf{z}_1 \cdots d\mathbf{z}_n \\
&= \int_{\mathscr{Z}} \cdots \int_{\mathscr{Z}} \left\{ \prod_{j=1}^{n} p(\mathbf{z}_j) \left[x r(\mathbf{z}_s, \mathbf{z}_j) + 1 - r(\mathbf{z}_s, \mathbf{z}_j) \right] \right\} d\mathbf{z}_1 \cdots d\mathbf{z}_n \\
&= \int_{\mathscr{Z}} p(\mathbf{z}_s) \left\{ \int_{\mathscr{Z}} p(\mathbf{z}_j) \left[x r(\mathbf{z}_s, \mathbf{z}_j) + 1 - r(\mathbf{z}_s, \mathbf{z}_j) \right] d\mathbf{z}_j \right\}^{n-1} d\mathbf{z}_s \\
&= \int_{\mathscr{Z}} p(\mathbf{z}_s) \left\{ x \int_{\mathscr{Z}} p(\mathbf{z}_j)\, r(\mathbf{z}_s, \mathbf{z}_j)\, d\mathbf{z}_j + 1 - \int_{\mathscr{Z}} p(\mathbf{z}_j)\, r(\mathbf{z}_s, \mathbf{z}_j)\, d\mathbf{z}_j \right\}^{n-1} d\mathbf{z}_s \\
&= \int_{\mathscr{Z}} p(\mathbf{z}_s) \left[x\theta(\mathbf{z}_s) + 1 - \theta(\mathbf{z}_s) \right]^{n-1} d\mathbf{z}_s.
\end{aligned} \tag{A 2}$$

Q3. The r-th factorial moment of D_s corresponds to the r-th derivative of G evaluated in 1:

$$\begin{aligned}\frac{\partial^r G}{\partial x^r}(x) &= \int_{\mathscr{Z}} p(\mathbf{z}_s) \frac{\partial^r}{\partial x^r} \left[x\theta(\mathbf{z}_s) + 1 - \theta(\mathbf{z}_s)\right]^{n-1} d\mathbf{z}_s \\ &= \int_{\mathscr{Z}} p(\mathbf{z}_s)(n-1)\cdots(n-r)\,\theta(\mathbf{z}_s)^r \left[x\theta(\mathbf{z}_s) + 1 - \theta(\mathbf{z}_s)\right]^{n-r-1} d\mathbf{z}_s \\ &= \frac{(n-1)!}{(n-r-1)!} \int_{\mathscr{Z}} p(\mathbf{z}_s)\,\theta(\mathbf{z}_s)^r \left[x\theta(\mathbf{z}_s) + 1 - \theta(\mathbf{z}_s)\right]^{n-r-1} d\mathbf{z}_s;\end{aligned} \tag{A 3}$$

and the final formula evaluated in $x = 1$ gives (11).

Q4. The average degree is the first factorial moment, thus:

$$\bar{k} = G'(1) = \frac{(n-1)!}{(n-2)!} \int_{\mathscr{Z}} p(\mathbf{z}_s)\,\theta(\mathbf{z}_s)\,d\mathbf{z}_s = (n-1)\int_{\mathscr{Z}} p(\mathbf{z}_s)\,\theta(\mathbf{z}_s)\,d\mathbf{z}_s. \tag{A 4}$$

Q5. The distribution of the degree of a random node can be recovered by differentiating G as well. Indeed, using (A 3), for every k:

$$p_k = \frac{1}{k!}\frac{\partial^r G}{\partial x^r}(0) = \binom{n-1}{k} \int_{\mathscr{Z}} p(\mathbf{z}_s)\,\theta(\mathbf{z}_s)\left[1 - \theta(\mathbf{z}_s)\right]^{n-k-1} d\mathbf{z}_s. \tag{A 5}$$

Q6. Define the PGF for the degree of a random node once its latent information is fixed to $\mathbf{z}_s$:

$$\begin{aligned}\tilde{G}(x;\mathbf{z}_s) &= \sum_{k=0}^{n-1} x^k Pr(D_s = k|\mathbf{z}_s) \\ &= \int_{\mathscr{Z}} \cdots \int_{\mathscr{Z}} \left[\prod_{\substack{j=1 \\ j\neq s}}^{n} p(\mathbf{z}_j)\right] \mathbb{E}\left[x^{D_s}\middle|\mathscr{M}\right] d\mathbf{z}_{-s} \\ &= \left\{\int_{\mathscr{Z}} p(\mathbf{z}_j)\left[xr(\mathbf{z}_s,\mathbf{z}_j) + 1 - r(\mathbf{z}_s,\mathbf{z}_j)\right] d\mathbf{z}_j\right\}^{n-1} \\ &= \left\{x\theta(\mathbf{z}_s) + 1 - \theta(\mathbf{z}_s)\right\}^{n-1};\end{aligned} \tag{A 6}$$

which is simply the PGF of a binomial random variable with parameters $n-1$ and $\theta(\mathbf{z}_s)$. Hence its average degree is $\bar{k}(\mathbf{z}_s) = (n-1)\,\theta(\mathbf{z}_s)$. Note that $d\mathbf{z}_{-s} = \prod_{j\neq s} d\mathbf{z}_j$.

Q7. We now write down the PGF for the degree of a random neighbour of a node located in $\mathbf{z}_s$.

$$\begin{aligned}H(x;\mathbf{z}_s) &= \sum_{k=0}^{n-1} x^k Pr(D_j = k\middle|y_{sj} = 1, \mathbf{z}_s) \\ &= \int_{\mathscr{Z}} p(\mathbf{z}_j\middle|y_{sj} = 1, \mathbf{z}_s) \sum_{k=0}^{n-1} x^k Pr(D_j = k\middle|y_{sj} = 1, \mathbf{z}_s, \mathbf{z}_j)\,d\mathbf{z}_j \\ &= \int_{\mathscr{Z}} p(\mathbf{z}_j\middle|y_{sj} = 1, \mathbf{z}_s)\,\mathbb{E}\left[x^{D_j}\middle|y_{sj} = 1, \mathbf{z}_s, \mathbf{z}_j\right] d\mathbf{z}_j.\end{aligned} \tag{A 7}$$

Note that $\mathbb{E}\left[x^{D_j}\middle|y_{sj}=1,\mathbf{z}_s,\mathbf{z}_j\right]$ corresponds to the PGF for the so called excess degree (Newman et al. 2001), i.e. the degree of a node at one extreme of an edge picked at random. Hence, such PGF is equal to $\frac{x\tilde{G}'(x;\mathbf{z}_j)}{\tilde{G}(1;\mathbf{z}_j)}$, where $\tilde{G}$ has been defined in (A 6). Then:

$$\begin{aligned} H(x;\mathbf{z}_s) &= \int_{\mathscr{Z}} p\left(\mathbf{z}_j\middle|y_{sj}=1,\mathbf{z}_s\right)\frac{x\tilde{G}'(x;\mathbf{z}_j)}{\tilde{G}(1;\mathbf{z}_j)}d\mathbf{z}_j \\ &= \int_{\mathscr{Z}} \frac{Pr\left(y_{sj}=1\middle|\mathbf{z}_j,\mathbf{z}_s\right)p(\mathbf{z}_j)}{Pr\left(y_{sj}=1\middle|\mathbf{z}_s\right)}\left\{x\left[x\theta\left(\mathbf{z}_j+1-\theta\left(\mathbf{z}_j\right)\right)\right]^{n-2}\right\}d\mathbf{z}_j \\ &= \frac{1}{\theta(\mathbf{z}_s)}\int_{\mathscr{Z}} p(\mathbf{z}_j)\,r(\mathbf{z}_j,\mathbf{z}_s)\left\{x\left[x\theta\left(\mathbf{z}_j+1-\theta\left(\mathbf{z}_j\right)\right)\right]^{n-2}\right\}d\mathbf{z}_j. \end{aligned} \tag{A 8}$$

Its average degree is then given by:

$$\begin{aligned} \bar{k}_{nn}(\mathbf{z}_s) = H'(1;\mathbf{z}_s) &= \frac{1}{\theta(\mathbf{z}_s)}\int_{\mathscr{Z}} p(\mathbf{z}_j)\,r(\mathbf{z}_j,\mathbf{z}_s)\left\{1+(n-2)\,\theta(\mathbf{z}_j)\right\}d\mathbf{z}_j \\ &= 1+\frac{(n-2)}{\theta(\mathbf{z}_s)}\int_{\mathscr{Z}} p(\mathbf{z}_j)\,r(\mathbf{z}_j,\mathbf{z}_s)\,\theta(\mathbf{z}_j)\,d\mathbf{z}_j. \end{aligned} \tag{A 9}$$

Q8. The PGF for the degree of a neighbour of a node with degree k is given by:

$$\begin{aligned} \tilde{H}(x;k) &= \sum_{r=0}^{n-1} x^r Pr\left(D_j=r\middle|D_s=k,y_{sj}=1\right) \\ &= \sum_{r=0}^{n-1} x^r \int_{\mathscr{Z}} p(\mathbf{z}_s|D_s=k)\,Pr\left(D_j=r\middle|\mathbf{z}_s,y_{sj}=1\right)d\mathbf{z}_s \\ &= \frac{1}{p_k}\int_{\mathscr{Z}} p(\mathbf{z}_s)\,Pr(D_s=k|\mathbf{z}_s)\,H(x;\mathbf{z}_s)\,d\mathbf{z}_s \\ &= \frac{1}{p_k}\int_{\mathscr{Z}} p(\mathbf{z}_s)\left[\frac{\partial^k}{\partial x^k}\tilde{G}(0;\mathbf{z}_s)\right]H(x;\mathbf{z}_s)\,d\mathbf{z}_s \\ &= \frac{1}{p_k}\int_{\mathscr{Z}} p(\mathbf{z}_s)\binom{n-1}{k}\theta(\mathbf{z}_s)^k\left[1-\theta(\mathbf{z}_s)\right]^{n-k-1}H(x;\mathbf{z}_s)\,d\mathbf{z}_s; \end{aligned} \tag{A 10}$$

and its first derivative evaluated in $x=1$ yields:

$$\bar{k}_{nn}(k) = \frac{1}{p_k}\int_{\mathscr{Z}} p(\mathbf{z}_s)\binom{n-1}{k}\theta(\mathbf{z}_s)^k\left[1-\theta(\mathbf{z}_s)\right]^{n-k-1}\bar{k}_{nn}(\mathbf{z}_s)d\mathbf{z}_s. \tag{A 11}$$

A.1.1 Proof for Corollary 1

Recall that a convolution of two Gaussian densities is still a Gaussian density:

$$\int_{\mathbb{R}^d} f_d(\mathbf{z}_i;\boldsymbol{\mu}_1,\gamma_1)\,f_d(\mathbf{z}_j-\mathbf{z}_i;\boldsymbol{\mu}_2,\gamma_2)\,d\mathbf{z}_i = f_d(\mathbf{z}_j;\boldsymbol{\mu}_1+\boldsymbol{\mu}_2,\gamma_1+\gamma_2), \tag{A 12}$$

for every $\mathbf{z}_i,\mathbf{z}_j,\boldsymbol{\mu}_1,\boldsymbol{\mu}_2$ in $\mathbb{R}^d$ and every positive real numbers γ_1 and γ_2.

That being said:

Q1.

$$\begin{aligned}\theta(\mathbf{z}_s) &= \int_{\mathbb{R}^d} f_d(\mathbf{z}_j;\mathbf{0},\gamma)\,\tau\,(2\pi\varphi)^{\frac{d}{2}}\, f_d(\mathbf{z}_s-\mathbf{z}_j;\mathbf{0},\varphi)\,d\mathbf{z}_j \\ &= \tau\,(2\pi\varphi)^{\frac{d}{2}}\, f_d(\mathbf{z}_s;\mathbf{0},\gamma+\varphi) \\ &= \tau\left(\frac{\varphi}{\gamma+\varphi}\right)^{\frac{d}{2}} \exp\left\{-\frac{1}{2(\gamma+\varphi)}\mathbf{z}_s^t\mathbf{z}_s\right\}.\end{aligned} \tag{A 13}$$

Q3.

$$\begin{aligned}\frac{\partial^r G}{\partial x^r}(1) &= \frac{(n-1)!}{(n-r-1)!}\int_{\mathbb{R}^d} f_d(\mathbf{z}_s;\mathbf{0},\gamma)\,\theta(\mathbf{z}_s)^r d\mathbf{z}_s \\ &= \frac{(n-1)!}{(n-r-1)!}\tau^r\left(\frac{\varphi}{\gamma+\varphi}\right)^{\frac{rd}{2}}\int_{\mathbb{R}^d} f_d(\mathbf{z}_s;\mathbf{0},\gamma)\exp\left\{-\frac{r}{2\,(\gamma+\varphi)}\mathbf{z}_s^t\mathbf{z}_s\right\}d\mathbf{z}_s \\ &= \frac{(n-1)!}{(n-r-1)!}\tau^r\left(\frac{\varphi}{\gamma+\varphi}\right)^{\frac{rd}{2}}\left\{2\pi\frac{(\gamma+\varphi)}{r}\right\}^{\frac{d}{2}}\times \\ &\qquad \times\int_{\mathbb{R}^d} f_d(\mathbf{z}_s;\mathbf{0},\gamma)\, f_d\left(\mathbf{z}_s;\mathbf{0},\frac{\gamma+\varphi}{r}\right)d\mathbf{z}_s \\ &= \frac{(n-1)!}{(n-r-1)!}\tau^r\left(\frac{\varphi}{\gamma+\varphi}\right)^{\frac{rd}{2}}\left\{2\pi\frac{(\gamma+\varphi)}{r}\right\}^{\frac{d}{2}}\left\{2\pi\frac{[(r+1)\,\gamma+\varphi]}{r}\right\}^{-\frac{d}{2}} \\ &= \frac{(n-1)!}{(n-r-1)!}\tau^r\left\{\frac{\varphi^r}{(\gamma+\varphi)^{r-1}\,[(r+1)\gamma+\varphi]}\right\}^{\frac{d}{2}}\end{aligned} \tag{A 14}$$

Q4.

$$\bar{k} = G'(1) = (n-1)\tau\left\{\frac{\varphi}{2\gamma+\varphi}\right\}^{\frac{d}{2}} \tag{A 15}$$

Q7.

$$\begin{aligned}
\bar{k}_{nn}(\mathbf{z}_s) &= 1 + \frac{(n-2)}{\theta(\mathbf{z}_s)} \int_{\mathbb{R}^d} p(\mathbf{z}_j)\, r(\mathbf{z}_s, \mathbf{z}_j)\, \theta(\mathbf{z}_j)\, d\mathbf{z}_j \\
&= 1 + \frac{(n-2)}{\theta(\mathbf{z}_s)} \tau^2 (2\pi\varphi)^d \times \\
&\qquad \times \int_{\mathbb{R}^d} f_d(\mathbf{z}_j; \mathbf{0}, \gamma)\, f_d(\mathbf{z}_j; \mathbf{0}, \gamma+\varphi)\, f_d(\mathbf{z}_s - \mathbf{z}_j; \mathbf{0}, \varphi)\, d\mathbf{z}_j \\
&= 1 + \frac{(n-2)}{\theta(\mathbf{z}_s)} \tau^2 (2\pi\varphi)^d \left\{2\pi(2\gamma+\varphi)\right\}^{-\frac{d}{2}} \times \\
&\qquad \times \int_{\mathbb{R}^d} f_d\left(\mathbf{z}_j; \mathbf{0}, \frac{\gamma(\gamma+\varphi)}{2\gamma+\varphi}\right) f_d(\mathbf{z}_s - \mathbf{z}_j; \mathbf{0}, \varphi)\, d\mathbf{z}_j \\
&= 1 + (n-2)\tau \left(\frac{\varphi}{2\gamma+\varphi}\right)^{\frac{d}{2}} \frac{f_d\left(\mathbf{z}_s; \mathbf{0}, \varphi + \frac{\gamma(\gamma+\varphi)}{2\gamma+\varphi}\right)}{f_d(\mathbf{z}_s; \mathbf{0}, \gamma+\varphi)} \\
&= 1 + \bar{k}\left(\frac{n-2}{n-1}\right) \frac{f_d\left(\mathbf{z}_s; \mathbf{0}, \frac{\gamma^2+3\gamma\varphi+\varphi^2}{2\gamma+\varphi}\right)}{f_d(\mathbf{z}_s; \mathbf{0}, \gamma+\varphi)}.
\end{aligned} \tag{A 16}$$

A.1.2 Proof for Corollary 2

Q1.

$$\begin{aligned}
\theta(\mathbf{z}_s) &= \int_{\mathbb{R}^d} \sum_{g=1}^{G} \pi_g f_d\left(\mathbf{z}_j; \boldsymbol{\mu}_g, \gamma_g\right) \tau (2\pi\varphi)^{\frac{d}{2}} f_d(\mathbf{z}_s - \mathbf{z}_j; \mathbf{0}, \varphi)\, d\mathbf{z}_j \\
&= \tau (2\pi\varphi)^{\frac{d}{2}} \sum_{g=1}^{G} \pi_g \int_{\mathbb{R}^d} f_d\left(\mathbf{z}_j; \boldsymbol{\mu}_g, \gamma_g\right) f_d(\mathbf{z}_s - \mathbf{z}_j; \mathbf{0}, \varphi)\, d\mathbf{z}_j \\
&= \tau (2\pi\varphi)^{\frac{d}{2}} \sum_{g=1}^{G} \pi_g f_d\left(\mathbf{z}_s; \boldsymbol{\mu}_g, \gamma_g + \varphi\right).
\end{aligned} \tag{A 17}$$

Q4.

$$\begin{aligned}
\bar{k} &= (n-1) \int_{\mathbb{R}^d} \sum_{g=1}^{G} \pi_g f_d\left(\mathbf{z}_s; \boldsymbol{\mu}_g, \gamma_g\right) \tau (2\pi\varphi)^{\frac{d}{2}} \sum_{h=1}^{G} \pi_h f_d(\mathbf{z}_s; \boldsymbol{\mu}_h, \gamma_h + \varphi)\, d\mathbf{z}_s \\
&= (n-1)\tau (2\pi\varphi)^{\frac{d}{2}} \sum_{g=1}^{G} \sum_{h=1}^{G} \pi_g \pi_h \int_{\mathbb{R}^d} f_d\left(\mathbf{z}_s; \boldsymbol{\mu}_g, \gamma_g\right) f_d(\mathbf{z}_s; \boldsymbol{\mu}_h, \gamma_h + \varphi)\, d\mathbf{z}_s \\
&= (n-1)\tau (2\pi\varphi)^{\frac{d}{2}} \sum_{g=1}^{G} \sum_{h=1}^{G} \pi_g \pi_h f_d\left(\boldsymbol{\mu}_g - \boldsymbol{\mu}_h; \mathbf{0}, \gamma_g + \gamma_h + \varphi\right).
\end{aligned} \tag{A 18}$$

While **Q7** is straightforward from (13).

A.2 Proof of Proposition 2

First, we recall a few properties of the Gaussian distribution through a Lemma:

Lemma A.1

Let $f_d(\cdot;\boldsymbol{\mu},\gamma)$ denote the d-dimensional Gaussian density centred in $\boldsymbol{\mu}$, with covariance matrix $\gamma\mathbf{I}_d$. Let also $\mathbf{x},\mathbf{u},\mathbf{v}\in\mathbb{R}^d$ and $a,b,\alpha\in\mathbb{R}^+$. Then:

$$f_d(\mathbf{x};\mathbf{u},a)\, f_d(\mathbf{x};\mathbf{v},b) = f_d(\mathbf{u}-\mathbf{v};\mathbf{0},a+b)\, f_d\left(\mathbf{x};\frac{b\mathbf{u}+a\mathbf{v}}{a+b},\frac{ab}{a+b}\right); \tag{A 19}$$

$$f_d(\alpha\mathbf{x};\mathbf{u},a) = \alpha^{-d} f_d\left(\mathbf{x};\frac{\mathbf{u}}{\alpha},\frac{a}{\alpha^2}\right). \tag{A 20}$$

Here follows the proof of Proposition 2 by mathematical induction on k. If $k=1$, then:

$$\xi_1(\mathbf{z}_i,\mathbf{z}_j) = h_1 f_d(\mathbf{z}_j-\alpha_1\mathbf{z}_i;\mathbf{0},\omega_1) = \tau(2\pi\varphi)^{\frac{d}{2}} f_d(\mathbf{z}_j-\mathbf{z}_i;\mathbf{0},\varphi) = r(\mathbf{z}_i,\mathbf{z}_j). \tag{A 21}$$

Now assume that $\xi_k(\mathbf{z}_i,\mathbf{z}_j) = h_k f_d(\mathbf{z}_j-\alpha_k\mathbf{z}_i;\mathbf{0},\omega_k)$, then we need to prove that

$$\xi_{k+1}(\mathbf{z}_i,\mathbf{z}_j) = h_{k+1} f_d(\mathbf{z}_j-\alpha_{k+1}\mathbf{z}_i;\mathbf{0},\omega_{k+1}),$$

where $h_{k+1},\alpha_{k+1},\omega_{k+1}$ are defined recursively by (27).

$$\begin{aligned}
\xi_{k+1}(\mathbf{z}_i,\mathbf{z}_j) &= \int_{\mathscr{Z}}\cdots\int_{\mathscr{Z}} p(\mathbf{z}_1)\dots p(\mathbf{z}_k) r(\mathbf{z}_i,\mathbf{z}_1)\cdots r(\mathbf{z}_k,\mathbf{z}_j)\, d\mathbf{z}_1\cdots d\mathbf{z}_k \\
&= \int_{\mathscr{Z}} p(\mathbf{z}_k)\, r(\mathbf{z}_k,\mathbf{z}_j) \int_{\mathscr{Z}}\cdots\int_{\mathscr{Z}} p(\mathbf{z}_1)\dots p(\mathbf{z}_{k-1})\times \\
&\quad \times r(\mathbf{z}_i,\mathbf{z}_1)\cdots r(\mathbf{z}_{k-1},\mathbf{z}_k)\, d\mathbf{z}_1\cdots d\mathbf{z}_k \\
&= \int_{\mathscr{Z}} p(\mathbf{z}_k)\, r(\mathbf{z}_k,\mathbf{z}_j)\, I_k(\mathbf{z}_i,\mathbf{z}_k)\, d\mathbf{z}_k \\
&= \int_{\mathscr{Z}} p(\mathbf{x})\, r(\mathbf{x},\mathbf{z}_j)\, I_k(\mathbf{z}_i,\mathbf{x})\, d\mathbf{x}.
\end{aligned} \tag{A 22}$$

Now, we introduce the Gaussian LPM assumptions and use the results of the Lemma A.1:

$$\begin{aligned}
\xi_{k+1}(\mathbf{z}_i,\mathbf{z}_j) &= \tau(2\pi\varphi)^{\frac{d}{2}} h_k \int_{\mathbb{R}^d} f_d(\mathbf{x};\mathbf{0},\gamma)\, f_d(\mathbf{x}-\mathbf{z}_j;\mathbf{0},\varphi)\, f_d(\mathbf{x}-\alpha_k\mathbf{z}_i;\mathbf{0},\omega_k)\, d\mathbf{x} \\
&= \tau(2\pi\varphi)^{\frac{d}{2}} h_k \times \\
&\quad \times \int_{\mathbb{R}^d} f_d(\mathbf{x}-\mathbf{z}_j;\mathbf{0},\varphi)\, f_d(-\alpha_k\mathbf{z}_i;\mathbf{0},\omega_k+\gamma)\, f_d\left(\mathbf{x};\frac{\gamma\alpha_k\mathbf{z}_i}{\omega_k+\gamma},\frac{\omega_k\gamma}{\omega_k+\gamma}\right) d\mathbf{x} \\
&= \tau(2\pi\varphi)^{\frac{d}{2}} h_k \alpha^{-d} f_d\left(\mathbf{z}_i;\mathbf{0},\frac{\omega_k+\gamma}{\alpha_k^2}\right)\times \\
&\quad \times \int_{\mathbb{R}^d} f_d(\mathbf{x}-\mathbf{z}_j;\mathbf{0},\varphi)\, f_d\left(\mathbf{x};\frac{\gamma\alpha_k\mathbf{z}_i}{\omega_k+\gamma},\frac{\omega_k\gamma}{\omega_k+\gamma}\right) d\mathbf{x} \\
&= h_{k+1} f_d\left(\mathbf{z}_j;\frac{\gamma\alpha_k\mathbf{z}_i}{\omega_k+\gamma},\frac{\omega_k\gamma+\omega_k\varphi+\varphi\gamma}{\omega_k+\gamma}\right) \\
&= h_{k+1} f_d(\mathbf{z}_j-\alpha_{k+1}\mathbf{z}_i;\mathbf{0}.\omega_{k+1}).
\end{aligned} \tag{A 23}$$

A.3 Proof of Corollary 3

Let G be the PGF of the random variable D, denoting the degree of a node picked at random. Then the r-th derivative of G evaluated in 1 is equal to the r-th factorial moment of D, denoted here c_r:

$$c_r = \frac{\partial^r G}{\partial x^r}(1) = \mathbb{E}\left[D\,(D-1)\cdots(D-r+1)\right]. \tag{A 24}$$

In particular:

$$\begin{aligned} c_1 &= \mathbb{E}[D] = m_1 && \text{(A 25)}\\ c_2 &= \mathbb{E}[D\,(D-1)] = \mathbb{E}\left[D^2\right] - \mathbb{E}[D] = m_2 - m_1 && \text{(A 26)}\\ &\Longrightarrow m_2 = c_1 + c_2, && \text{(A 27)} \end{aligned}$$

where m_1 and m_2 denote the first two non-central moments of D. That being said, using Corollary 1 the dispersion index can be evaluated exactly:

$$\begin{aligned} \mathscr{D} &= \frac{\mathbb{E}\left[(D-m_1)^2\right]}{m_1} = \frac{m_2 - m_1^2}{m_1} = \frac{m_2}{m_1} - m_1 = 1 + \frac{c_2}{c_1} - c_1 \\ &= 1 + \frac{(n-1)\,(n-2)\,\tau^2 \left\{\frac{\varphi^2}{(\gamma+\varphi)(3\gamma+\varphi)}\right\}^{\frac{d}{2}}}{(n-1)\,\tau \left\{\frac{\varphi}{2\gamma+\varphi}\right\}^{\frac{d}{2}}} - (n-1)\,\tau \left\{\frac{\varphi}{2\gamma+\varphi}\right\}^{\frac{d}{2}} \\ &= 1 + (n-2)\,\tau \left\{\frac{\varphi\,(2\gamma+\varphi)}{(\gamma+\varphi)\,(3\gamma+\varphi)}\right\}^{\frac{d}{2}} - (n-1)\,\tau \left\{\frac{\varphi}{2\gamma+\varphi}\right\}^{\frac{d}{2}}, \end{aligned} \tag{A 28}$$

which proves the corollary. Also, when $d = 2$, the threshold between underdispersion and overdispersion is given by:

$$\frac{(n-2)\,(2\gamma+\varphi)}{(\gamma+\varphi)\,(3\gamma+\varphi)} - \frac{(n-1)}{(2\gamma+\varphi)} = 0. \tag{A 29}$$

Now, recalling that $\varphi > 0$ and $\gamma > 0$, this is equivalent to:

$$\begin{aligned} &(n-2)\,(2\gamma+\varphi)^2 - (n-1)\,(\gamma+\varphi)\,(3\gamma+\varphi) = 0 \\ \Rightarrow\, &\varphi^2 + 4\gamma\varphi + 5\gamma^2 - n\gamma^2 = 0 \\ \Rightarrow\, &\varphi = \gamma\left(-2 \pm \sqrt{n-1}\right). \end{aligned} \tag{A 30}$$

One solution is negative thus not feasible, then the threshold is given by:

$$\varphi = \gamma\left(\sqrt{n-1} - 2\right).$$

A.4 Proof of Proposition 1

We now show how to obtain the exact formula (25) under the Gaussian LPM. We solve the numerator $\mathscr{C}_N$ and the denominator $\mathscr{C}_D$ independently.

$$\begin{aligned}
\mathscr{C}_D &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} p(\mathbf{z}_i)p(\mathbf{z}_k)p(\mathbf{z}_j)r(\mathbf{z}_i,\mathbf{z}_k)\,r(\mathbf{z}_k,\mathbf{z}_j)\,d\mathbf{z}_k d\mathbf{z}_i d\mathbf{z}_j \\
&= \int_{\mathbb{R}^d} p(\mathbf{z}_k)\left\{\int_{\mathbb{R}^d} p(\mathbf{z}_i)r(\mathbf{z}_i,\mathbf{z}_k)\,d\mathbf{z}_i\right\}\left\{\int_{\mathbb{R}^d} p(\mathbf{z}_j)r(\mathbf{z}_k,\mathbf{z}_j)\,d\mathbf{z}_j\right\}d\mathbf{z}_k \\
&= \int_{\mathbb{R}^d} p(\mathbf{z}_k)\theta(\mathbf{z}_k)^2\,d\mathbf{z}_k \qquad \text{(A 31)}\\
&= \frac{G''(1)}{(n-1)(n-2)} \\
&= \tau^2\left\{\frac{\varphi^2}{(\gamma+\varphi)(3\gamma+\varphi)}\right\}^{\frac{d}{2}}
\end{aligned}$$

Now we solve the numerator.

$$\begin{aligned}
\mathscr{C}_N &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} p(\mathbf{z}_i)p(\mathbf{z}_k)p(\mathbf{z}_j)r(\mathbf{z}_i,\mathbf{z}_k)\,r(\mathbf{z}_k,\mathbf{z}_j)\,r(\mathbf{z}_j,\mathbf{z}_i)\,d\mathbf{z}_i d\mathbf{z}_k d\mathbf{z}_j \\
&= \int_{\mathbb{R}^d} p(\mathbf{z}_i)\int_{\mathbb{R}^d} p(\mathbf{z}_j)r(\mathbf{z}_j,\mathbf{z}_i)\left\{\int_{\mathbb{R}^d} p(\mathbf{z}_k)r(\mathbf{z}_i,\mathbf{z}_k)\,r(\mathbf{z}_k,\mathbf{z}_j)\,d\mathbf{z}_k\right\}d\mathbf{z}_j d\mathbf{z}_i \qquad \text{(A 32)}\\
&= \int_{\mathbb{R}^d} p(\mathbf{z}_i)\int_{\mathbb{R}^d} p(\mathbf{z}_j)r(\mathbf{z}_j,\mathbf{z}_i)\,\xi_2(\mathbf{z}_i,\mathbf{z}_j)\,d\mathbf{z}_j d\mathbf{z}_i \\
&= \int_{\mathbb{R}^d} p(\mathbf{z}_i)\xi_3(\mathbf{z}_i,\mathbf{z}_i)d\mathbf{z}_i
\end{aligned}$$

where $\xi_k(\mathbf{z}_i,\mathbf{z}_j)$ is defined in (26) for every $k\in\mathbb{N}^0$, $\mathbf{z}_i\in\mathbb{R}^d$ and $\mathbf{z}_j\in\mathbb{R}^d$.

For more clarity, we define the recurring quantity

$$\lambda = \varphi^2+3\gamma\varphi+\gamma^2. \qquad \text{(A 33)}$$

We first discover the quantities needed to write $\xi_3(\mathbf{z}_i,\mathbf{z}_i)$ explicitly:

$$\begin{cases}\alpha_1 &= 1\\ \omega_1 &= \varphi\\ h_1 &= \tau(2\pi\varphi)^{\frac{d}{2}}\end{cases}; \qquad \begin{cases}\alpha_2 &= \frac{\gamma}{\gamma+\varphi}\\ \omega_2 &= \frac{\varphi(2\gamma+\varphi)}{\gamma+\varphi}\\ h_2 &= \tau^2(2\pi\varphi)^d f_d(\mathbf{z}_i;\mathbf{0},\gamma+\varphi)\end{cases}; \qquad \text{(A 34)}$$

$$\alpha_3 = \frac{\alpha_2\gamma}{\omega_2+\gamma} = \frac{\gamma^2}{\lambda}; \qquad \text{(A 35)}$$

$$\omega_3 = \frac{\omega_2\varphi+\omega_2\gamma+\gamma\varphi}{\omega_2+\gamma} = \frac{\varphi(\gamma+\varphi)(3\gamma+\varphi)}{\lambda}; \qquad \text{(A 36)}$$

$$h_3 = \tau^3(2\pi\varphi)^{\frac{3}{2}d} f_d(\mathbf{z}_i;\mathbf{0},\gamma+\varphi)\left(\frac{\gamma+\varphi}{\gamma}\right)^d f_d\left(\mathbf{z}_i;\mathbf{0},\frac{\lambda(\gamma+\varphi)}{\gamma^2}\right). \qquad \text{(A 37)}$$

Now, for h_3, we use Lemma A.1 and join the two Gaussian densities:

$$\begin{aligned} h_3 &= \tau^3 (2\pi\varphi)^{\frac{3}{2}d} \left(\frac{\gamma+\varphi}{\gamma}\right)^d \left\{2\pi \frac{(\gamma+\varphi)^2 (2\gamma+\varphi)}{\gamma^2}\right\}^{-\frac{d}{2}} f_d\left(\mathbf{z}_i; \mathbf{0}, \frac{\lambda}{2\gamma+\varphi}\right) \\ &= \tau^3 (2\pi\varphi)^d \left\{\frac{\varphi}{2\gamma+\varphi}\right\}^{\frac{d}{2}} f_d\left(\mathbf{z}_i; \mathbf{0}, \frac{\lambda}{2\gamma+\varphi}\right). \end{aligned} \tag{A 38}$$

Also:

$$(1-\alpha_3) = \frac{\varphi(3\gamma+\varphi)}{\lambda} \tag{A 39}$$

$$\frac{\omega_3}{(1-\alpha_3)^2} = \frac{\lambda(\gamma+\varphi)}{\varphi(3\gamma+\varphi)} \tag{A 40}$$

$$\tag{A 41}$$

Then, it follows:

$$\begin{aligned} \xi_3(\mathbf{z}_i, \mathbf{z}_i) &= h_3 (1-\alpha_3)^{-d} f_d\left(\mathbf{z}_i; \mathbf{0}, \frac{\omega_3}{(1-\alpha_3)^2}\right) \\ &= \tau^3 (2\pi\varphi)^d \left\{\frac{\varphi}{2\gamma+\varphi}\right\}^{\frac{d}{2}} f_d\left(\mathbf{z}_i; \mathbf{0}, \frac{\lambda}{2\gamma+\varphi}\right) \times \\ &\quad \times \left\{\frac{\lambda}{\varphi(3\gamma+\varphi)}\right\}^d f_d\left(\mathbf{z}_i; \mathbf{0}, \frac{\lambda(\gamma+\varphi)}{\varphi(3\gamma+\varphi)}\right). \end{aligned} \tag{A 42}$$

Collapsing again the Gaussian densities:

$$\xi_3(\mathbf{z}_i, \mathbf{z}_i) = \tau^3 \left\{\frac{2\pi\varphi^2}{2(3\gamma+\varphi)}\right\}^{\frac{d}{2}} f_d\left(\mathbf{z}_i; \mathbf{0}, \frac{\gamma+\varphi}{2}\right) \tag{A 43}$$

We can now obtain the final result for the numerator:

$$\begin{aligned} \mathscr{C}_N &= \int_{\mathbb{R}^d} p(\mathbf{z}_i) \xi_3(\mathbf{z}_i, \mathbf{z}_i) d\mathbf{z}_i \\ &= \tau^3 \left\{\frac{2\pi\varphi^2}{2(3\gamma+\varphi)}\right\}^{\frac{d}{2}} \int_{\mathbb{R}^d} f_d(\mathbf{z}_i; \mathbf{0}, \gamma) f_d\left(\mathbf{z}_i; \mathbf{0}, \frac{\gamma+\varphi}{2}\right) d\mathbf{z}_i \\ &= \tau^3 \left\{\frac{\varphi^2}{(3\gamma+\varphi)^2}\right\}^{\frac{d}{2}} \end{aligned} \tag{A 44}$$

The final formula for the clustering coefficient follows:

$$\mathscr{C} = \frac{\mathscr{C}_N}{\mathscr{C}_D} = \frac{\tau^3 \left\{\frac{\varphi^2}{(3\gamma+\varphi)^2}\right\}^{\frac{d}{2}}}{\tau^2 \left\{\frac{\varphi^2}{(\gamma+\varphi)(3\gamma+\varphi)}\right\}^{\frac{d}{2}}} = \tau \left(\frac{\gamma+\varphi}{3\gamma+\varphi}\right)^{\frac{d}{2}}. \tag{A 45}$$

A.5 Characterisation of the geodesic distances

Fronczak et al. (2004) focused on the family of fitness models for networks, which includes Erdős-Rényi random graphs and the preferential attachment model of Barabási and Albert

(1999). These models satisfy assumptions **A1** and **A2**, where the latent information is coded by a fitness value h_i, for every $i \in \mathscr{V}$. Then, edge probabilities are given by:

$$r(h_i, h_j) = \frac{h_i h_j}{\beta}, \tag{A 46}$$

where β is a suitable constant. The model includes Erdős-Rényi random graphs as a special case, for example if $h_i = \bar{k}$ for every i, and $\beta = \bar{k}(n-1)$.

Here, we exploit the fact that fitness models and LPMs both originate from LVMs, generalising the work of Fronczak et al. (2004) to a wider family of models. To study the connectivity of the networks and the path lengths' distribution, we focus on the quantities $\ell_k(\mathbf{z}_i, \mathbf{z}_j)$, defined as the probability that the shortest path between two nodes located in $\mathbf{z}_i$ and $\mathbf{z}_j$ has length k. We also define $r_k(\mathbf{z}_i, \mathbf{z}_j)$ as the probability that a path of length k exists between two nodes. In both definitions, and from now on, we condition on the fact that the two nodes are connected, i.e. there exists a finite-length path that has the two nodes as extremes. Such an assumption is natural since usually statistics of path lengths are defined only for sets of connected nodes. Note that $\xi_k(\mathbf{z}_i, \mathbf{z}_j)$ differs from $r_k(\mathbf{z}_i, \mathbf{z}_j)$ in that the latter is the probability that there is at least one k-step path between the two nodes. We now describe a way to evaluate $\ell_k(\mathbf{z}_i, \mathbf{z}_j)$ efficiently, as a function of the model parameters of a Gaussian LPM.

A5. *The graphs considered are dense enough, such that for every $(i, j) \in \mathscr{U}$, if there exists a path of length k between nodes i and j, then a path of length t exists between the same nodes for every $t = k+1, \dots, n-1$.*

Proposition A.1
Under the Gaussian LPM and assumption **A5**, for any two nodes i and j, the following two statements are equivalent:

- The geodesic distance between i and j is less than k.
- There exists a k-step path between i and j.

The proof of Proposition A.1 relies heavily on **A5** and is straightforward. From Proposition A.1 it follows that, for any i and j:

$$r_k(\mathbf{z}_i, \mathbf{z}_j) = \sum_{t=1}^{k} \ell_t(\mathbf{z}_i, \mathbf{z}_j). \tag{A 47}$$

Moreover, since $\ell_1(\mathbf{z}_i, \mathbf{z}_j) = r_1(\mathbf{z}_i, \mathbf{z}_j) = r(\mathbf{z}_i, \mathbf{z}_j)$, the following holds:

$$\ell_k(\mathbf{z}_i, \mathbf{z}_j) = r_k(\mathbf{z}_i, \mathbf{z}_j) - r_{k-1}(\mathbf{z}_i, \mathbf{z}_j). \tag{A 48}$$

Hence, we aim at characterising $r_k(\mathbf{z}_i, \mathbf{z}_j)$, thereby deducing the properties of $\ell_k(\mathbf{z}_i, \mathbf{z}_j)$.

Each possible path of length k from i to j can be thought of as a Bernoulli random variable, having a success if all the edges involved in the path appear, or not having a success if any of those edges fail to appear. For an Erdős-Rényi random graph with average degree $\bar{k} = (n-1)p$, the parameter of such a random variable is p^k. For Gaussian LPMs, the success probability is $\xi_k(\mathbf{z}_i, \mathbf{z}_j)$, which has been characterised in Proposition 2.

However, we are interested in $r_k(\mathbf{z}_i, \mathbf{z}_j)$, which is the probability of the union of all the k-steps paths from i to j. Unfortunately, these variables are not independent, since different paths will have edges in common. We circumvent this issue by pretending that all such paths are mutually independent, following the reasoning of Fronczak et al. (2004). This assumption makes sense when k is much smaller than n. In fact, for the purpose of the study of shortest path lengths, estimates of $r_k(\mathbf{z}_i, \mathbf{z}_j)$ will be needed only for small ks, since in the general case $\ell_k(\mathbf{z}_i, \mathbf{z}_j)$ will drop to zero very quickly.

Using the results of Proposition 2 and Lemma 1 of Fronczak et al. (2004), we can approximate (A 48) by:

$$\ell_k(\mathbf{z}_i, \mathbf{z}_j) \approx \exp\left\{-n^{k-1}\xi_{k-1}(\mathbf{z}_i, \mathbf{z}_j)\right\} - \exp\left\{-n^k \xi_k(\mathbf{z}_i, \mathbf{z}_j)\right\}. \qquad \text{(A 49)}$$

Equation (32) gives a general formula to evaluate the distribution of the geodesic distance $\ell_k(\mathbf{z}_i, \mathbf{z}_j)$ for every $k << n$ for dense Gaussian LPM networks.

References

Barabási, A. L., and R. Albert. 1999. "Emergence of scaling in random networks". *Science* 286 (5439): 509–512.

Fronczak, A., P. Fronczak, and J. A. Hołyst. 2004. "Average path length in random networks". *Physical Review E* 70 (5): 056110.

Newman, M. E. J., S. H. Strogatz, and D. J. Watts. 2001. "Random graphs with arbitrary degree distributions and their applications". *Physical Review E* 64 (2): 026118.