APPENDIX 1 : DIVERSITY OF PARAMETER UPDATES

Here we provide proofs Theorem 1 from the main paper concerning the diversity of the parameter updates.

Theorem 8.1. *In a multi-agent evolution strategies update iteration t for a system with N agents with parameters $\Theta = \{\theta_1^{(t)}, ..., \theta_N^{(t)}\}$, agent communication matrix $A = \{a_{ij}\}$, agent-wise perturbations $\mathcal{E} = \{\epsilon_1^{(t)}, ..., \epsilon_N^{(t)}\}$, and parameter update $u_i^{(t)}$ given by the sparsely-connected update rule:*

$$u_i^{(t)} = \frac{\alpha}{N\sigma^2}\sum_{j=1}^{N} a_{ij} \cdot \left(R(\theta_j^{(t)} + \sigma\epsilon_j^{(t)}) \cdot ((\theta_j^{(t)} + \sigma\epsilon_j^{(t)}) - (\theta_i^{(t)}))\right)$$

The following relation holds:

$$\mathrm{Var}_i[u_i^{(t)}] \leq \frac{\max^2 R(\cdot)}{N\sigma^4}\Big\{\Big(\frac{\|A^2\|_F}{(\min_l |A_l|)^2}\Big) \cdot f(\Theta, \mathcal{E}) - \Big(\frac{\min_l |A_l|}{\max_l |A_l|}\Big)^2 \cdot g(\mathcal{E})\Big\} \quad (5)$$

Here, $|A_l| = \sum_j a_{jl}$, $f(\Theta, \mathcal{E}) = \Big(\sum_{j,k,m}^{N,N,N} \big((\theta_j^{(t)} + \sigma\epsilon_j^{(t)} - \theta_m^{(t)}) \cdot (\theta_k^{(t)} + \sigma\epsilon_k^{(t)} - \theta_m^{(t)})\big)^2\Big)^{\frac{1}{2}}$, and $g(\mathcal{E}) = \frac{\sigma^2}{N}\Big(\sum_{i,j}^{N,N} \epsilon_i^{(t)}\epsilon_j^{(t)}\Big)$.

Proof. From Equation 8.1, the update rule is given by:

$$u_i^{(t)} = \frac{\alpha}{N\sigma^2}\sum_{j=1}^{N} a_{ij} \cdot \left(R(\theta_j^{(t)} + \sigma\epsilon_j^{(t)}) \cdot ((\theta_j^{(t)} + \sigma\epsilon_j^{(t)}) - (\theta_i^{(t)}))\right) \quad (6)$$

The variance of $u_i^{(t)}$ can be written as:

$$\mathrm{Var}_i[u_i^{(t)}] = \mathbb{E}_{i\in\mathcal{A}}[(u_i^{(t)})^2] - (\mathbb{E}_{i\in\mathcal{A}}[(u_i^{(t)})])^2 \quad (7)$$

Expanding $\mathbb{E}_{i\in\mathcal{A}}[(u_i^{(t)})^2]$:

$$= \frac{1}{N}\sum_{i\in\mathcal{A}}\Big\{\frac{\gamma}{N\sigma^2}\sum_{j=1} a_{ij} \cdot R(\theta_j^{(t)} + \sigma\epsilon_j^{(t)}) \cdot (\theta_j^{(t)} + \sigma\epsilon_j^{(t)} - \theta_i^{(t)})\Big\}^2 \quad (8)$$

Simplifying:

$$= \frac{1}{N\sigma^4}\sum_{i,j,k}\Big(\frac{a_{ij}a_{ik}}{|A_i|^2} R(\theta_j^{(t)} + \sigma\epsilon_j^{(t)}) R(\theta_k^{(t)} + \sigma\epsilon_k^{(t)}) \cdot (\theta_j^{(t)} + \sigma\epsilon_j^{(t)} - \theta_i^{(t)}) \cdot (\theta_k^{(t)} + \sigma\epsilon_k^{(t)} - \theta_i^{(t)})\Big) \quad (9)$$

Since $R(\cdot) \leq \max R(\cdot)$, therefore:

$$\leq \frac{\max^2 R(\cdot)}{N\sigma^4}\sum_{i,j,k}\frac{a_{ij}a_{ik}}{|A_i|^2} \cdot (\theta_j^{(t)} + \sigma\epsilon_j^{(t)} - \theta_i^{(t)}) \cdot (\theta_k^{(t)} + \sigma\epsilon_k^{(t)} - \theta_i^{(t)}) \quad (10)$$

$$\leq \frac{\max^2 R(\cdot)}{N\sigma^4}\sum_{i,j,k}\frac{a_{ij}a_{ik}}{\min_l |A_l|^2} \cdot (\theta_j^{(t)} + \sigma\epsilon_j^{(t)} - \theta_i^{(t)}) \cdot (\theta_k^{(t)} + \sigma\epsilon_k^{(t)} - \theta_i^{(t)}) \quad (11)$$

By the Cauchy-SchDwarz Inequality:

$$\mathbb{E}_{i\in\mathcal{A}}[(u_i^{(t)})^2] \leq \frac{\max^2 R(\cdot)}{N\sigma^4}\Big(\sum_{i,j,k}\frac{(a_{ij}a_{ik})^2}{\min_l |A_l|^4}\Big)^{\frac{1}{2}} \cdot \Big(\sum_{i,j,k}\big((\theta_j^{(t)} + \sigma\epsilon_j^{(t)} - \theta_i^{(t)}) \cdot (\theta_k^{(t)} + \sigma\epsilon_k^{(t)} - \theta_i^{(t)})\big)^2\Big)^{\frac{1}{2}} \quad (12)$$

Since $a_{ij} \in \{0, 1\} \forall\ (i, j), (a_{ij}a_{ik})^2 = a_{ij}a_{ik}\ \forall (i, j, k)$. Additionally, we know that $a_{ij} = a_{ji}$, since A is symmetric. Therefore, $\sum_i a_{ij}a_{ik} = \sum_i a_{ji}a_{ik} = A^2_{jk}$. Using this:

$$\mathbb{E}_{i\in\mathcal{A}}[(u_i^{(t)})^2] \leq \frac{\max^2 R(\cdot)}{N\sigma^4} \cdot \Big(\frac{|A^2|^{\frac{1}{2}}}{\min_l |A_l|^2}\Big) \cdot \Big(\sum_{i,j,k}\big((\theta_j^{(t)} + \sigma\epsilon_j^{(t)} - \theta_i^{(t)}) \cdot (\theta_k^{(t)} + \sigma\epsilon_k^{(t)} - \theta_i^{(t)})\big)^2\Big)^{\frac{1}{2}} \quad (13)$$

Replacing $\Big(\sum_{i,j,k}\big((\theta_j^{(t)} + \sigma\epsilon_j^{(t)} - \theta_i^{(t)}) \cdot (\theta_k^{(t)} + \sigma\epsilon_k^{(t)} - \theta_i^{(t)})\big)^2\Big)^{\frac{1}{2}} = f(\Theta, \mathcal{E})$, where $\Theta = \{\theta_i^{(t)}\}_{i=1}^N, \mathcal{E} = \{\epsilon_i\}_{i=1}^N$ for compactness, we obtain:

$$\mathbb{E}_{i\in\mathcal{A}}[(u_i^{(t)})^2] \leq \frac{\max^2 R(\cdot)}{N\sigma^4} \cdot \Big(\frac{|A^2|^{\frac{1}{2}}}{\min_l |A_l|^2}\Big) \cdot f(\Theta, \mathcal{E}) \quad (14)$$

Similarly, the squared expectation of $(u_i^{(t)})$ over all agents can be given by:

$$(\mathbb{E}_{i\in\mathcal{A}}[u_i^{(t)}])^2 = \Big(\frac{1}{N}\sum_{i\in\mathcal{A}}\Big\{\frac{\gamma}{N\sigma^2}\sum_{j=1} a_{ij} \cdot R(\theta_j^{(t)} + \sigma\epsilon_j^{(t)}) \cdot (\theta_j^{(t)} + \sigma\epsilon_j^{(t)} - \theta_i^{(t)})\Big\}\Big)^2 \quad (15)$$

$$= \frac{1}{N^2\sigma^4}\Big(\sum_{i\in\mathcal{A}}\Big\{\frac{1}{|A_i|}\sum_{j=1} a_{ij} \cdot R(\theta_j^{(t)} + \sigma\epsilon_j^{(t)}) \cdot (\theta_j^{(t)} + \sigma\epsilon_j^{(t)} - \theta_i^{(t)})\Big\}\Big)^2 \quad (16)$$

$$= \frac{1}{N^2\sigma^4}\Big(\sum_{i,j}\Big\{\frac{a_{ij}}{|A_i|} \cdot R(\theta_j^{(t)} + \sigma\epsilon_j^{(t)}) \cdot (\theta_j^{(t)} + \sigma\epsilon_j^{(t)} - \theta_i^{(t)})\Big\}\Big)^2 \quad (17)$$

Since $R(\cdot) \geq \min R(\cdot)$, therefore:

$$\geq \frac{\min^2 R(\cdot)}{N^2\sigma^4}\Big(\sum_{i,j}\Big\{\frac{a_{ij}}{|A_i|} \cdot (\theta_j^{(t)} + \sigma\epsilon_j^{(t)} - \theta_i^{(t)})\Big\}\Big)^2 \quad (18)$$

$$\geq \frac{\min^2 R(\cdot)}{N^2\sigma^4 \max_l |A_l|^2}\Big(\sum_{i,j}\Big\{a_{ij} \cdot (\theta_j^{(t)} + \sigma\epsilon_j^{(t)} - \theta_i^{(t)})\Big\}\Big)^2 \quad (19)$$

Since A is symmetric, $\sum_{i,j}^{N,N} a_{ij} \cdot (\theta_j^{(t)} + \sigma\epsilon_j - \theta_i^{(t)}) = \sum_{i,j}^{N,N} a_{ij} \cdot (\theta_i^{(t)} + \sigma\epsilon_i - \theta_j^{(t)})$. Therefore:

$$= \frac{\min^2 R(\cdot)}{N^2\sigma^4 \max_l |A_l|^2}\Big(\sum_{i,j}\frac{1}{2}\Big\{a_{ij} \cdot (\theta_j^{(t)} + \sigma\epsilon_j^{(t)} - \theta_i^{(t)}) + a_{ij} \cdot (\theta_i^{(t)} + \sigma\epsilon_i^{(t)} - \theta_j^{(t)})\Big\}\Big)^2 \quad (20)$$

Therefore,

$$(\mathbb{E}_{i\in\mathcal{A}}[u_i^{(t)}])^2 = \frac{\min^2 R(\cdot)}{N^2\sigma^2 \max_l |A_l|^2}\Big(\sum_{i,j}\frac{1}{2}\big\{a_{ij}\cdot(\epsilon_j^{(t)}+\epsilon_i^{(t)})\big\}\Big)^2 \quad (21)$$

Using the symmetry of A, we have that $\sum_{i,j}^{N,N} a_{ij}\epsilon_i = \sum_{i,j}^{N,N} a_{ij}\epsilon_j$. Therefore:

$$= \frac{\min^2 R(\cdot)}{N^2\sigma^2 \max_l |A_l|^2}\Big(\sum_{i,j} a_{ij}\cdot\epsilon_j^{(t)}\Big)^2 \quad (22)$$

$$= \frac{\min^2 R(\cdot)}{N^2\sigma^2 \max_l |A_l|^2}\Big(\sum_{j} |A_j|\cdot\epsilon_j^{(t)}\Big)^2 \quad (23)$$

$$\geq \frac{\min^2 R(\cdot)\min_l |A_l|^2}{N^2\sigma^2 \max_l |A_l|^2}\Big(\sum_{i,j} \epsilon_i^{(t)}\epsilon_j^{(t)}\Big) \quad (24)$$

Combining both terms of the variance expression, and using the normalization of the iteration rewards that ensures $\min R(\cdot) = -\max R(\cdot)$, we can obtain (using $g(\mathcal{E}) = \frac{\sigma^2}{N}\Big(\sum_{i,j}\epsilon_i^{(t)}\epsilon_j^{(t)}\Big)$):

$$\mathsf{Var}_{i\in\mathcal{A}}[u_i^{(t)}] \leq \frac{\max^2 R(\cdot)}{N\sigma^4}\Big\{\Big(\frac{|A^2|^{\frac{1}{2}}}{\min_l |A_l|^2}\Big)\cdot f(\Theta,\mathcal{E}) - \Big(\frac{\min_l |A_l|^2}{\max_l |A_l|^2}\Big)\cdot g(\mathcal{E})\Big\} \quad (25)$$

□

APPENDIX 2 : APPROXIMATING REACHABILITY AND HOMOGENEITY FOR LARGE ERDOS-RENYI GRAPHS

Recall that a Erdos-Renyi graph is constructed in the following way

(1) Take n nodes
(2) For each pair of nodes, link them with probability p

The model is simple, and we can infer the following:

- The average degree of a node is $p(n-1)$
- The distribution of degree for the nodes is the Binomial distribution of $n-1$ events with probability p, $B(n-1,p)$.
- The (average) number of paths of length 2 from one node i to a node $j \neq i$ ($n_{ij}^{(2)}$) can be calculated this way: a path of length two between i and j involves a third node k. Since there are $n-2$ of them, the maximun number of paths between i and j is $n-2$. However, for that path to exists there has to be a link between i and k and k and j, an event with probability p^2. Thus, the average number of paths between i and j is $p^2(n-2)$

Estimating Reachability

We can then estimate Reachability:

$$Reachability = \frac{||A^2||_F}{(min_l|A_l|)^2} = \frac{\sqrt{\sum_{i,j} n_{ij}^{(2)}}}{k_{min}^2}$$

where $k_{min} = (min_l|A_l|)$ is the minimum degree in the network. Given the above calculations we can approximate

$$\sum_{i,j} n_{ij}^{(2)} = \sum_i n_{ii}^{(2)} + \sum_{i\neq j} n_{ij}^{(2)} \approx n\times[p(n-1)] + n(n-1)\times[p^2(n-2)]$$

where the first term is the number of paths of length 2 from i to i summed over all nodes, i.e. the sum of the degrees in the network. The second term is the sum of $p^2(n-2)$ for the terms in which $i \neq j$. For large n we have that

$$\sum_{i,j} n_{ij}^{(2)} \approx p^2 n^3$$

and thus,

$$||A^2||_F \approx \sqrt{p^2n^3}. \quad (26)$$

For the denominator k_{min} we could use the distribution of the minimum of the binomial distribution $B(n-1,p)$. However, since it is a complicated calculation we can approximate this way: since the binomial distribution $B(n-1,p)$ looks like a Gaussian, we can say that the minimum of the distribution is closed to the mean minus two times the standard deviation:

$$k_{min} \approx p(n-1) - 2\sqrt{p(n-1)(1-p)} \quad (27)$$

Once again in the case of large n we have

$$k_{min} \approx pn$$

Thus

$$Reachability \approx \frac{\sqrt{p^2n^3}}{[p(n-1) - 2\sqrt{p(n-1)(1-p)}]^2} \quad (28)$$

Assuming that n is large, we can approximate

$$Reachability \approx \frac{pn^{3/2}}{p^2n^2} = \frac{1}{pn^{1/2}}$$

Thus the bound decreases with increasing n and p. Note that the density of the Erdos-Renyi graph (the number of links over the number of possible links) is p. And thus for a fixed n more sparse networks $p \simeq 0$ have larger Reachability than more connected networks $p \simeq 1$.

Estimating Homogeneity

The Homogeneity is defined as

$$Homogeneity = \Big(\frac{k_{min}}{k_{max}}\Big)^2$$

As before we can approximate

$$k_{max} \approx p(n-1) + 2\sqrt{p(n-1)(1-p)}$$

And thus

$$Homogeneity \approx \Bigg(\frac{p(n-1) - 2\sqrt{p(n-1)(1-p)}}{p(n-1) + 2\sqrt{p(n-1)(1-p)}}\Bigg)^2$$

For large p we can approximate it to be

$$Homogeneity \approx 1 - 8\frac{\sqrt{1-p}}{\sqrt{np}} \quad (29)$$

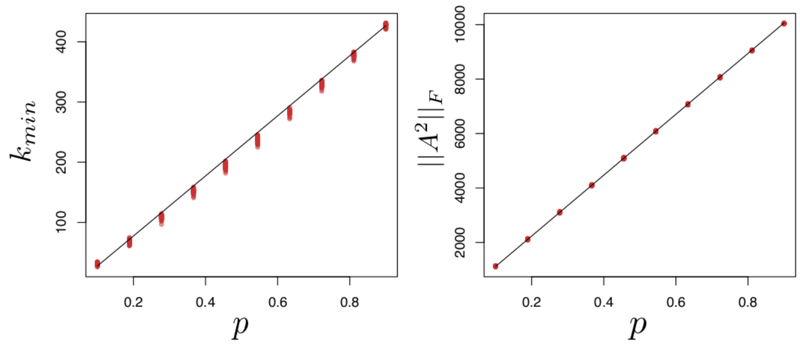

Figure 6: Comparison between the values of k_{min}, $||A^2||_F$, and Reachability as a function of p for different realizations of the Erdos-Renyi model (points) and their approximations given in Equations (27), (26) and (28) respectively (lines).

which shows that for $p \simeq 1$ we have that Homogeneity grows as a function of p. Thus for fixed number of nodes n, increasing p we get larger values of the Homogeneity.