Supplementary Materials for

Socioeconomic correlations and stratification in social-communication networks

Y. Leo, E. Fleury, J. I. Alvarez-Hamelin, C. Sarraute, M. Karsai*

*Corresponding author email: marton.karsai@ens-lyon.fr

1 Degree and wealth correlations

In the main text we studied a social network where each individual was assigned with a socioeconomic indicator defined as their average monthly purchase (AMP) (see Eq.1 in the main text). We used these indicators to estimate the socioeconomic status of individuals and group them into 9 exclusive socioeconomic classes. By analysing the social network and the assigned socioeconomic classes we observed that individuals tend to connect to similar others from their own or neighbouring socioeconomic classes, while social ties with people from remote classes are less frequent. We argued that this observed stratification in the social structure (see in Fig.3a-c in the main text) is due to the entangled effects of socioeconomic imbalances and status homophily, i.e., the tendency of people to connect to others with similar socioeconomic status.

However, one can argue that the observed stratified structure can be simply the consequence of simultaneously present degree-degree and degree-wealth correlations. More precisely, if the degree of an individual is highly correlated with its economic status (wealth) and the network is strongly assortative (i.e., people prefer to connect to other people with similar degrees) we may observe similar effects as in Fig.1 (in the main text). To close out this possibility we present here a correlation analysis and a null model study where we carefully define random reference models to remove the correlations in focus in a controlled way and check their effects on the quantitative observations.

1.1 Degree-degree correlations

The simplest way to characterise degree-degree correlations in a network is by computing the Pearson correlation coefficient between two random variables identified as the degrees of nodes connected in the network structure [1]. After calculating the Pearson correlation coefficient in the investigated social network we found that it is $r \approx -0.00813$ ($p < 0.001$, $SE = 7.13 \times 10^{-4}$), suggesting that the network shows no (or very weak disassortative) degree-degree correlations. However, since the Pearson correlation coefficient gives only an overall characteristic measure and assumes that correlations are linear, we further investigate degree-degree correlations with another metric, which is conventionally used to characterise degree-degree correlations. We measure the $k_{nn}(k)$ average nearest neighbour degree for each degree class k in the network [1]. This function (shown in Fig.S1a) disclosed a somewhat more sophisticated picture about degree-degree correlations. First of all it is not a monotonous function but it assigns mixed effects of assortative and disassortative mixing. It shows positive (assortative) correlations up to $k \simeq 10$, which after it indicates negative (disassortative) correlations, and becomes flat for the largest degrees. Consequently our network does not show strong assortative correlations over its whole degree range, which suggests that degree correlations may not evidently play a deterministic role in the observed stratified structure even if they are correlated with wealth. Note that this type of complex functional scaling of $k_{nn}(k)$ commonly characterises non-mutualised directed networks as discussed in [2]. This is in line with our case where the network was not mutualised but directed links were kept in the structure, which were assumed to be indirected after the network construction for the convenience of our study.

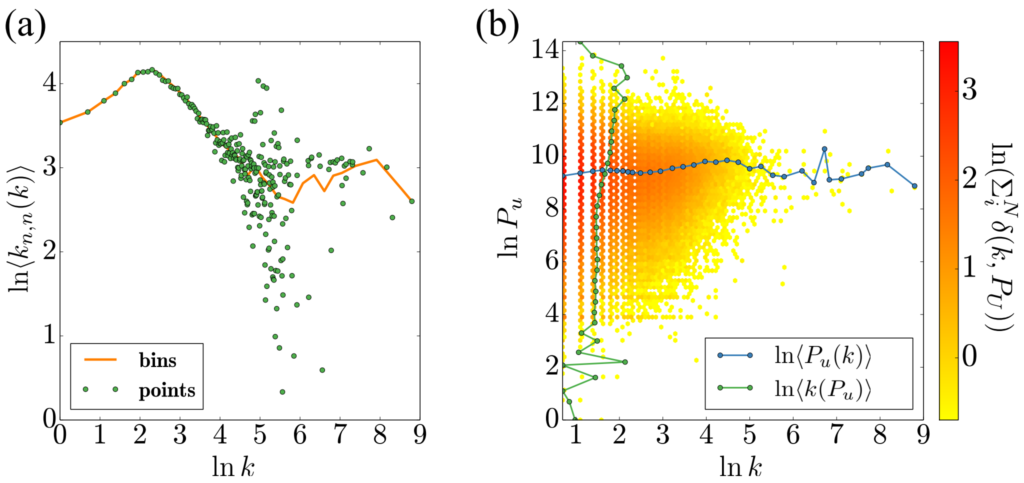

Figure S1. Degree-degree and degree-wealth correlations in the social-economical network. (a) The $k_{nn}(k)$ function computed in the social network. (b) Correlation plot (shown as a heat map) of degree-wealth correlations. Blue (resp. green) horizontal (resp. vertical) solid line and symbols show the average wealth (resp. degree) as the function of degree (resp. wealth).

1.2 Degree-wealth correlations

Entangled with degree correlations, dependencies between node degree and economic status (wealth) can also contribute to the emergence of the observed stratified structure. To characterise this correlation, first we measure again the Pearson correlation coefficient between the degree k and AMP value P_u of each node. This correlation turns out to be small with coefficient $r \approx 0.0357$ ($p < 0.001$, $SE = 9.71 \times 10^{-1}$). To obtain a more complete picture about their dependencies we simply show in Fig.1b the binned scatter plot, as a heat map, of these two variables and in addition calculate the average value of wealth (resp. degree) as the function of degree (resp. wealth). These results indicate weak dependencies between these variables for their whole range and although they un-disclose some non-monotonous dependencies between node degree and economic status they do not indicate evident strong positive correlations between them.

1.3 Null model study

In the main text by measuring the fraction $L(s_i, s_j) = |E(s_i, s_j)|/|E_{rn}(s_i, s_j)|$ (see also in Eq.4 in the main text) we concluded that the social structure is stratified in terms of socioeconomic status. In this measure we used a null model structure defined as the configuration network model of the empirical network (here we call null model 1 (NM1)), where we take the original social network, select random pairs of links and swap them without allowing multiple links and self loops. In order to remove any residual correlations we repeated this procedure $5 \times |E|$ times (where $|E|$ being the number of links in the social network) and calculated averages over 100 independent random realisations. This randomisation keeps the number of links, individual economic indicators P_u, and the assigned class of people unchanged, but destroys degree-degree correlations, possibly present community structure (for a summary of present correlations see Table S1). Since it destroys all possible structural correlations in the network (apart from correlations due to unavoidable finite size effects) as a consequence it eliminates the socioeconomic layers as well. Results shown in Fig.S2b appears with a diagonal component, which evidently assigns strong connectivity between neighbouring socioeconomic classes, i.e., a stratified structure. This measure simply assigns *how many times people of different classes are more connected as compared to the case where they are connected by chance.* The observed diagonal component indeed assigns the significance of this correlations what we associated to status homophily.

To further investigate potential reasons behind this observation, let's take a null model to directly address the possible effects of degree homophily. In this null model (NM2) we destroy the potential community structure and all other structural correlations including socioeconomic layers, but conserve degree-degree and degree-wealth correlations (see Table S1). NM2 is defined as a modification of the configuration network model, where instead of selecting link pairs randomly to swap, we select a link and one of its end randomly, and choose another link randomly where the degree of one of the ending nodes is equal to the degree of the selected end of the first link. Swapping the other ends of the links (with potentially different degrees) will result yet two links between nodes of the original degrees but connected randomly otherwise. We do not allow self loops and multiple links and skip to swap links where both node degrees are unique in the network. We found 22 such cases from $\sim 2M$ links thus we assume this condition will not bias our shuffling considerably. We swapped randomly link pairs $5 \times |E|$ just as for the NM1 model and computed averages over 100 realisations.

	Original	NM1	NM2
degree-degree	✓	×	✓
degree-wealth	✓	✓	✓
communities	✓	×	×

Table S1. Correlations present in different null models. We consider degree-degree, degree-wealth, and higher-order structural (communities) correlations in the original network (Orig) and two null models (for definitions see text).

Our hypothesis is that if the present degree-degree correlations in NM2 would explain the observed stratified structure, then after using the corresponding $|E_{rn}^{NM2}(s_i, s_j)|$ link density matrix in the normalisation of $L(s_i, s_j)$ (see Eq.4 in the main text) the resulting matrix should become flat. This would mean that the actually present correlations could explain (reproduce) the empirical observations. However, this is not the case here as seen in Fig.S2b. The $L(s_i, s_j)$ matrix normalised by the corresponding NM2 matrix appears to be almost identical than the one normalised by the NM1 matrix. This suggests that degree-degree correlations and degree homophily do not play a role here, thus it cannot explain the emergence of the stratified structure.

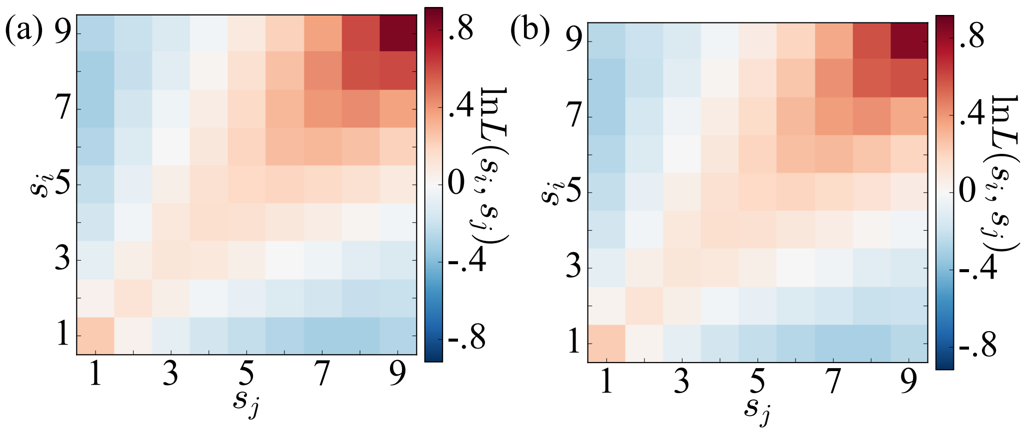

Figure S2. $L(s_i, s_j)$ normalised socioeconomic class connectivity matrices in case of two null models (for definition see Eq.4 in the main text). In each case the numerator was taken as the socioeconomic class connectivity matrices of the original network, while the denominator was measured from a null model structure of (a) NM1, (b) NM2.

References

[1] Newman M. 2010 *Networks: An Introduction.* (Oxford University Press, USA), 1 edn.

[2] Li M.-X., et. al. 2014 A comparative analysis of the statistical properties of large mobile phone calling networks. *Sci. Rep.* **4,** 5132 . (DOI: 10.1038/srep05132)