

# Disentangling degree and tie strength heterogeneity in egocentric social networks

Sara Heydari<sup>1\*</sup>, Gerardo Iñiguez<sup>2,3,1,4</sup>, János Kertész<sup>3,5</sup>, and Jari Saramäki<sup>1\*</sup>

<sup>1</sup>*Department of Computer Science, Aalto University School of Science, 00076 Aalto, Finland*

<sup>2</sup>*Faculty of Information Technology and Communication Sciences, Tampere University, 33720 Tampere, Finland*

<sup>3</sup>*Department of Network and Data Science, Central European University, 1100 Vienna, Austria*

<sup>4</sup>*Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, 04510 Ciudad de México, Mexico*

<sup>5</sup>*Complexity Science Hub, 1080 Vienna, Austria*

*\*Corresponding authors' email: sara.heydari@aalto.fi, jari.saramaki@aalto.fi*

## Abstract

The structure of personal networks reflects how we organise and maintain social relationships. The distribution of tie strengths in personal networks is heterogeneous, with a few close, emotionally intense relationships and a larger number of weaker ties. Recent results indicate this feature is universal across communication channels. Within this general pattern, there is a substantial and persistent inter-individual variation that is also similarly distributed among channels. The reason for the observed universality is yet unclear—one possibility is that people's traits determine their personal network features on any channel. To address this hypothesis, we need to compare an individual's personal networks across channels, which is a non-trivial task: while we are interested in measuring the differences in tie strength heterogeneity, personal network size is also expected to vary a lot across channels. Therefore, for any measure that compares personal networks, one needs to understand the sensitivity with respect to network size. Here, we study different measures of personal network similarity and show that a recently introduced alter-preferentiality parameter and the Gini coefficient are equally suitable measures for tie strength heterogeneity, as they are fairly insensitive to differences in network size. With these measures, we show that the earlier observed individual-level persistence of personal network structure cannot be attributed to network size stability alone, but that the tie strength heterogeneity is persistent too. We also demonstrate the effectiveness of the two measures on multichannel data, where tie strength heterogeneity in personal networks is seen to moderately correlate for the same users across two communication channels (calls and text messages).

**Keywords**— social network analysis, personal networks, egocentric networks, tie strength heterogeneity, social signatures, persistence

## 1 Introduction

Social relationships play a fundamentally important role in our lives. On an individual level, they constitute a person's social capital [30]: they offer social support and engagement, transmit social influence, and provide access to resources and commodities [3]. At the system level, they keep society connected

and functional [7]. Social relationships vary in their purposes and levels of emotional closeness [10, 27]. Personal social networks typically consist of a handful of close and emotionally intense relationships (strong ties), essential for our well-being and health [12, 13, 22, 41], and a larger number of less close relationships (weak ties) that provide us with diversity and opportunities. Weak ties are crucial for the integrity of society since they function as bridges between otherwise separate parts of the social web [8, 10, 18, 35].

Tie strength diversity can be quantified by analyzing the features of egocentric networks [11, 25, 34]. An egocentric network (ego-net) consists of an individual (the ego) and the social ties connecting the ego to its direct neighbors (alters) [23]. Such networks can be readily constructed from communication data, where the frequency of interaction is typically used as a proxy of tie strength [28]. Several studies have provided evidence of tie strength heterogeneity in ego-nets inferred from mobile phone data [11, 21, 34], with recent findings on multiple online platforms indicating strong similarities in ego-net structural patterns across communication channels [14].

The extent of tie strength heterogeneity in ego-nets varies from person to person. Whatever the communication channel, there are distinct individual differences that are similarly distributed across channels [14]. These individual patterns persist over time [11, 21, 34], even if there are major changes in the composition of alters. In other words, people have *social signatures* [34] and their networks tend to retain their characteristic feature of heterogeneity, independently of who their alters are.

One candidate explanation for this surprising universality – of both tie strength heterogeneity and of the distribution of individual variation across channels – is that it reflects some ego-level latent variable, whose distribution in the population determines the variety of ego-net tie strengths. Here, personality traits [6] are an obvious candidate to investigate. However, people communicate on different channels with varying levels of activity [37], resulting in variations in network size that might obscure the role of personality traits in ego-net evolution.

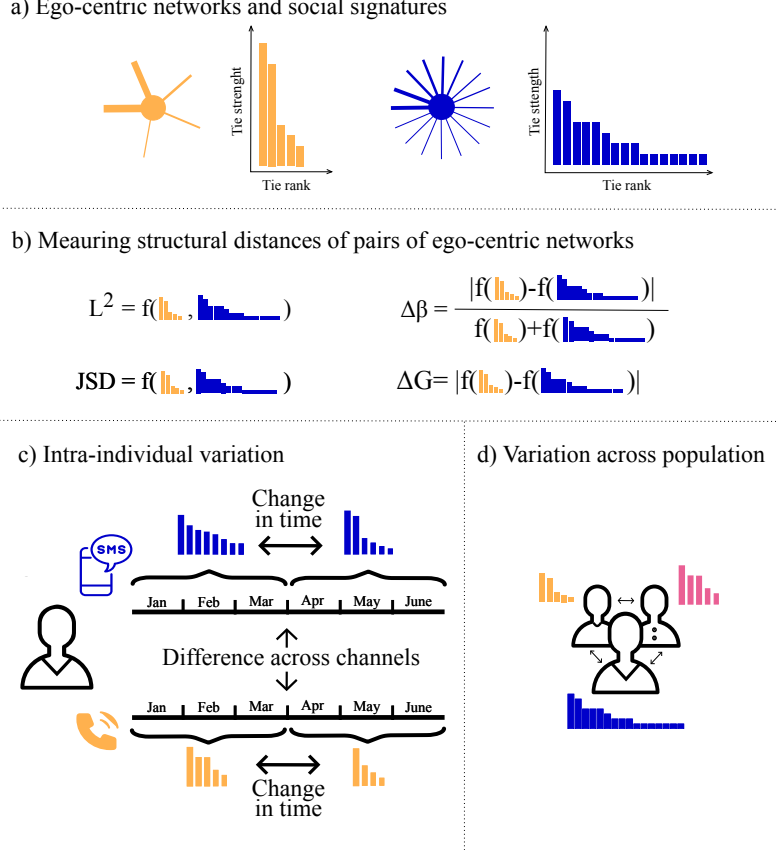
Thus, we need a way of comparing personal networks that captures the similarity of tie strength heterogeneity despite differences in network size. This would help us clarify whether the methods used in earlier work (e.g., [34]) really capture the persistence of tie strength heterogeneity, rather than the persistence of network size alone. To this end, we explore four methods of measuring differences in heterogeneity between pairs of ego-nets and examine their sensitivity to ego-net size (i.e., the network degrees of the egos).

We demonstrate that some of the studied measures are less sensitive to ego-net degree and are therefore a better option for comparing the levels of tie strength heterogeneity of networks of different sizes. We also apply two of the measures to multichannel data on mobile telephone call and text message (SMS) logs, finding that the heterogeneity levels of egos are moderately correlated across these two channels. This observation supports the notion of an individual-level latent variable shaping ego-net characteristics and evolution.

## 2 Data and Methods

### 2.1 Data

We use two mobile phone datasets in this study, both including call and short message (SMS) logs. The larger dataset includes communication logs of more than five million users over a six-month period and



**Figure 1. Ego-nets and distance functions for measuring their structural differences.** (a) Two weighted ego-nets and their corresponding non-normalized social signatures. We construct an ego’s social signature by sorting ties in descending order of strength. (b) Distance functions for measuring structural differences between ego-net pairs. The  $L^2$  and JSD distances directly assign a value to each ego-net pair. In contrast, when measuring the distance using the other two functions, preferentiality  $\beta$  and the Gini coefficient  $G$ , a heterogeneity value is first calculated for each ego-net, and then the distances are measured. (c) Intra-individual variations (self-distances) of an ego are calculated by measuring the distances between its ego-nets in two consecutive windows (change in time) and by measuring the distance between its ego-nets across communication channels (difference across channels). (d) To determine whether the self-distances are smaller or larger than expected, we need to compare them with a reference. We make the reference distance distribution by calculating the distance between ego-nets of random pairs of egos in the population.

dates back to 2007, a time when mobile phone calls and SMS were more widely used, as smartphone-app-based alternatives did not yet exist. We refer to the larger data as the Large Mobile Phone (LMP) dataset and use it to examine the sensitivity of ego-nets’ structural similarity measures to ego-nets’ size difference. The smaller dataset is known as the Copenhagen Networks Study (CNS) dataset [33] and consists of communication logs between a group of around one thousand university students in 2012-2013. After determining the robustness of the measures for ego-net comparison, we use both datasets to compare heterogeneity across different communication channels. For a detailed description of these two datasets, see Supplementary Information (SI) SI Section S1.1 and SI Section S1.2.

## 2.2 Egocentric networks and social signatures

An egocentric network consists of the ego and the ego’s alters, i.e., friends and family members that have direct social ties with the ego. In weighted ego-nets, the links are associated with weights that represent

tie strength, thus being proxies of relationship intensity. There are several approaches for defining the weights using communication logs [11, 39]. In this paper, we use the accumulated frequencies of communication events (numbers of calls or messages) during a certain time window as the link weights, similarly to approaches used in [25, 28, 34, 40]. See Fig. 1 for schematic visualizations of two example ego-nets.

People vary in how they distribute their communication resources among their alters. To capture this inhomogeneity, the notion of *social signatures* was introduced [34] as the fraction of communication dedicated to alters when they are ranked in decreasing order of communication frequency. For the weighted ego-net of ego  $e$ , where  $w_{ei}$  denotes the link weight of the  $i$ th most contacted alter, the vector representing the social signature is defined as

$$s_e = (w_{e1}/\sum_{i=1}^k w_{ei}, \dots, w_{ek}/\sum_{i=1}^k w_{ei}), \quad (1)$$

where  $k$  is the degree (total number of alters) of the ego  $e$ . Social signatures have been shown to be persistent over time, implying that people’s networking patterns have distinct signatures [34]. Moreover, people’s call and SMS signatures have been shown to resemble each other, providing further evidence for the existence of distinct individual social signatures [11].

### 2.3 Measures of ego-net structural distance

People vary in their social network size, in how actively they use a channel for communication, and in how they distribute their communication resources among their alters. Here, we will review four different functions that have been used in the literature to measure the structural distances between ego-nets. These functions take tie strength heterogeneity into account in different ways. In the subsequent Sections (3.2 and S3), we will investigate whether the observed phenomena of ego-net persistence and across-channel similarity (reported in [11, 14, 21, 34]) hold true, irrespective of the extent to which the degrees of the ego-nets change over time (or differ across channels) as well as the distance function used for measuring the similarity.

#### *L<sup>2</sup> distance*

To measure the  $L^2$  distance (or Euclidean distance) between two social signatures, we first zero-pad the shorter signature, appending zeroes to its end in order to make sure that both signatures are of the same length. Then, the  $L^2$  distance between social signatures  $s_1$  and  $s_2$  is defined as

$$L^2(s_1, s_2) = \sqrt{\sum_{r=1}^k |f_{1r} - f_{2r}|^2}, \quad (2)$$

where  $f_{1r}$  is the fraction of communication that the alter of rank  $r$  in signature  $s_1$  receives.

#### *Jensen-Shannon distance*

To measure the Jensen-Shannon distance between two social signatures, we again first zero-pad the shorter one so that both of the signatures are of the same length. Then, the JSD distance  $d_{\text{JSD}}$  between social signatures  $s_1$  and  $s_2$  is defined as

$$d_{\text{JSD}}(s_1, s_2) = \left[ H\left(\frac{1}{2}s_1 + \frac{1}{2}s_2\right) - \frac{1}{2}(H(s_1) + H(s_2)) \right]^{\frac{1}{2}}, \quad (3)$$

where  $H(s_1)$ , the Shannon entropy of  $s_1$ , is

$$H(s_1) = - \sum_{r=1}^k f_{1r} \ln f_{1r}, \quad (4)$$

with  $k$  the maximum rank and  $f_{1r}$  the fraction of communication dedicated to alter of rank  $r$ . ( $f \ln f$  is taken as 0 for  $f = 0$ .)

#### *Gini Coefficient distance*

The Gini coefficient [9] is commonly used to measure the deviation of wealth distribution from perfect equality, with values ranging from 0 (indicating complete equality) to 1 (indicating complete inequality). In our study, we utilize the Gini coefficient to quantify the inequality in the distribution of communication among alters of an ego, similarly to [4].

If the link weights of an ego-net with degree  $k$  are sorted in ascending order ( $w_1 < w_2 < \dots < w_k$ ), then the Gini coefficient  $G$  is calculated as

$$G = \frac{2 \sum_{i=1}^k i w_i}{k \sum_{i=1}^k w_i} - \frac{k+1}{k}. \quad (5)$$

We define the metric distance between a pair of Gini coefficients  $G_1$  and  $G_2$  as

$$\Delta G(G_1, G_2) = |G_1 - G_2|. \quad (6)$$

#### *Alter-preferentiality distance*

The alter-preferentiality parameter  $\beta$  [14] is a measure defined for an ego-net growth model that incorporates a tuning parameter that determines the tendency of an ego to recurrently contact previously contacted alters, resulting in a “rich-gets-richer” phenomenon and leading to a heterogeneous ego-net, as opposed to choosing an alter uniformly at random resulting in a homogeneous ego-net. The range of  $\beta$  is from 0 to infinity, where  $\beta < 1$  corresponds to the homogeneous regime and  $\beta > 1$  corresponds to the heterogeneous regime.

To estimate the alter preferentiality  $\beta$  associated with a weighted ego-net, we utilize a maximum likelihood estimation method and a goodness-of-fit (GOF) test, as described in Ref. [14]. Our GOF test uses Kolmogorov-Smirnov statistic [29] with a p-value threshold of 0.1. Not all the fitted values pass the test, and therefore, the preferentiality value cannot be assigned to all the egos. The percentages of egos with valid fits are presented in Table S1. After fitting the  $\beta$  values, we filter out the outliers by excluding the largest five percentile of  $\beta$  values for all the analyses with the exception of Figure 5 for which because of the small size of the CNS dataset we skip this step.

Unlike the Gini coefficient, which has a bounded range,  $\beta$  has no upper bound. Therefore, we propose a normalized distance function for measuring the distance between a pair of alter-preferentiality parameters,  $\beta_1$  and  $\beta_2$ , as

$$\Delta\beta(\beta_1, \beta_2) = \frac{|\beta_1 - \beta_2|}{\beta_1 + \beta_2}. \quad (7)$$

Similarly to Eq. (6), this normalized distance is bounded between zero and one.

## 2.4 Persistence, revisited

Ego-nets are dynamic: people initiate new social relationships and lose contact with old acquaintances over time [31, 34]. Moreover, social ties usually display bursty patterns of activity: there are bursts of frequent contact that are separated by longer periods of silence [2, 5, 15, 26]. Despite the constant changes and turnover in the sets of alters, several studies have reported structural stability in the shape of ego-nets [11, 21, 34]. There are two main aspects to the shape of ego-nets: the size or degree (the number of alters) and the manner in which the limited communication resources are distributed among the alters. However, it is not entirely clear whether the ego-net persistence reported in the literature is merely due to stability in ego-net size or whether the ego’s pattern of tie strength heterogeneity is also persistent.

To answer this question, we must first define and quantify the persistence of ego-net structure. Persistence is commonly measured by comparing the temporal changes in the structures of individual ego-nets with the variation in population (e.g. the distribution of inter-individual distances) to determine if the changes are small or large (a method used in [6, 11, 14, 17, 21, 34]). In this study, we take a similar approach but with some modifications which enable us to disentangle the effects of degree stability and tie-heterogeneity persistence.

To measure the temporal change for each ego and communication channel, we divide the timeline into two consecutive windows. In this study, instead of having consecutive windows with equal duration, we divided the timeline so there are equal numbers of communication events in each window. Then, similarly to the previous studies, we construct the weighted ego-net associated with each window and measure the structural difference between the two ego-nets,  $d(ew_1, ew_2)$ , using the distance functions of our choice (distance functions are listed in Section 2.3).

To investigate the effect of the degree change, we calculate the persistence relative to two different reference distributions: a general reference with no constraint (which is the approach taken in the earlier studies [6, 11, 14, 17, 21, 34]) and a degree-stratified reference consisting exclusively of the ego pairs that have a degree difference equal to the degree change of the ego for whom we are calculating the persistence. In the first approach, the reference distribution comprises the ego-net distances between randomly selected egos  $x$  in the first period and randomly selected egos  $y$  in the second period, without paying attention to their degrees (with  $x \neq y$ ). In the second method, we form the reference distribution of a subset of these instances, namely the pairs  $(x, y)$  that satisfy the condition  $|k_{e1} - k_{e2}| = |k_{x1} - k_{y2}|$ , where  $k_{e1}$  and  $k_{e2}$  are the degrees of ego  $e$  in the different time windows, and  $k_{x1}$  and  $k_{y2}$  are the degrees of egos  $x$  and  $y$  in their respective time windows.

To quantify the structural persistence of the network of ego  $e$ , we calculate the  $z$ -score, a signed dimensionless quantity that measures the deviation of  $d(ew_1, ew_2)$  from the mean value of the reference distribution of choice, ref, in units of the standard deviation of ref:

$$z(d(ew_1, ew_2), \text{ref}) = \frac{d(ew_1, ew_2) - \text{avg}(\text{ref})}{\text{std}(\text{ref})}. \quad (8)$$

A negative value means that the temporal change for the ego is less than the average distance between a random ego in the first window and another random ego in the second window, and this non-randomness is taken as indicative of ego-net persistence.

### 3 Results

Earlier studies have produced some evidence of ego-net persistence and of their similarities across communication channels, as indicated by intra-individual distances being smaller than the average distances between random individuals [6, 11, 14, 17, 21, 34]. Here, we ask if the observed persistence is merely caused by the stability of ego-net size or if egos are also persistent in the manner in which they distribute their limited communication resources among their alters. We use the LMP dataset (see S1.1) which includes communication logs of over four million users for our investigations. This size facilitates computing degree-stratified reference distributions, as pairs of egos with a chosen degree difference can almost always be found.

First, in Section 3.1, we establish our motivation for probing the notion of ego-net stability by showing that all different distance functions introduced in Section 2.3 are to some extent sensitive to the degree differences of the ego-net pairs. We also show that as compared with the variation in the population, people tend to have stable degrees in time and, to a smaller extent, similar degrees across the call and SMS channels.

Having established the motivation for revisiting the persistence of ego-nets, in Section 3.2 we show that the structural persistence reported in the literature holds when measured as compared to references with similar degree differences and is not a mere side effect of degree stability.

Moreover, we observe that among the four structural distance functions used in the literature (see Section 2.3), the distance functions based on the Gini coefficient and the alter preferentiality  $\beta$  are less sensitive to network size differences. Thus, they are more appropriate measures for comparing ego-nets of different sizes if we wish to focus on tie strength heterogeneity.

Finally, armed with measures relatively insensitive to ego-net size, we investigate the similarity of networks of egos across call and SMS communication channels. For this purpose, we use the CNS and LMP datasets (see Sections S1.2 and S1.1 for data description). The comparison over these channels suggests that the heterogeneity levels of ego-nets are moderately correlated across call and SMS channels.

#### 3.1 Why do we need to disentangle the persistence of ego-net size and tie strength heterogeneity?

We are interested in measuring the persistence of tie strength heterogeneity of the ego-nets because we hypothesise that the distribution of tie strength is rather stable and possibly determined to a large extent by individual personality traits. To prove this, we need to compare ego-nets of the same individual across time and different channels. However, measures applied to quantify the difference between ego-nets may be sensitive not only to how tie strength is distributed in the ego-nets but also to their degrees. Ego-nets of similar degrees may appear similar purely because of the small degree difference. We wish to minimise this effect and we do so by investigating the sensitivity of the distance functions on degree variation in order to choose the less sensitive measures.

We empirically investigate how the distance measures (introduced in Section 2.3) depend on the degree difference of the ego-net pairs that are being compared. To this end, we take a random sample of approximately 7,000 egos from the LMP dataset and form all possible ordered ego pairs ( $\sim 7000 \times 6999$  pairs). For each ego pair  $(e_1, e_2)$ , we construct an ego-net for  $e_1$  based on the first half of its call log and an ego-net for  $e_2$  using the second half of  $e_2$ 's call log. Then we calculate the distance between these

pairs of ego-nets using each of the four distance functions defined in Section 2.3 (notice that these are the same ego-nets that we will use to form the reference distribution in Section 3.2).

To investigate how the measured distance depends on the degrees, we group the ego-net pairs based on their degrees  $(k_1, k_2)$  and calculate the average distance for each group and each distance function. Fig. 2 shows the average distances for each group as a function of the degree of the ego-nets in four heatmaps, one for each distance function. Bins with too few pairs (less than 200 pairs) are not shown. The plots reveal that for all distance functions, the measured distance tends to be smaller when the ego-nets have similar degrees. However, the effect is stronger for the two signature-based distance functions (JSD and  $L^2$  distances).

As the structural distance of ego-nets and degree difference appear correlated, the next natural step is to study the individual and population variations of degree. Fig. 3 shows complementary cumulative distributions of both the amount of individual degree change of ego-nets in time and the individual degree differences of ego-nets across channels, as compared to the degree differences of random ego-net pairs (the population variation). We observe that compared to the population variation, egos tend to have stable degrees in time and, to a smaller extent, similar degrees across the call and SMS channels.

The degree sensitivity of the distance functions, along with the degree stability of the egos, prompt us to revisit the notions of ego-net persistence and similarity and quantify how degree differences affect ego-net persistence as measured with different distance functions (see Sections 3.2 and S3).

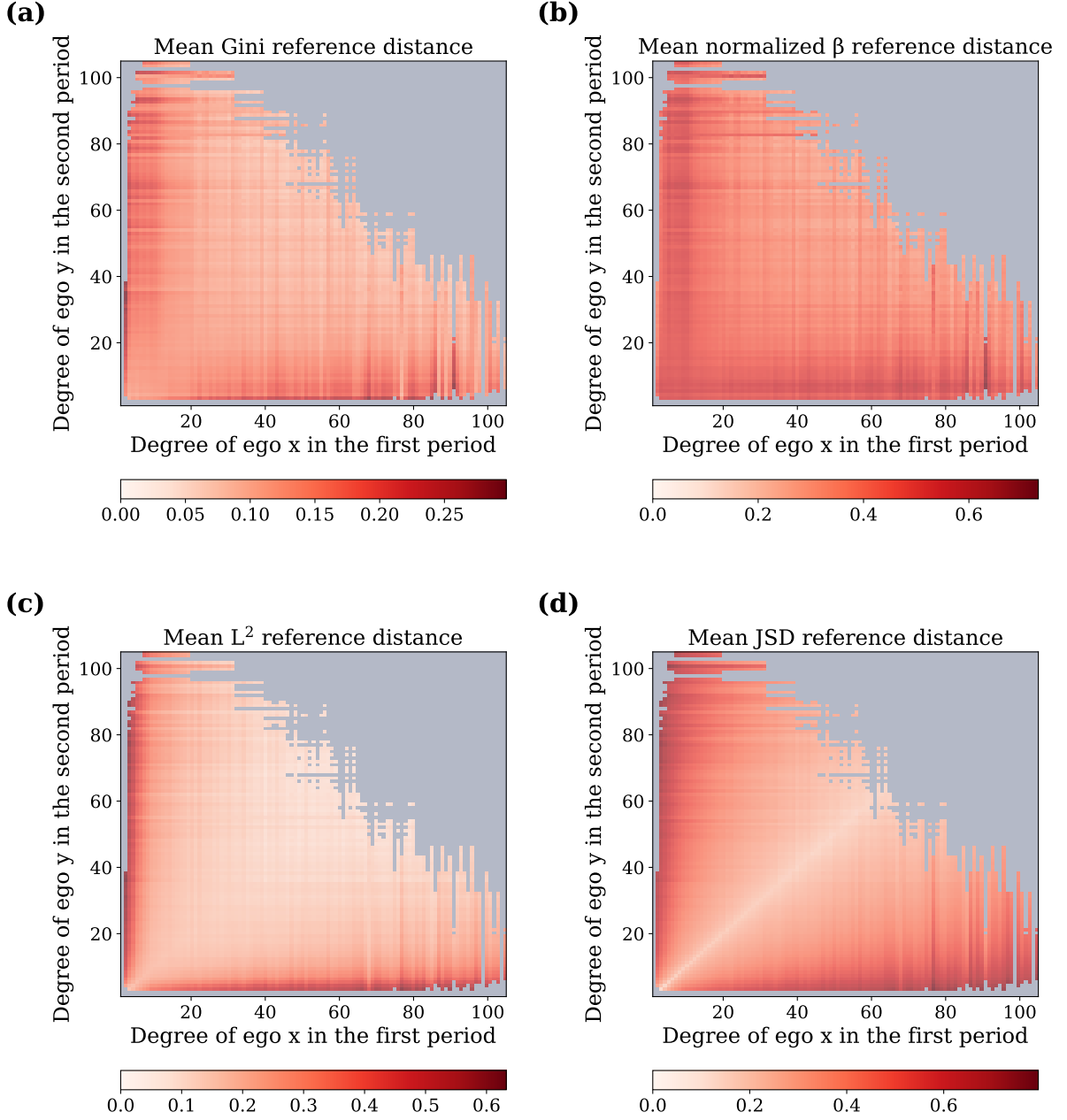
### 3.2 Impact of degree change on the structural persistence of ego-nets

Next, we wish to investigate the impact of degree changes on the persistence of ego-nets when using different functions to measure structural distance (see Section 2.3 for a description of different distance functions and Section 2.4 on how to quantify persistence). For this analysis, we utilize the call log data from the LMP dataset. The extensive size of the dataset enables us to draw a sufficiently large sample for establishing a representative reference distribution of distances in the population. To quantify persistence, we compare the self-distances that represent the structural change of each ego-net from one period to another against the reference distribution and calculate the corresponding  $z$ -scores [see Eq. (8)].

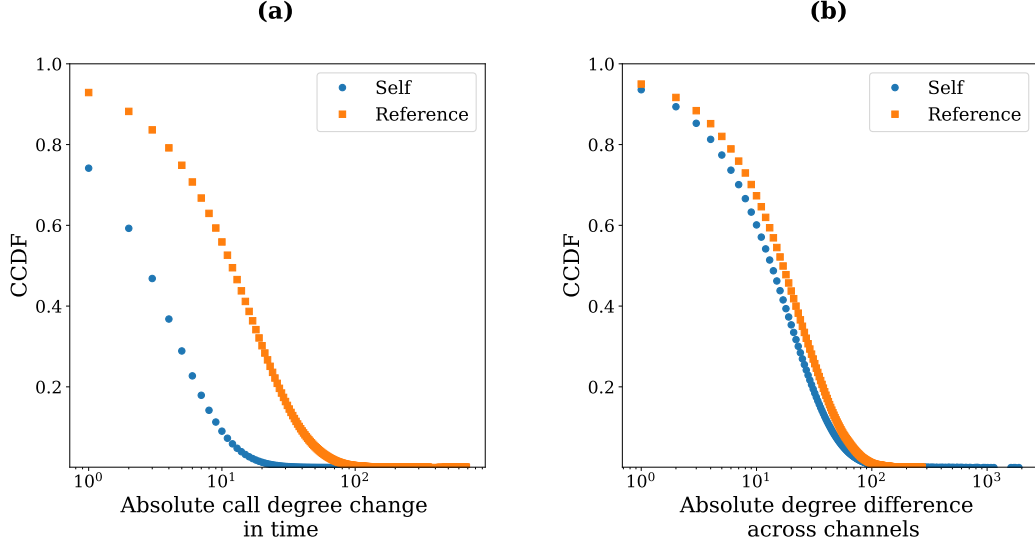
When calculating the self distances, we limit ourselves to the egos to whom we can assign both  $\beta$  and Gini heterogeneity values, i.e. those ego-nets that pass the goodness-of-fit test for  $\beta$  and have degrees larger than one so it is possible to calculate their Gini coefficient. This leaves us with around 3 million ego-nets for whom we calculate the structural distance between their ego-nets in two consecutive time windows, according to different measures.

To determine if the self-distances are smaller or larger than expected, as explained in Section 2.4, we use a general, degree-agnostic reference and  $|\Delta k|$ -specific references. The sample for making the degree-agnostic reference is the same as in 3.1 ( $\sim 49\text{M}$  pairs). As for  $|\Delta k|$ -specific references, we use a specific subset of the sample according to the degree change of the ego whose ego-net persistence we are assessing (See 2.4).

Having the self-distances and the reference distributions, using Eq. (8), we can measure the persistence score of each ego. Fig. 4 shows the distribution of persistence scores of call ego-nets among the population (around three million egos) in the LMP dataset. The persistence values are calculated using all four distance functions introduced in Section 2.3 and both against a single degree-agnostic reference [panel



**Figure 2. Average measured distance is smaller for pairs with similar degrees.** The distribution of average reference distances (distance between the network of a randomly selected ego,  $x$ , in the first period and the network of another randomly selected ego,  $y$ , in the second period) measured by different distance functions [(a) Gini distance, (b) normalised  $\beta$  distance, (c)  $L^2$  distance, and (d) Jensen-Shannon divergence distance]. For all the distance functions, the average reference distance tends to increase as the degree difference becomes larger. However, the effect is sharper for the JSD and  $L^2$  distance functions.



**Figure 3. Egos have stable degrees in time and to a smaller extent across call and text channels.** Left plot: The complementary cumulative distribution (CCDF) of degree change over time of call ego-nets in the large mobile phone dataset (blue circles) vs. the CCDF of degree difference of the random ego-net pairs (orange squares). The comparison of the two curves shows that egos have persistent call degrees as compared to the call degree variation in the population. **(right)** CCDF of the personal call and SMS degree difference of egos in the LMP dataset (blue circles) vs. the degree difference of the call and SMS ego-nets belonging to random pairs of egos (orange squares). The comparison shows that egos tend to have smaller differences between their call and SMS degrees as compared to the random pairs. However, the effect is weaker than the call degree persistence case.

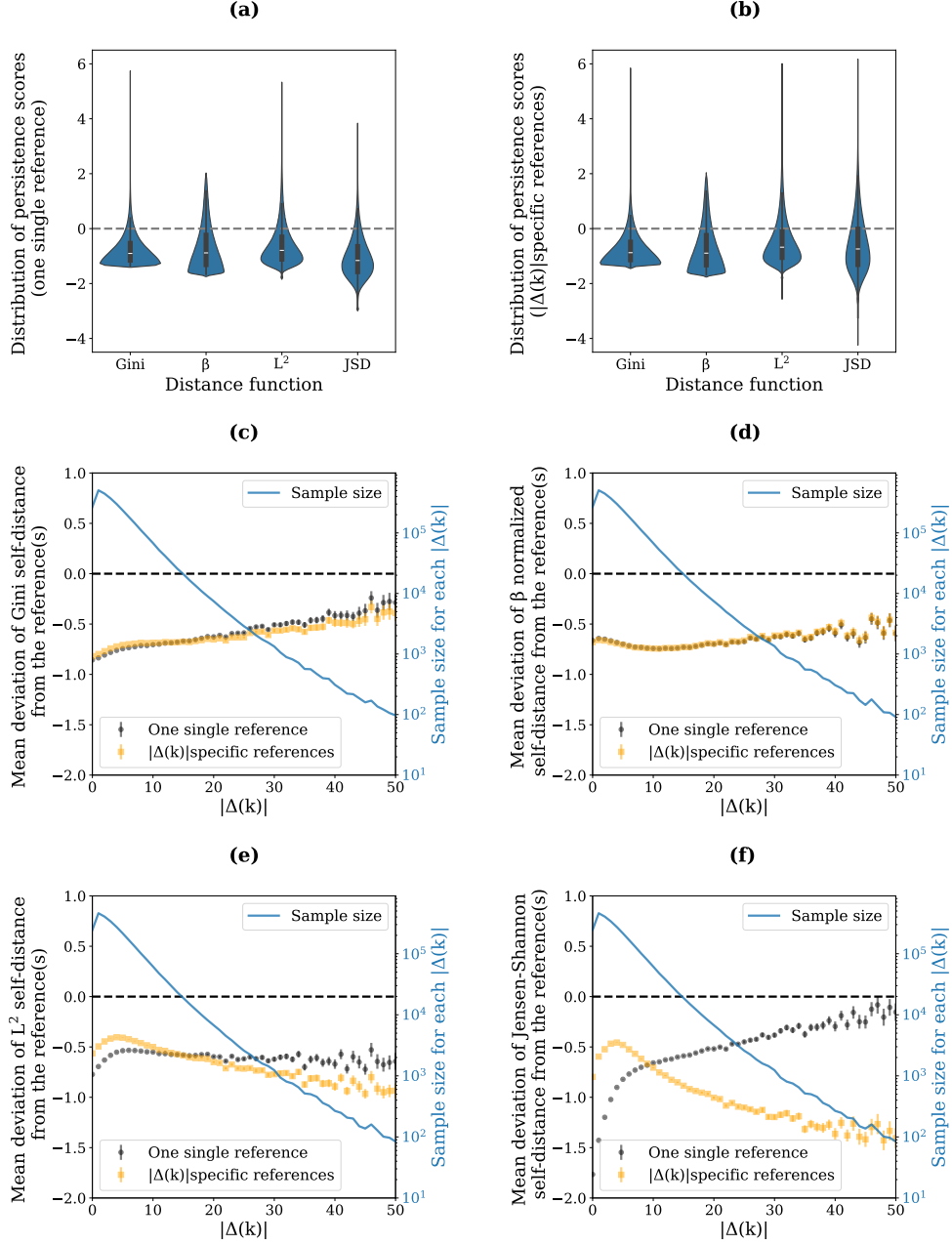
(a)] as well as  $|\Delta k|$ -specific references [panel (b)]. We observe that for all these cases, most egos have negative persistence scores, indicating the stability of the ego-nets.

So far, we have quantified the level of persistence whose existence was statistically observed in the previous studies ([6, 11, 21, 34]) and also shown that the reported persistence holds when controlling for different levels of degree change and thus it is not merely a side effect of degree stability. To examine the impact of degree change on the persistence of the ego-nets, we take a closer look at the persistence scores by grouping the egos based on the absolute value of their degree change, denoted as  $|\Delta(k)|$ , and calculate the average persistence for each group separately (see Fig. 4). We observe that the average of the persistence scores measured against both a single reference (the grey scatter plots) and  $|\Delta k|$ -specific references (the orange scatter plots) remains relatively constant as a function of  $|\Delta(k)|$  for the normalized  $\beta$  distance and  $L^2$  distances, whereas a clear dependence on  $|\Delta(k)|$  can be observed for the JSD distance.

Moreover, we observe that when switching from one single reference to  $\Delta k$ -specific references, the persistence scores for the  $L^2$  and JSD distances change more significantly as compared to the  $\beta$  and Gini persistence scores (see the orange scatterplots in Fig. 4 in comparison to the grey scatterplots). This result indicates that the alter preferentiality  $\beta$  and the inequality measure  $G$  are both fairly insensitive to degree difference and, thus, can be considered more suitable for comparing ego-nets with varying sizes.

### 3.3 Application: similarity of ego-nets across call and SMS channels using the Gini coefficient and the alter preferentiality $\beta$

Finally, we apply our understanding of the proper measures for ego-net comparison to analyse the cross-channel similarity of ego-nets. This is motivated by the fact that an ego-net reconstructed from data on



**Figure 4. Structural persistence of call ego-nets in the LMP dataset using different functions for measuring structural distance.** (a) Distribution of persistence scores in the LMP dataset when measured against one single reference and in (b) when measured against  $|\Delta(k)|$ -specific references. In both cases, most of the scores are negative, which indicates the stability of call ego-nets. Average persistence as a function of  $|\Delta(k)|$  (absolute degree change) is shown in (c) using the Gini distance function; in (d) using the normalised  $\beta$  distance; in (e) using the  $L^2$  distance; and in (f) using the JSD distance. The average persistence scores (the deviation of self-distances from reference distances) are visualised in grey if they are measured against one single reference distribution (regardless of  $|\Delta(k)|$ ) and in orange if against  $|\Delta(k)|$ -specific reference distributions. See Eq. 8 for the definition of the persistence score. All four heterogeneity measures have negative average values across the degree range. Moreover, the persistence plots for the Gini coefficient and  $\beta$  do not change much when switching from a single reference to  $|\Delta(k)|$ -specific references. This suggests that Gini and  $\beta$  are better measures for comparing ego-nets of different sizes. The blue curves in plots (a), (b), (c), and (d) show the number of egos with a particular absolute degree change ( $|\Delta(k)|$ ) from the first period to the second period (values are on the right axes).

communication on one single channel only represents an incomplete reflection of the underlying social network. Then, the accuracy and completeness of such reflections are determined by how exclusively and actively the ego uses the communication channel and what portion of observed events are related to actual social relationships (e.g., calling a service line or calling a wrong number are not).

We start by using the LMP dataset to assess the similarity of ego-nets across channels with an approach that controls for different levels of degree difference (similar to our approach in analyzing the persistence of ego-nets in Section 3.2). We conclude that call and SMS ego-nets resemble each other, but the effect is smaller than the observed persistence of ego-nets in time (see Section S3 for details).

Next, as the Gini coefficient and the alter preferentiality  $\beta$  do not require the use of reference distributions, we directly correlate the Gini and  $\beta$  values for call and SMS based ego-nets, for both of our datasets (LMP and CNS), as shown in Fig. 5. We observe moderate correlations across the modalities, indicating that there are clear similarities between an individual’s tie strength heterogeneity across the channels.

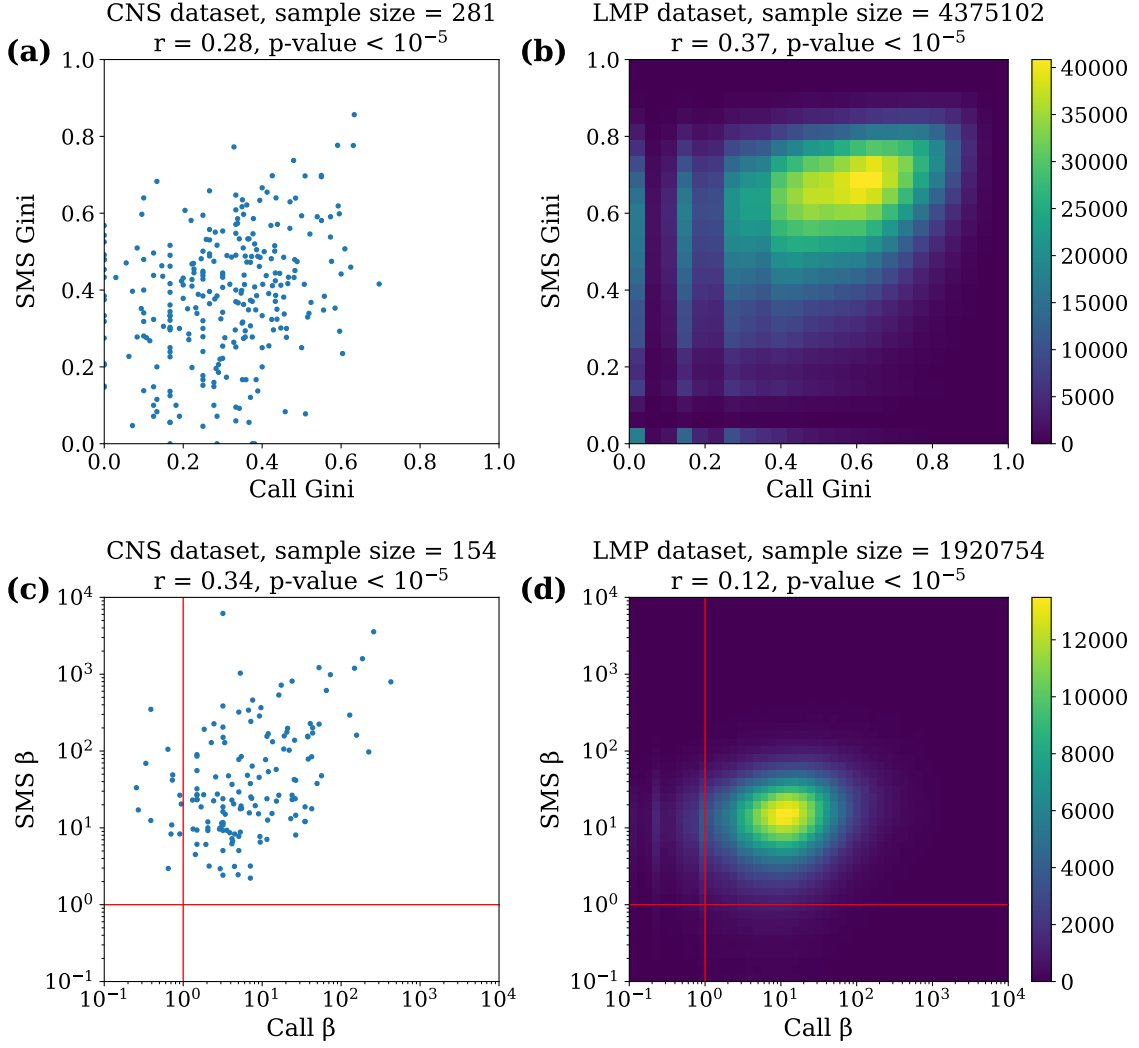
Taken together, the above results support the hypothesis of a latent variable, such as a personality trait, giving rise to similar ego-net shapes for an individual across different channels. However, as the correlation is only moderate, this latent variable might not be the only determinant of tie strength heterogeneity in an ego-net.

### 3.4 Discussion

Egocentric networks are the building blocks of societal-level networks. They comprise a wide variety of relationships ranging from romantic partners and family to friends, colleagues, and acquaintances. This diversity of relationship types is reflected in the structure of ego-nets: not all social ties are equally strong, but tie strength is typically heterogeneous with a few strong and a larger number of weaker ties. This characteristic pattern has been observed in social networks inferred from various communication datasets [11,14,25,34]. Within this broad pattern, there is, however, individual-level variation, as people express different levels of heterogeneity in their personal networks. Besides tie strength heterogeneity, ego-nets vary in size, i.e., network degrees (see, e.g., [28]). To some extent, the degree of an ego-net reflects the size of the ego’s underlying social circle, but it is also affected by empirical issues such as sampling (the communication channel used, and the length of the observation window [20,37]).

The heterogeneity of tie strength appears to be a universal characteristic of ego-nets [14]: its distribution is similar in ego-nets derived from multiple communication channels. However, as the datasets in Ref. [14] do not comprise the same individuals, it is not known what causes this similarity. On one hand, the similarity may be a population-level consequence of people having different channel preferences and behaving differently on different channels so that the mixture of these different behaviours is behind the distribution of heterogeneity. Alternatively, it may arise from latent-individual level variables, such as personality traits, which determine people’s ego-net structure on any channel.

Both ego-net size and tie strengths are encapsulated in so-called social signatures [34] that quantify the share of communication that alters receive as a function of their rank in an ego-net. These signatures are rather persistent for an individual from one time period to another, even when the combination of alters of the ego changes [34]. An individual’s signatures also resemble each other across communication channels [11]. However, until now, it has been unclear whether the persistence is due to degree stability alone or tie strength heterogeneity persists as well. In this paper, we carried out an analysis taking into account the sensitivity of distance functions used for comparing ego-nets on their degree differences. We



**Figure 5. Call and SMS ego-nets are correlated in their heterogeneity levels.** Gini values associated with tie-strength heterogeneity of SMS ego-nets of the same individuals are plotted as a function of their call Gini coefficient in (a) the CNS dataset and in (b) the LMP dataset (colours in the heatmap indicate the number of egos falling into each bin regarding their call and SMS Gini values). Panels (c) and (d) show similar plots for the alter parameter ( $\beta$ ). The red lines indicate  $\beta = 1$ , the border between homogeneous and heterogeneous regimes. We observe that most of the egos fall into the heterogeneous regime. The positive Pearson correlation values in all four plots show that SMS and call heterogeneity values are correlated.

showed that the observed persistence of social signatures and their individual-level similarity across the call and SMS channels are not merely a side effect due to the stability of the network size. While the effect is somewhat smaller, it is still statistically significant when controlling for degree differences. Our results indicate that although part of the persistence of the signatures of individuals indeed stems from the stability of the size of the ego-nets over time and channels, there is a considerable contribution to this persistence related to other individual features, possibly to personality traits.

Our comparison of the distance functions also revealed that two measures are far less sensitive to degree differences: the alter preferentiality  $\beta$  and the Gini coefficient. These two are thus better options for assessing the persistence of tie strength heterogeneity when network sizes are also heterogeneous and for understanding whether the same individuals have similar ego-net shapes on different communication channels, where network sizes may generally vary quite extensively. These measures are highly correlated and theoretically related (see S2.1), and therefore, either of them can be used. The Gini coefficient is a well-known statistical indicator and is also straightforward to compute. The Gini coefficient can be calculated for any ego-net with a degree larger than one (which is also a prerequisite for any tie strength heterogeneity to be possible). On the other hand, the alter-preferentiality parameter  $\beta$  has more explanatory power since it is associated with an ego-network growth model but is obtained via a rather cumbersome fitting procedure. Moreover, the fitting procedure automatically filters out ego-nets of too low degree and entirely homogeneous networks (see Fig. S3).

Another advantage of both coefficients is that they attribute a heterogeneity value to each ego-net, facilitating a more straightforward comparison than when using signature-based distance functions (e.g. JSD and  $L^2$  distances). This is particularly relevant for comparing ego-nets across different communication modalities and for testing the latent-variable hypothesis. While some previous studies have examined the similarities and differences of ego-nets across different communication channels, the results have not been conclusive. Multichannel studies have shown that examining a single channel provides only a partial description of the ego’s interactions, and a holistic approach is needed to gain a more realistic picture [11, 37, 42]. A study comparing call and SMS ego-nets pointed out that the composition and rankings of the alters in these two channels can be very different [11], even though the ego-nets of one individual are similarly shaped. Another study [1] compared temporal communication patterns of calling and texting and reported significant differences, which may indicate their different functionalities.

To this end, we applied the Gini and  $\beta$  measures to investigate if the call and SMS ego-nets of an individual are correlated in terms of tie strength heterogeneity. We used two datasets that contain both calls and SMS messages and observed that the tie strength heterogeneity measures show a moderate correlation for call and SMS ego-nets, which is in line with the latent-variable/personality-trait hypothesis, where an individual tends to maintain similar networks on any media. However, the moderate level of correlation also leaves room for population-level explanations for the observed universality in heterogeneity distributions, as the hypothetical latent variable cannot be the sole determinant of tie strength heterogeneity in an ego-net. This points out that to properly understand the reasons for tie strength heterogeneity across different communication channels, there is a need for data involving the same subjects interacting on even more channels—one might also envision experimental setups where people repeatedly build and maintain networks on online channels and the tie strength heterogeneity in the resulting networks is investigated in detail.

## 4 Declaration

### 4.1 Availability of data and materials

The CNS data analysed in this paper is publicly available via *figshare* [32, 33]. While the LMP data is not publicly available and shared with us via a non-disclosure agreement, it has been extensively studied in the literature (see, for example, [11, 15, 16, 19, 27, 28, 38]).

### 4.2 Competing interests

The authors declare that they have no competing interests.

### 4.3 Funding

This study was part of the NetResilience consortium funded by the Strategic Research Council at the Academy of Finland (grant numbers 345188 and 345183). G.I. and J.K. acknowledge support from AFOSR (Grant No. FA8655-20-1-7020), project EU H2020 Humane AI-net (Grant No. 952026), and CHIST-ERA project SAI (Grant No. FWF I 5205-N). J.K. acknowledges support from European Union’s Horizon 2020 research and innovation programme under grant agreement ERC No 810115 - DYNASNET.

### 4.4 Author’s Contributions

S.H., G.I., J.K., and J.S. conceived, designed, and developed the study. S.H. analysed the empirical data. G.I. derived the analytical relations in Section S2.2. S.H., G.I., J.K., and J.S. wrote the paper.

### 4.5 Acknowledgments

We acknowledge the computational resources provided by the Aalto Science-IT project.

### 4.6 Abbreviations

**Ego-net:** egocentric network, **SMS:** a type of text message transmitted via Short Messaging Service, **LMP:** Large Mobile Phone, **CNS:** Copenhagen Networks Study, **SI:** Supplementary Information, **JSD:** Jensen-Shannon Divergence, **GOF:** goodness-of-fit, **CCDF:** Complementary cumulative distribution function.

## S1 Communication data

### S1.1 The Large Mobile Phone (LMP) dataset

The Large Mobile Phone (LMP) dataset consists of time-stamped logs of outgoing communication of approximately 20% of the population of an undisclosed European country. The user IDs are anonymised, and the logs span six months in 2007. The dataset was initially introduced in [27]. To ensure data quality, we have excluded self-communication events and records made by IDs listed under “family contracts” with the operator company, which could involve multiple users sharing the same phone line. This leaves us with more than 5 million users, 1.3 billion calls, and 613 million short messages (see Table S1 for exact numbers). It is important to note that the data is directional, and we only use the outgoing communication events when constructing ego-nets, which is unlike our approach for the CNS dataset, where we also include incoming call/SMS to the ego-nets (see section S1.2). This is because the LMP dataset includes two types of individuals: company-users, for whom we have outgoing communication logs and construct the ego-nets, and non-company users, who only appear in the data if a company-user contacts them. Therefore, if all events are considered regardless of direction, the ties to non-company alters would systematically have lower weights as we lack information on outgoing communication events of those alters. While the data is not publicly available, it has been extensively studied in the literature (see, for example, [11, 15, 16, 19, 27, 28, 38]).

### S1.2 Copenhagen Networks Study (CNS) dataset

The Copenhagen Networks Study (CNS) [33, 36] is a multichannel dataset collected via mobile phone devices distributed among around one thousand volunteer university students in 2012-2013. Data includes communication logs of calls and text messages (SMS) exchanged between the students as well as data on their physical proximity (captured by Bluetooth sensors) [36]. Here, we use the portion of mobile phone communication data that is publicly available as described in [33], e.g. one month’s worth of call and SMS logs between anonymised users. Data is publicly available via *figshare* in [32].

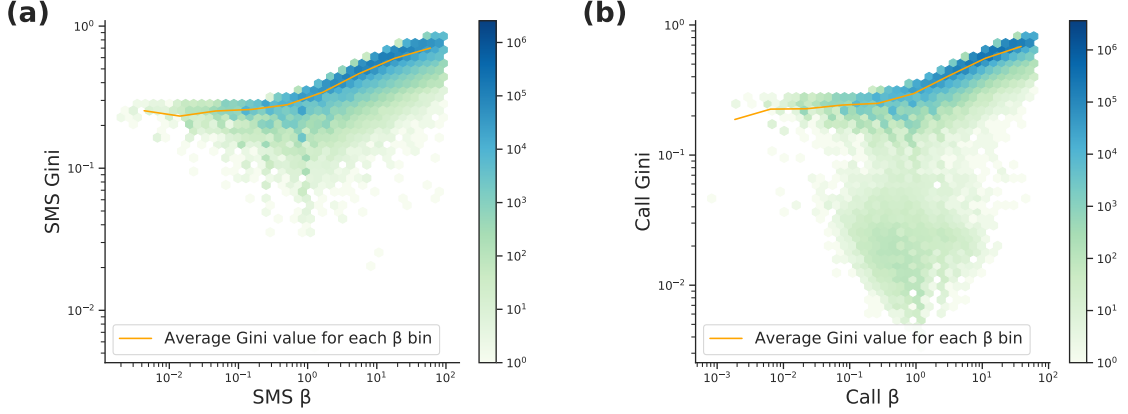
The call and SMS logs include only the events between the subjects in the study. We disregard the missed calls when making the ego-nets. Table S1 shows the number of remaining egos and events in each channel. We define indirectional call and SMS ego-nets by adding a call or SMS between  $x$  to  $y$  to both  $x$ ’s and  $y$ ’s ego-net, regardless of who has initiated it. We define the tie strengths simply as the number of events of the corresponding type between the ego-alter pairs.

| Dataset | Communication channel | Number of egos | Number of events | % of egos with valid $\beta$ |
|---------|-----------------------|----------------|------------------|------------------------------|
| LMP     | Phone calls           | 5994967        | 1342862618       | 64%                          |
| LMP     | Short messages        | 5387745        | 613751054        | 50%                          |
| CNS     | Phone calls           | 525            | 3234             | 37%                          |
| CNS     | Short messages (SMS)  | 568            | 24333            | 53%                          |

**Table S1.** A summary of the datasets used in this study. The large mobile phone (LMP) dataset is used in sections 3.1, 3.2, 3.3, S2, and S3, while the Copenhagen Network (CNS) study dataset is only used in section 3.3.

## S2 Gini and $\beta$ : The less sensitive heterogeneity measures

In Section 2.4, we observed that the Gini coefficient and the alter preferentiality  $\beta$  are better choices for comparing ego-nets of different sizes. Here, we take a closer look at the two measures and compare them empirically and theoretically. We observe that these two measures are closely related both according to their mathematical definitions and also based on the empirical data.



**Figure S1. Gini and  $\beta$  are heavily correlated.** The colour intensity indicates the number of egos which fall into each bin regarding their  $\beta$  and Gini coefficients. The ego-nets are made based on the call and SMS events in the LMP dataset (dataset described in SI Section S1.2). We observe that these two variables are highly correlated for both call and SMS ego-nets, with a Pearson correlation coefficient equal to 0.84 for the call ego-nets (sample size = 3626453 and  $p < 10^{-5}$ ), and 0.8 for SMS ego-nets (sample size = 2553079 and  $p < 10^{-5}$ ). We have excluded outlier  $\beta$  values (the top five percentile).

## S2.1 The empirical correlation

To compare the heterogeneity values measured by the Gini coefficient and the alter preferentiality  $\beta$ , we check the Pearson correlation coefficient between the measured heterogeneity values of the same egos in the LMP dataset for both call and SMS channels. Correlation coefficients of  $\sim 0.8$  show that these two values are strongly correlated, and the large sample size of a few million individuals and  $p$ -values smaller than  $10^5$  show the statistical significance of the observed correlations (See Fig. S1 demonstrating the correlation between these two measures for both call and SMS ego-nets in the LMP dataset).

## S2.2 The theoretical relation

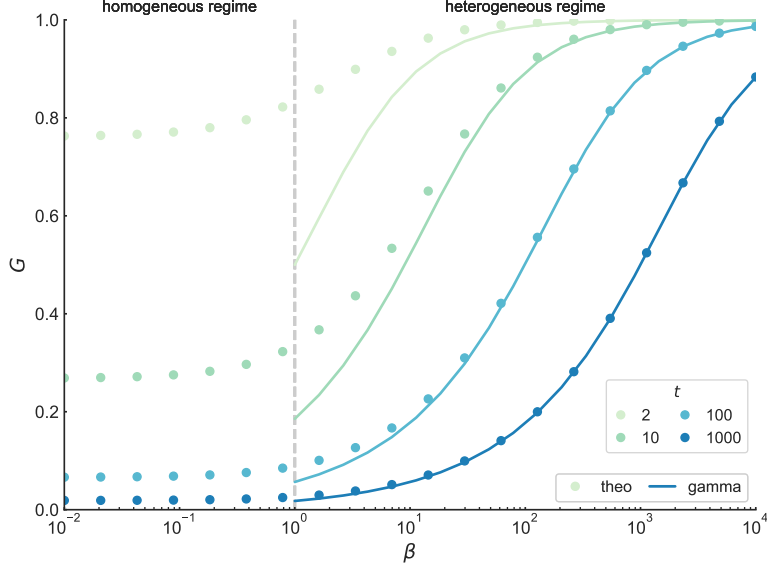
In the minimal model of ego-net dynamics introduced in Ref. [14], individuals allocate interactions via cumulative advantage and a tunable amount of random choice. At an initial time  $\tau_0 = k a_0$  with  $k$  the degree of the ego-net, all alters have minimal activity  $a_0$ . At any time  $\tau \geq \tau_0$ , the probability that an alter with activity  $a$  becomes active at time  $\tau + 1$  (and thus  $a \mapsto a + 1$ ) is

$$\pi_a = \frac{a_r/t_r + \beta^{-1}}{k(1 + \beta^{-1})}, \quad (9)$$

with  $a_r = a - a_0$ ,  $t_r = t - a_0$ , and  $t = \tau/k$  the mean alter activity. The alter preferentiality  $\beta = t_r/\alpha_r$  (with  $\alpha_r = \alpha + a_0$  and  $\alpha$  a tunable parameter of the model) interpolates between two regimes: random alter choice ( $\beta \rightarrow 0$  and  $\pi_a \sim 1/k$ ), and preferential alter selection ( $\beta \rightarrow \infty$  and  $\pi_a \sim a_r/\tau_r$  with  $\tau_r = \tau - \tau_0$ ). An analytical treatment of the model leads to the following expression for the activity distribution  $p_a(t)$  (the probability that a randomly chosen alter has activity  $a$ ),

$$p_a(t) = p_0 \frac{a_r^{-1}}{B(a_r, \alpha_r)} \left(1 + \frac{1}{\beta}\right)^{-a_r}, \quad (10)$$

where  $p_0 = (1 + \beta)^{-\alpha_r}$  and  $B(a_r, \alpha_r)$  is the Euler beta function [14].



**Figure S2. Theoretical relation between Gini coefficient and alter preferentiality.** Gini coefficient  $G$  [Eq. (5)] as a function of preferentiality  $\beta$  in the alter activity model of Ref. [14]. Curves shown for several values of mean alter activity  $t$ , both in terms of the cumulative of the alter activity distribution  $p_a$  [Eq. (11), theo], and via an approximation valid in the heterogeneous regime of  $\beta \rightarrow \infty$  [Eq. (13), gamma]. Gini increases with preferentiality, indicating how alter activities become progressively more heterogeneous (i.e. unequal).

The Gini coefficient  $G$  of Eq. (5) can be written in terms of the cumulative distribution function of alter activity,  $P_a(t) = \sum_{a'=a_0}^a p_{a'}(t)$ , by approximating  $a$  as a continuous variable, writing the Lorenz curve of  $P_a$  as an integral, and discretizing  $a$  once more,

$$G = 1 - \frac{1}{t} \sum_{a=a_0}^{a_m} (1 - P_a)^2, \quad (11)$$

where  $p_a$  follows Eq. (10). In the heterogeneous regime of preferential alter selection, the activity distribution approaches a gamma distribution with shape  $\alpha_r$  and scale  $\beta$  [14],

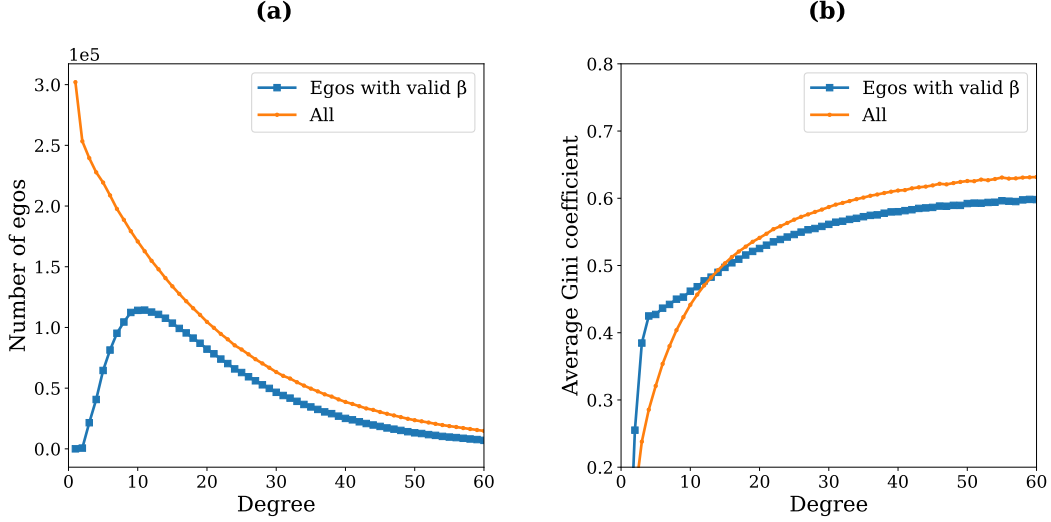
$$p_a(t) = \frac{1}{\beta^{\alpha_r} \Gamma(\alpha_r)} a_r^{\alpha_r-1} e^{-a_r/\beta}, \quad \beta \rightarrow \infty, \quad (12)$$

meaning that the Gini coefficient can be approximated as [24]

$$G = \frac{\Gamma(t_r/\beta + 1/2)}{\sqrt{\pi} \Gamma(t_r/\beta + 1)}, \quad \beta \rightarrow \infty, \quad (13)$$

with  $\Gamma$  the gamma function.

In the ego-net model of [14], and following Eqs. (11)-(13), the Gini coefficient of alter activity increases monotonically with preferentiality (Fig. S2). In the homogeneous regime of random alter choice ( $\beta \rightarrow 0$ ),  $G$  is low, in consistency with a homogeneous (Poissonian) activity distribution. As  $\beta$  increases and preferential alter selection becomes dominant, Gini approaches the value  $G = 1$  corresponding to a heterogeneous activity distribution and complete inequality (one alter dominates all activity). The approximation of Eq. (13) works well in the heterogeneous regime, but progressively fails as  $\beta$  becomes small, depending on the value of the mean alter activity  $t$ .



**Figure S3. The Gini coefficient can be calculated for any ego-net with two or more alters, while not all ego-nets can be assigned an alter preferentiality  $\beta$ .** (left) Total number of call ego-nets in the LMP dataset for each degree is shown by orange dots. Blue squares show the number of egos for which we can fit a  $\beta$ . The parameter  $\beta$  can be fitted only for a small fraction of low-degree egos. (right) Gini coefficient averaged over egos of a specific degree as a function of their degrees. The average values over all the egos and only the egos with valid  $\beta$  are shown respectively in orange and blue.

### S2.3 Pros and cons of each measure

In the previous Sections (S2.1 and S2.2), we observe that the Gini coefficient and alter preferentiality  $\beta$  are closely related. Here, we review their differences and discuss the pros and cons of using each measure.

The Gini coefficient can be calculated for any ego-net with two or more alters. In contrast, we might not necessarily find a valid alter preferentiality  $\beta$  for all ego-nets. For instance, the model used for obtaining  $\beta$  is not well defined for fully homogeneous networks where all alters have exactly the same communication amount (equivalent to a zero Gini coefficient). This prerequisite filters out those low-activity egos whose ego-nets consist of a few once-contacted alters. Fig. S3 shows the number of call ego-nets in the LMP dataset for which we can fit  $\beta$  as a function of their degrees. We observe that most of the failed cases are ego-nets with low degrees.

## S3 Measuring across-channel similarity using $\Delta k$ -specific references

The LMP dataset used in our study is also used in [11], where the authors showed that call and SMS social signatures are not only persistent but also resemble each other for each ego. The study used Jensen-Shannon as a distance function. Here, we replicate the same study by using three other distance functions besides JSD distance and also by controlling for different levels of degree change (See Section 2.3 for the definition of the distance functions). We want to check if the similarity of call and SMS ego-nets observed in [11] is more than a side effect of people having similar degrees across these channels as compared with the population degree variation (See Fig. 3 for evidence on the degree similarity).

To measure the self-distance for ego  $e$ , we construct a weighted ego-net using the call log during the whole observation window (denoted as  $e_c$ ) and another using the SMS log (denoted as  $e_s$ ) and use the

distance functions introduced in 2.3 to measure the heterogeneity difference of ego  $e$  across the channels denoted as  $d(e_c, e_s)$ .

Then, we calculate the across-channel similarity values both using a single degree-agnostic reference and  $|\Delta(k)|$ -specific references. To construct the degree-agnostic reference, we choose a random sample of  $\sim 7000$  egos and calculate the heterogeneity distance between the call ego-net of ego  $x$  and the SMS ego-net of ego  $y$  ( $y \neq x$ ) for all the possible pairs in the sample. To form  $|\Delta(k)|$ -specific reference, we take the subset of the sampled pairs which satisfy  $|k_{e,c} - k_{e,s}| = |k_{x,c} - k_{x,s}|$ , where  $k_{e,c}$  and  $k_{e,s}$  are the degrees of ego  $e$  in call and SMS communication channels, and  $k_{x,c}$  and  $k_{y,s}$  are the degrees of egos  $x$  and  $y$  respectively in channels  $c$  and  $S$ . We define the across-channel similarity for ego  $e$  by a signed dimensionless quantity defined similarly to the measure of persistence in Section 2.4

$$z(d(e_c, e_s), \text{ref}) = \frac{d(e_c, e_s) - \text{avg}(\text{ref})}{\text{std}(\text{ref})}, \quad (14)$$

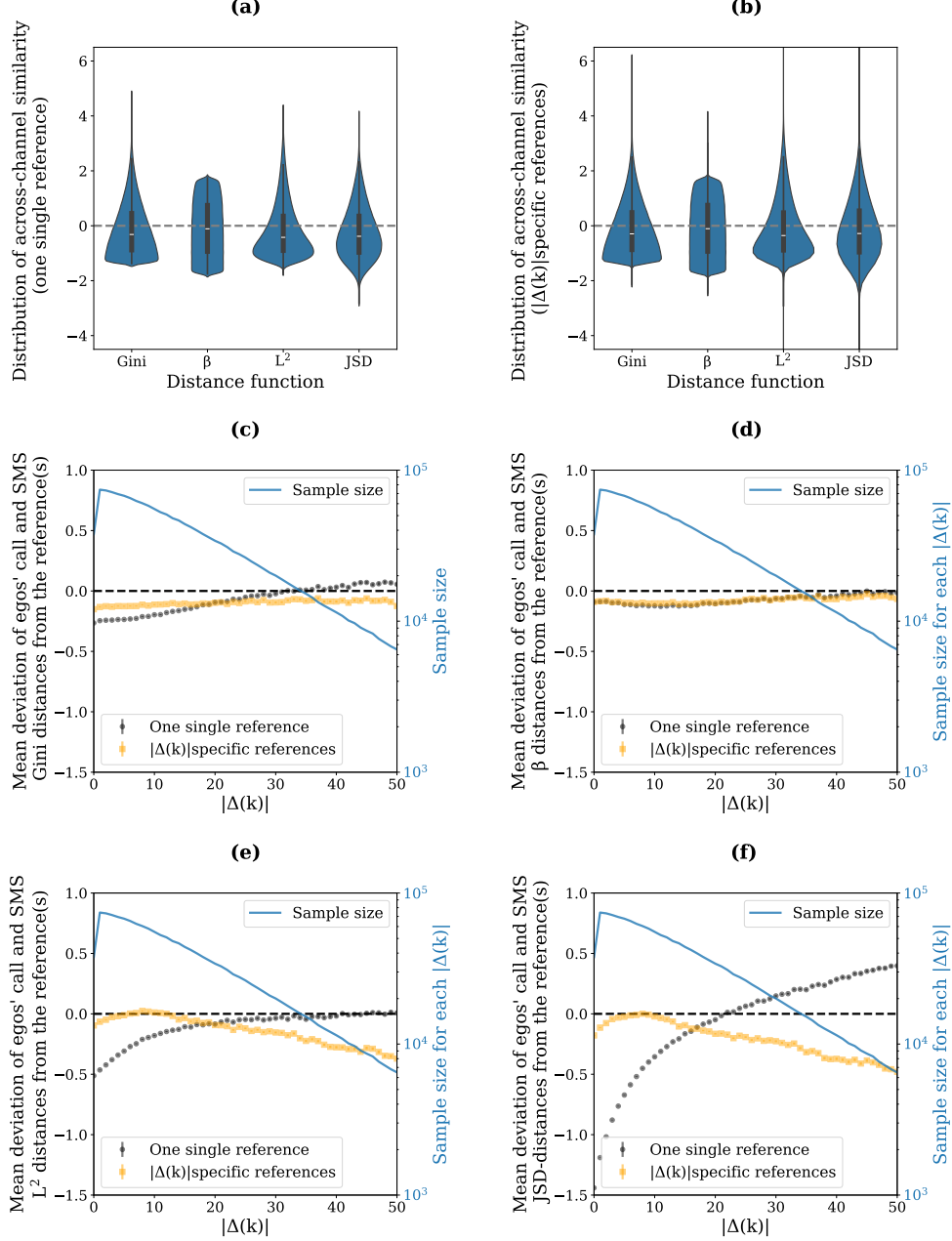
where  $\text{avg}(\text{ref})$  and  $\text{std}(\text{ref})$  denote the average and the standard deviation of the reference distribution of choice. Thus, across-channel similarity is measured as the deviation of  $d(e_c, e_s)$  from the mean value of the ref distribution in units of standard deviation of ref. A negative value indicates that the difference in the ego-net of ego  $e$  across two channels is less than the average distance between the network of a random ego in channel  $c$  and another random ego in channel  $s$ .

For all four distance functions, the across-channel similarity observation is confirmed when controlling for different levels of degree change (by using  $|\Delta(k)|$ -specific references). The average across-channel similarity values are negative across the  $|\Delta(k)|$  range (See Fig. S4). However, the deviations suggesting across-channel similarity are smaller than the persistence deviations (shown in Section 3.2), indicating a weaker effect. Similarly to the case for persistence measures, the across-channel similarity plots for the Gini coefficient and  $\beta$  do not change much when switching from the degree-agnostic reference to  $|\Delta(k)|$ -specific references. This is another evidence that  $\beta$  and Gini are better measures for comparing ego-nets of different sizes.

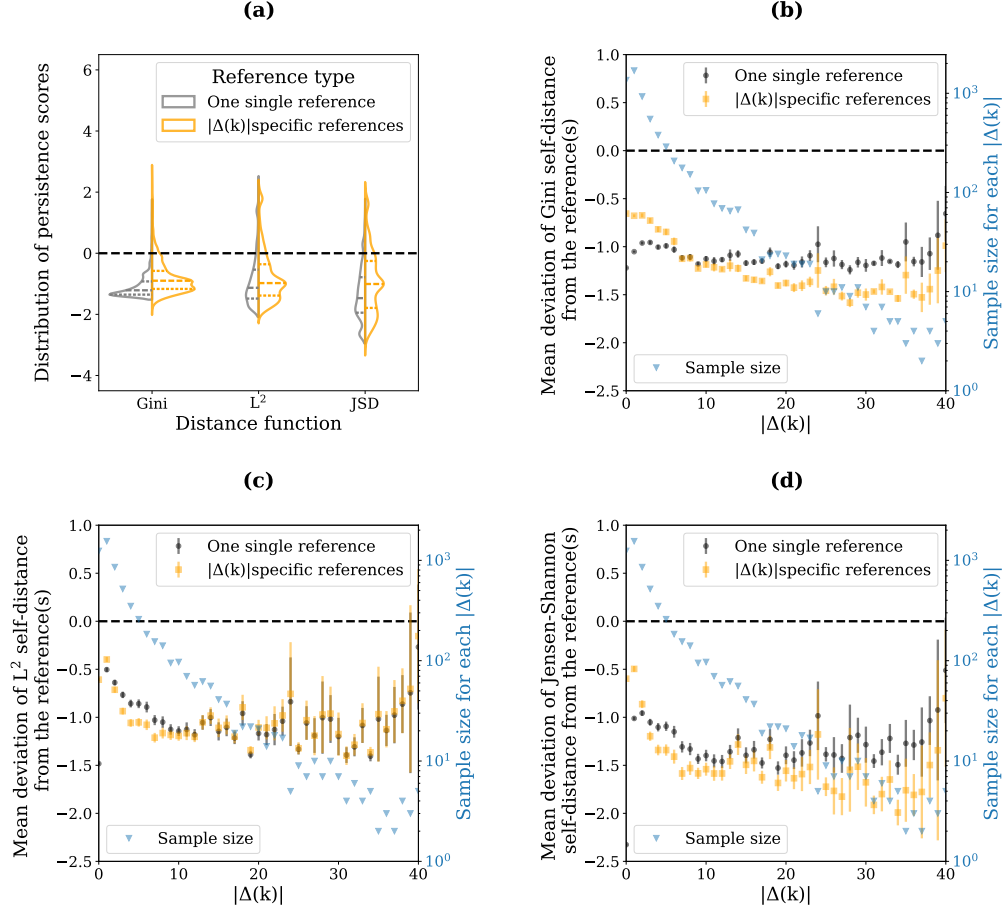
## S4 Are only ego-nets with valid alter-preferentiality values persistent?

As shown in Figure S3, the alter-preferentiality parameter  $\beta$ , which is associated with the ego-net growth model introduced in [14], cannot be assigned to all the ego-nets. Particularly, the goodness-of-fit test fails for a large fraction of low-degree ego-nets. The persistence plots in Figure 4 only include the ego-nets that pass the test, because one of the objectives of our study is to compare the distance function based on  $\beta$  with the other heterogeneity distance functions.

This raises a question: Are only ego-nets with valid  $\beta$ s persistent? To answer this, we repeat the study using the other three distance functions for the egos that can be assigned Gini values to their call ego-nets (degree larger than one in both periods) but not valid  $\beta$  (at least in one of the periods). There are 2708055 egos that fulfil this criterion in the LMP dataset. We observe that the ego-nets are persistent for these egos as compared to both the degree-agnostic and  $|\Delta(k)|$ -specific references (See Figure S5).



**Figure S4. Call and SMS ego-net similarity for the egos in the LMP dataset using different measures and distance functions.** (a) Distribution of similarity scores with a single reference. (b) Distribution of similarity scores with  $|\Delta(k)|$  specific references. In both cases and for all the measures, the majority of the scores are negative, indicating across-channel similarity. In panels (c) to (f) we see average call-SMS ego-net similarity scores as a function of  $|\Delta(k)|$  (absolute degree difference of call and SMS ego-nets) using Gini index in (c); alter preferentiality  $\beta$  in (d);  $L^2$  distance in (e); and Jensen-Shannon distance in (f). The average across-channel similarity scores are visualized in black if they are measured with respect to one single reference distribution (regardless of  $|\Delta(k)|$ ) and in orange if  $|\Delta(k)|$ -specific reference distributions are used (see Eq. 14 for definition). We see that all four heterogeneity measures suggest a slight amount of across-channel similarity, at least for the lower  $|\Delta(k)|$  values. However, the effect is not as strong as what we observe for the persistence of ego-nets (compared with Fig. 4). Similarly to the case for the persistence curves, we observe that call-SMS similarity curves for the Gini coefficient and  $\beta$  do not change much when we switch from a single reference to  $|\Delta(k)|$ -specific references. This is another evidence showing that when comparing ego-nets of different sizes, Gini and  $\beta$  are better choices than  $L^2$  and JSD distances. The blue curves in plots (c)-(f) show the number of egos with a particular absolute call and SMS degree-difference ( $|\Delta(k)|$ ) (the values are on the right axes).



**Figure S5. Structural persistence of call ego-nets for egos in the LMP dataset that can be assigned Gini values but not valid  $\beta$ .** (a) Distribution of persistence scores with regard to a single degree-agnostic reference (grey violin-plots on the left) and with regard to  $|\Delta k|$ -specific references (orange violin-plots on the right). Most values for both references and all three distance functions are negative, indicating the persistence of the call ego-nets for these egos. Average persistence as a function of  $|\Delta(k)|$  (absolute degree change) is shown in (b) using the Gini distance function; in (c) using the  $L^2$  distance; and in (d) using the JSD distance. The average persistence scores are visualized in grey if they are measured against one single degree-agnostic reference distribution (regardless of  $|\Delta(k)|$ ) and in orange if against  $|\Delta(k)|$ -specific reference distributions. (See Eq. 8 for the definition of the persistence score.) The error bars illustrate the standard error of the mean of the samples. All three heterogeneity measures have negative average values across the degree range. The blue triangles in plots (b)-(d) show the number of egos with each particular absolute degree change ( $|\Delta(k)|$ ) (the values are on the right axes).

## References

- [1] ALEDAVOOD, T., LÓPEZ, E., ROBERTS, S. G., REED-TSOCHAS, F., MORO, E., DUNBAR, R. I., AND SARAMÄKI, J. Channel-specific daily patterns in mobile phone communication. In *Proceedings of ECCS 2014: European Conference on Complex Systems* (2016), Springer, pp. 209–218.
- [2] BARABASI, A.-L. The origin of bursts and heavy tails in human dynamics. *Nature* *435*, 7039 (2005), 207–211.
- [3] BERKMAN, L., GLASS, T., BRISSETTE, I., AND SEEMAN, T. From social integration to health: Durkheim in the new millennium. *Social Science & Medicine* *51* (2000), 843–857.
- [4] BHATTACHARYA, K., GHOSH, A., MONSIVAIS, D., DUNBAR, R. I., AND KASKI, K. Sex differences in social focus across the life cycle in humans. *Royal Society open science* *3*, 4 (2016), 160097.
- [5] CANDIA, J., GONZÁLEZ, M. C., WANG, P., SCHOENHARL, T., MADEY, G., AND BARABÁSI, A.-L. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical* *41*, 22 (2008), 224015.
- [6] CENTELLEGHER, S., LÓPEZ, E., SARAMÄKI, J., AND LEPRI, B. Personality traits and ego-network dynamics. *PloS one* *12*, 3 (2017), e0173110.
- [7] COLEMAN, J. S. *Foundations of social theory*. Belknap Press of Harvard University Press, Cambridge, Massachusetts, USA, 1990.
- [8] CSERMELY, P. *Weak links: Stabilizers of complex systems from proteins to social networks*. Springer-Verlag, 2006.
- [9] GINI, C. Measurement of inequality of incomes. *The Economic Journal* *31*, 121 (1921), 124–125.
- [10] GRANOVETTER, M. S. The strength of weak ties. *American Journal of Sociology* *78*, 6 (1973), 1360–1380.
- [11] HEYDARI, S., ROBERTS, S. G., DUNBAR, R. I., AND SARAMÄKI, J. Multichannel social signatures and persistent features of ego networks. *Applied Network Science* *3*, 1 (2018), 8.
- [12] HOLT-LUNSTAD, J., SMITH, T. B., AND LAYTON, J. B. Social relationships and mortality risk: a meta-analytic review. *PLoS Medicine* *7*, 7 (2010), e1000316.
- [13] HOUSE, J. S., LANDIS, K. R., AND UMBERSON, D. Social relationships and health. *Science* *241*, 4865 (1988), 540–545.
- [14] IÑIGUEZ, G., HEYDARI, S., KERTÉSZ, J., AND SARAMÄKI, J. Universal patterns in egocentric communication networks. *Nature Communications* *14* (2023), 5217.
- [15] KARSAI, M., KIVELÄ, M., PAN, R. K., KASKI, K., KERTÉSZ, J., BARABÁSI, A.-L., AND SARAMÄKI, J. Small but slow world: How network topology and burstiness slow down spreading. *Physical Review E* *83*, 2 (2011), 025102.
- [16] KIVELÄ, M., PAN, R. K., KASKI, K., KERTÉSZ, J., SARAMÄKI, J., AND KARSAI, M. Multiscale analysis of spreading in a large communication network. *Journal of Statistical Mechanics* *2012*, 03 (2012), P03005.

- [17] KOLTSOVA, O. Y., MARARITSA, L. V., TERPILOVSKII, M. A., AND SINYAVSKAYA, Y. E. Social signature in an online environment: Stability and cognitive limits. *Computers in Human Behavior* 122 (2021), 106856.
- [18] KORTE, C., AND MILGRAM, S. Acquaintance linking between white and negro populations: Application of the small world problem. *Journal of Personality and Social Psychology* 15, 2 (1970), 101–108.
- [19] KOVANEN, L., KASKI, K., KERTÉSZ, J., AND SARAMÄKI, J. Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences. *Proceedings of the National Academy of Sciences* 110, 45 (2013), 18070–18075.
- [20] KRINGS, G., KARSAI, M., BERNHARDSSON, S., BLONDEL, V. D., AND SARAMÄKI, J. Effects of time window size and placement on the structure of an aggregated communication network. *EPJ Data Science* 1, 1 (2012), 4.
- [21] LI, Y., AND BOND, R. M. Evidence of the persistence and consistency of social signatures. *Applied Network Science* 7, 1 (2022), 10.
- [22] LITWIN, H., AND STOECKEL, K. J. Confidant network types and well-being among older europeans. *The Gerontologist* 54 (2014), 762–772.
- [23] MARSDEN, P. Network data and measurement. *Annual Review of Sociology* 16 (1990), 435–463.
- [24] McDONALD, J. B., AND JENSEN, B. C. An analysis of some properties of alternative measures of income inequality based on the gamma distribution function. *Journal of the American Statistical Association* 74, 368 (1979), 856–860.
- [25] MIRITELLO, G., MORO, E., LARA, R., MARTÍNEZ-LÓPEZ, R., BELCHAMBER, J., ROBERTS, S. G., AND DUNBAR, R. I. Time as a limited resource: Communication strategy in mobile phone networks. *Social Networks* 35, 1 (2013), 89–95.
- [26] OLIVEIRA, J. G., AND BARABÁSI, A.-L. Human dynamics: Darwin and einstein correspondence patterns. *Nature* 437, 7063 (2005), 1251–1251.
- [27] ONNELA, J.-P., SARAMÄKI, J., HYVÖNEN, J., SZABÓ, G., DE MENEZES, M. A., KASKI, K., BARABÁSI, A.-L., AND KERTÉSZ, J. Analysis of a large-scale weighted network of one-to-one human communication. *New journal of physics* 9, 6 (2007), 179.
- [28] ONNELA, J.-P., SARAMÄKI, J., HYVÖNEN, J., SZABÓ, G., LAZER, D., KASKI, K., KERTÉSZ, J., AND BARABÁSI, A.-L. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences* 104, 18 (2007), 7332–7336.
- [29] PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T., AND FLANNERY, B. P. *Numerical recipes in C*. Cambridge University Press, 1988.
- [30] PUTNAM, R. D. *Bowling alone: The collapse and revival of the American society*. Simon and Schuster, 2000.
- [31] ROY, C., BHATTACHARYA, K., DUNBAR, R. I., AND KASKI, K. Turnover in close friendships. *Scientific Reports* 12, 1 (2022), 11018.
- [32] SAPIEZYNSKI, P., STOPCZYNSKI, A., LASSEN, D. D., AND JØRGENSEN, S. L. The Copenhagen Networks Study interaction data. *figshare* (2019). URL: <https://doi.org/10.6084/m9.figshare.7267433>.

- [33] SAPIEZYNSKI, P., STOPCZYNSKI, A., LASSEN, D. D., AND LEHMANN, S. Interaction data from the Copenhagen Networks Study. *Scientific Data* 6, 1 (2019), 1–10.
- [34] SARAMÄKI, J., LEICHT, E. A., LÓPEZ, E., ROBERTS, S. G., REED-TSOCHAS, F., AND DUNBAR, R. I. Persistence of social signatures in human communication. *Proceedings of the National Academy of Sciences* 111, 3 (2014), 942–947.
- [35] SIMMEL, G. Conflict (kh wolff, trans.). *Glencoe, IL: The Free Press. (Original work published 1908)* (1955).
- [36] STOPCZYNSKI, A., SEKARA, V., SAPIEZYNSKI, P., CUTTONE, A., MADSEN, M. M., LARSEN, J. E., AND LEHMANN, S. Measuring large-scale social networks with high resolution. *PloS ONE* 9, 4 (2014), e95978.
- [37] TÖRÖK, J., MURASE, Y., JO, H.-H., KERTÉSZ, J., AND KASKI, K. What big data tells: Sampling the social network by communication channels. *Physical Review E* 94 (2016), 052319.
- [38] UNICOMB, S., IÑIGUEZ, G., AND KARSAI, M. Threshold driven contagion on weighted networks. *Scientific Reports* 8, 1 (2018), 1–10.
- [39] UREÑA-CARRION, J., SARAMÄKI, J., AND KIVELÄ, M. Estimating tie strength in social networks using temporal communication data. *EPJ Data Science* 9, 1 (2020), 37.
- [40] WANG, D., PEDRESCHI, D., SONG, C., GIANNOTTI, F., AND BARABASI, A.-L. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2011), ACM, pp. 1100–1108.
- [41] WITTIG, R. M., CROCKFORD, C., LEHMANN, J., WHITTEN, P. L., SEYFARTH, R. M., AND CHENEY, D. L. Focused grooming networks and stress alleviation in wild female baboons. *Hormones and Behavior* 54, 1 (2008), 170–177.
- [42] ZIGNANI, M., QUADRI, C., GAITO, S., AND ROSSI, G. P. Calling, texting, and moving: multidimensional interactions of mobile phone users. *Computational Social Networks* 2 (2015), 1–24.