Temporal stability in human interaction networks

Renato Fabbri,[1, a)] Ricardo Fabbri,[b)] Deborah C. Antunes,[c)] Marilia M. Pisani,[d)] Leonardo P. Maia,[e)] and Osvaldo N. Oliveira Jr.[f)]

São Carlos Institute of Physics, University of São Paulo (IFSC/USP), PO Box 369, 13560-970, São Carlos, SP, Brazil

(Dated: 2 November 2015)

This paper reports on stable properties of human interaction networks, with benchmarks derived from public email lists. Activity along time and topology evolution were observed by snapshots in a timeline and at different scales. Our analysis show that the activity across timescales, ranging from seconds to months, is practically the same for all networks. The most important metrics to the dispersion of participants in the topological measures space are shown to be are centrality measurements (degree, strength and betweenness), followed by symmetry-related metrics and then clustering coefficient. The observed activity of participants follows the expected scale-free trace, thus yielding the hub, intermediary and periphery classes of vertices by comparison against the Erdös-Rényi model. The relative sizes of these three sectors are shown to be essentially the same for all email lists and the same along time. Typically, 3-12% of the vertices are hubs, 15-45% are intermediary and 44-81% are peripheral vertices. Similar results for the distribution of participants in the three sectors and for the relative importance of the topological metrics were obtained for 12 additional networks from Facebook, Twitter and Participabr. These properties are consistent with expectations derived from the literature and may be general for human interaction networks, which has important implications for establishing a typology of participants based on quantitative criteria.

PACS numbers: 89.75.Fb,05.65.+b,89.65.-s
Keywords: complex networks, pattern recognition, statistics, social network analysis, anthropological physics, social psychology of big data, typology

'The reason for the persistent plausibility of the typological approach, however, is not a static biological one, but just the opposite: dynamic and social.' - *Adorno et al, 1969, p. 747*

I. INTRODUCTION

First studies explicitly about human interaction networks date from the nineteenth century while the foundation of social network analysis is generally attributed to the psychiatrist Jacob Moreno in mid twentieth century[1,2]. With the increasing availability of data related to human interactions, research about these networks has grown continuously. Contributions can now be found in a variety of fields, from social sciences and humanities[3] to computer science[4] and physics[5,6], given the multidisciplinary nature of the topic. One of the approaches from an exact science perspective is to represent interaction networks as complex networks[5,6], with which several features of human interaction have been revealed. For example, the topology of human interaction networks exhibits a scale-free trace, which points to the existence of a small number of highly connected hubs and a large number of poorly connected nodes. The dynamics of complex networks representing human interaction has also been addressed[7,8], but only to a limited extent, since research is normally focused on a particular metric or task, such as accessibility or community detection[9,10].

In this paper we analyze the evolution of human interaction networks. Interaction networks from email lists were the most convenient for deriving results and for benchmarking while networks from Facebook, Twitter and Participabr enhanced the generalization. Using a timeline of activity snapshots with a constant number of contiguous messages, we found remarkable stability for important network properties. For instance, the activity along different timescales exhibit pronounced patterns; the most basic topological measures always combine into very characteristic principal components; and the fractions of participants in each of the hubs, intermediady and periphery Erdös sectors are unshaken. This is not an intuitive result, given that participants constantly transition in network structure. Because these properties were shared by networks from various sources, and are consistent with the complex networks literature, we advocate that the conclusions might be valid for general classes of interaction networks. In particular, this allows us to

a) http://ifsc.usp.br/~fabbri/; Electronic mail: fabbri@usp.br
b) http://www.lems.brown.edu/~rfabbri/; Electronic mail: rfabbri@iprj.uerj.br; Instituto Politécnico, Universidade Estadual do Rio de Janeiro (IPRJ)
c) http://lattes.cnpq.br/1065956470701739; Electronic mail: deborahantunes@gmail.com; Curso de Psicologia, Universidade Federal do Cerá (UFC)
d) http://lattes.cnpq.br/6738980149860322; Electronic mail: marilia.m.pisani@gmail.com;
e) http://www.ifsc.usp.br/~lpmaia/; Electronic mail: lpmaia@ifsc.usp.br ; Also at IFSC-USP
f) www.polimeros.ifsc.usp.br/professors/professor.php?id=4; Electronic mail: chu@ifsc.usp.br; Also at IFSC-USP

bridge the gap between data analysis and social sciences in the discussion of types of networks and of participants. Noteworthy, typologies are the canon of scientific literature for the classification of human agents, with pragmatic standards[11] and critical paradigms[12,13].

Section I A describes related work, while data, scripts and methods of analysis are given in Section II and Section III. Section IV reports results and discussion, leading to Section V for conclusions. Supplementary data analysis, including directions for video and sound mappings of network structures, and numeric detailed results for networks from Twitter, Facebook and Participabr, are provided in the Supporting Information document.

A. Related work

Research on network evolution is often restricted to network growth, in which there is a monotonic increase in the number of events[7]. Network types have been discussed with regard to the number of participants, intermittence of their activity and network longevity[7]. Two topologically different networks emerged from human interaction networks, depending on whether the frequency of interactions follows a generalized power law or an exponential connectivity distribution[14]. In email list networks, scale-free properties were reported with $\alpha \approx 1.8$[4] (as in web browsing and library loans[5]), and different linguistic traces were related to weak and strong ties[15].

The fact that unreciprocated edges often exceed 50% in human interaction networks[8] motivated the inclusion of symmetry metrics in our analysis. No correlation of topological characteristics and geographical coordinates was found[16], therefore geographical positions were not considered in our study. Gender related behavior in mobile phone datasets was indeed reported[17] but is not relevant for the present manuscript because email messages and addresses have no gender related metadata[18].

II. DATA AND SCRIPTS

Email list messages were obtained from the Gmane email archive[18], which consists of more than $20,000$ email lists (discussion groups) and more than 130×10^6 messages[19]. These lists cover a variety of topics, mostly technology-related. The archive can be described as a corpus along with message metadata, including sent time, place, sender name, and sender email address. The usage of the Gmane database in scientific research is reported in studies of isolated lists and of lexical innovations[4,15].

We analyzed many email lists together with data from Twitter, Facebook and Participabr and selected four email lists for a thorough analysis, from which general properties can be inferred:

- Linux Audio Users list[20], with participants from different countries with artistic and technological interests. English is the prevailing language. Abbreviated as LAU from now on.

- Linux Audio Developers list[21], with participants from different countries; a more technical and less active version of LAU. English is the prevailing language. Abbreviated as LAD from now on.

- Developer's list for the standard C++ library[22], with computer programmers from different countries. English is the prevailing language. Abbreviated as CPP from now on.

- List of the MetaReciclagem project[23], a Brazilian digital culture interested email list. Portuguese is the prevailing language, although Spanish and English are also incident. Abbreviated as MET from now on.

TABLE I. Columns $date_1$ and $date_M$ have dates of first and last messages from the 20,000 messages considered in each email list. N is the number of participants (number of different email addresses), Γ is the number of discussion threads (count of messages without antecedent), $\overline{M}$ is the number of messages missing in the 20,000 collection ($100\frac{23}{20000} = 0.115$ percent in the worst case).

list	$date_1$	$date_M$	N	Γ	$\overline{M}$
LAU	2003-06-29	2005-07-23	1181	3372	5
LAD	2003-06-30	2009-10-07	1268	3109	4
MET	2005-08-01	2008-03-07	492	4607	23
CPP	2002-03-12	2009-08-25	1052	4506	7

The first 20,000 messages of each list were considered, with basic attributes of total timespan, authors, threads and missing messages indicated in Table I. We considered 140 additional email lists to report on the interdependence between the number of participants and the number of discussion threads. Furthermore, 12 networks from Facebook (8), Twitter (2) and Participabr (2) are scrutinized in the Supporting Information document for better hypothesizing about the generality of the results.

A. Availability

All data and scripts needed to derive results, figures, tables and this article itself are publicly available. Email messages are downloadable from the Gmane public database[19]. Data annotated from Facebook and Twitter are in a public repository[24]. Data from Participabr was used from the linked data/semantic web RDF triples reported[25] and available[26]. Computer scripts are delivered through a public domain Python PyPI package and an open Git repository[18]. This open approach to both data and scripts reinforces the scientific aspect of the contribution[27] and mitigates ethical and moral issues involved in researching systems constituted of human individuals[28,29].

III. METHODS

A. Time activity statistics

Messages were counted over time as histograms in the scales of seconds, minutes, hours, days of the week, days of the month, and months of the year. Most standard measures of location and dispersion, e.g. the usual mean and standard deviation, hold little meaning in a compact Riemannian manifold, such as the recurrent time periods that we are interested in. Equivalent measures were taken using circular statistics, in which each measurement t is represented as a unit complex number, $z = e^{i\theta} = \cos(\theta) + i\sin(\theta)$, where $\theta = t\frac{2\pi}{T}$, and T is the period in which the counting is repeated. For example, $\theta = 12\frac{2\pi}{24} = \pi$ for a message sent at $t = 12h$ and given $T = 24h$ for days. The moments m_n, lengths of moments R_n, mean angles θ_μ, and rescaled mean angles θ'_μ are defined as:

$$\begin{aligned} m_n &= \frac{1}{N}\sum_{i=1}^{N} z_i^n \\ R_n &= |m_n| \\ \theta_\mu &= Arg(m_1) \\ \theta'_\mu &= \frac{T}{2\pi}\theta_\mu \end{aligned} \tag{1}$$

θ'_μ is used as the measure of location. Dispersion is measured using the circular variance $Var(z)$, the circular standard deviation $S(z)$, and the circular dispersion $\delta(z)$:

$$\begin{aligned} Var(z) &= 1 - R_1 \\ S(z) &= \sqrt{-2\ln(R_1)} \\ \delta(z) &= \frac{1 - R_2}{2R_1^2} \end{aligned} \tag{2}$$

Also, the ratio $r = \frac{b_l}{b_h}$ between the lowest b_l and the highest b_h incidences on the histograms served as a further clue of how close the distribution was to being uniform. As expected, a positive correlation was found in all $r, Var(z)$, $S(z)$ and $\delta(z)$ dispersion measures, which can be noticed in the Section SI A of the Supporting Information document. The circular dispersion $\delta(z)$ was found more sensitive and therefore preferred in the discussion of results.

B. Interaction networks

Edges in interaction networks can be modeled both as weighted or unweighted, both as directed or undirected[4,30,31]. Networks in this paper are directed and weighted, the most informative of the possibilities. We did not investigate directed unweighted, undirected weighted, and undirected unweighted representations of the interaction networks.

The interaction networks were obtained as follows: a direct response from participant B to a message from participant A yields an edge from A to B, as information went from A to B. The reasoning is: if B wrote a response to a message from A, he/she read what A wrote and formulated a response, so B assimilated information from A, thus $A \to B$. Edges in both directions are allowed. Each time an interaction occurs, the value of one is added to the edge weight. Selfloops were regarded as non-informative and discarded. Inverting edge direction canonically yields the status network: B read the message and considered what A wrote worth responding, giving status to A, thus $B \to A$. This paper considers by convention the information network as described above ($A \to B$) and depicted in Figure 1. These interaction networks are reported in the literature as exhibiting scale-free and small-world properties, as expected for a number of social networks[2,4].

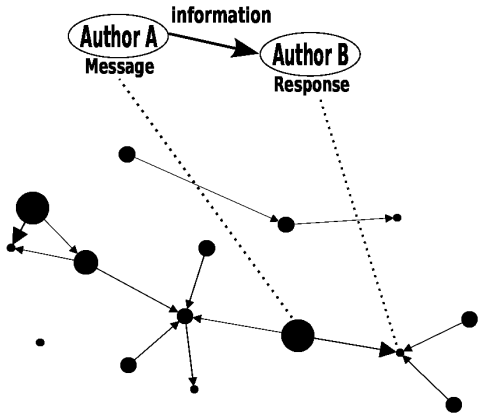

FIG. 1. The formation of interaction networks from exchanged messages. Each vertex represents a participant. A reply message from author B to a message from author A is regarded as evidence that B received information from A and yields a directed edge. Multiple messages add "weight" to a directed edge. Further details are given in Section III B.

1. Topological metrics

The topology of the networks was characterized from a small selection of the most basic and fundamental measurements for each vertex, as follows:

- Degree k_i: number of edges linked to vertex i.
- In-degree k_i^{in}: number of edges ending at vertex i.
- Out-degree k_i^{out}: number of edges departing from vertex i.

- Strength s_i: sum of weights of all edges linked to vertex i.
- In-strength s_i^{in}: sum of weights of all edges ending at vertex i.
- Out-strength s_i^{out}: sum of weights of all edges departing from vertex i.
- Clustering coefficient cc_i: fraction of pairs of neighbors of i that are linked, i.e. the standard clustering coefficient metric for undirected graphs.
- Betweenness centrality bt_i: fraction of geodesics that contain vertex i. The betweenness centrality index was computed for weighted digraphs as specified in[32].

The non-standard metrics bellow were formulated to capture symmetries in the activity of participants:

- Asymmetry of vertex i: $asy_i = \frac{k_i^{in} - k_i^{out}}{k_i}$.
- Average asymmetry of edges at vertex i: $\mu_i^{asy} = \frac{\sum_{j \in J_i} e_{ji} - e_{ij}}{|J_i|}$, where e_{ij} is 1 if there is an edge from i to j, and 0 otherwise, and J_i is the set of neighbors of vertex i.
- Standard deviation of asymmetry of edges: $\sigma_i^{asy} = \sqrt{\frac{\sum_{j \in J_i} [\mu_i^{asy} - (e_{ji} - e_{ij})]^2}{|J_i|}}$.
- Disequilibrium: $dis_i = \frac{s_i^{in} - s_i^{out}}{s_i}$.
- Average disequilibrium of edges: $\mu_i^{dis} = \frac{\sum_{j \in J_i} \frac{w_{ji} - w_{ij}}{w_{ji} - w_{ij}}}{|J_i|}$, where w_{xy} is the weight of edge $x \to y$ and zero if there is no such edge.
- Standard deviation of disequilibrium of edges: $\sigma_i^{dis} = \sqrt{\frac{\sum_{j \in J_i} \left[\mu_i^{dis} - \frac{w_{ji} - w_{ij}}{w_{ji} + w_{ij}}\right]^2}{|J_i|}}$.

Both standard and non-standard metrics are used for the Erdös sectioning (described in Secion III C) and for performing PCA (as described in Section III D).

C. Erdös sectioning

It is often useful to think of vertices as hubs, peripheral and intermediary. Following this pattern, the peripheral, intermediary and hub sectors of the empirical networks were derived from a comparison against an Erdös-Rényi network with the same number of edges and vertices, as depicted in Figure 2. We refer to this procedure as *Erdös sectioning*, with the resulting sectors regarded as *Erdös sectors*. The Erdös sectioning was theorized about by M. O. Jackson in his video lectures[33], but this might be the first time the method is reported to be applied.

The degree distribution $\widetilde{P}(k)$ of a scale-free network $\mathcal{N}_f(N, z)$ with N vertices and z edges has less average degree nodes than the distribution $P(k)$ of an Erdös-Rényi network with the same number of vertices and edges. Indeed, we define in this work the intermediary sector of a network to be the set of all the nodes whose degree is less abundant in the real network than on the Erdös-Rényi model:

$$\widetilde{P}(k) < P(k) \Rightarrow \text{k is intermediary degree} \tag{3}$$

If $\mathcal{N}_f(N, z)$ is directed and has no self-loops, the probability of the existence of an edge between two arbitrary vertices is $p_e = \frac{z}{N(N-1)}$. A vertex in the ideal Erdös-Rényi digraph with the same number of vertices and edges, and thus the same probability p_e for the presence of an edge, will have degree k with probability

$$P(k) = \binom{2(N-1)}{k} p_e^k (1 - p_e)^{2(N-1)-k} \tag{4}$$

The lower degree fat tail corresponds to the border vertices, i.e. the peripheral sector or periphery where $\widetilde{P}(k) > P(k)$ and k is lower than any intermediary sector value of k. The higher degree fat tail is the hub sector, i.e. $\widetilde{P}(k) > P(k)$ and k is higher than any intermediary sector value of k. The reasoning for this classification is as follows: vertices so connected that they are virtually inexistent in networks connected without distinction of the vertices, i.e. without preferential attachment and as in the Erdös-Rényi model, are correctly associated to the hub sector. Vertices with very few connections, which are way more abundant than expected in the Erdös-Rényi model, are assigned to the periphery. Vertices with degree values predicted as the most abundant in a Erdös-Rényi model, near the average, and less frequent in the real network, are classified as intermediary.

To ensure statistical validity of the histograms, bins can be chosen to contain at least η vertices of the real network. The range Δ of incident values should be partitioned in m parts $\Delta = \cup_{i=1}^{m} \Delta_i$, with $\Delta_i \cap \Delta_j = \emptyset \; \forall \; i \neq j$ and:

$$\begin{aligned} \Delta_i = \Bigg\{ k \;\; | \;\; & \overline{\Delta}_{i-1} < k \leq j \text{ and} \\ & \Bigg[N - \sum_{k=0}^{\overline{\Delta}_{i-1}} \eta_k < \eta \text{ or} \\ & \Bigg[\sum_{k=\overline{\Delta}_{i-1}+1}^{j} \eta_k \geq \eta \text{ and} \\ & \Big(\sum_{k=\overline{\Delta}_{i-1}+1}^{j-1} \eta_k < \eta \text{ or } j = \overline{\Delta}_{i-1} + 1 \Big) \Bigg] \Bigg] \Bigg\} \end{aligned} \tag{5}$$

where η_k is the number of vertices with degree k, while $\overline{\Delta}_i = max(\Delta_i)$, and $\overline{\Delta}_0 = -1$. Equation 3 can now be written in the form:

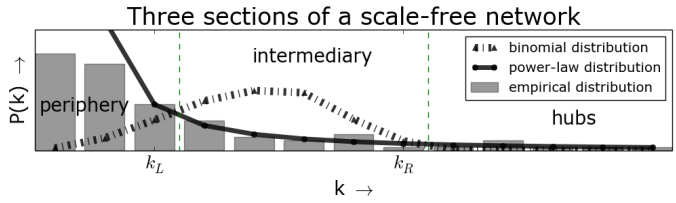

FIG. 2. Classification of vertices by comparing degree distributions[33]. The binomial distribution of the Erdös-Rényi network model exhibits more intermediary vertices, while a scale-free network, associated with the power-law distribution, has more peripheral and hub vertices. The sector borders are defined with respect to the intersections of the distributions. Characteristic degrees are in the compact intervals: $[0, k_L]$, $(k_L, k_R]$, $(k_R, k_{max}]$ for the periphery, intermediary and hub sectors, the "Erdös sectors". The connectivity distribution of empirical interaction networks, e.g. derived from email lists, can be sectioned by comparison against the associated binomial distribution with the same number of vertices and edges. In this figure, a snapshot of 1000 messages from CPP list yield the degree distribution of an interaction network of 98 nodes and 235 edges. A through exposition of the method is exposed in Section III C.

$$\sum_{x=min(\Delta_i)}^{\overline{\Delta}_i} \widetilde{P}(x) < \sum_{x=min(\Delta_i)}^{\overline{\Delta}_i} P(x) \Leftrightarrow \quad (6)$$
$$\Leftrightarrow \Delta_i \text{ spans intermediary degree values.}$$

If the strength s is used for comparison, P remains the same, but $P(\kappa_i)$ with $\kappa_i = \frac{s_i}{\overline{w}}$ should be used, where $\overline{w} = 2\frac{z}{\sum_i s_i}$ is the average weight of an edge and s_i the strength of vertex i. For in and out degrees (k^{in}, k^{out}), the real network should be compared against

$$\hat{P}(k^{way}) = \binom{N-1}{k^{way}} p_e^k (1-p_e)^{N-1-k^{way}}, \quad (7)$$

where *way* can be *in* or *out*. In and out strengths (s^{in}, s^{out}) are divided by $\overline{w}$ and compared also using $\hat{P}$. Note that p_e remains the same, as each edge yields an incoming (or outgoing) edge, and there are at most $N(N-1)$ incoming (or outgoing) edges, thus $p_e = \frac{z}{N(N-1)}$, as with the total degree.

In other words, let γ and ϕ be integers in the intervals $1 \leq \gamma \leq 6$, $1 \leq \phi \leq 3$, and each of the basic six Erdös sectioning possibilities $\{E_\gamma\}$ have three Erdös sectors $E_\gamma = \{e_{\gamma,\phi}\}$ defined as

$$\begin{aligned} e_{\gamma,1} &= \{\, i \mid \overline{k}_{\gamma,L} \geq \overline{k}_{\gamma,i}\} \\ e_{\gamma,2} &= \{\, i \mid \overline{k}_{\gamma,L} < \overline{k}_{\gamma,i} \leq \overline{k}_{\gamma,R}\} \\ e_{\gamma,3} &= \{\, i \mid \quad\quad \overline{k}_{\gamma,i} < \overline{k}_{\gamma,R}\}, \end{aligned} \quad (8)$$

where $\{\overline{k}_{\gamma,i}\}$ is

$$\begin{aligned} \overline{k}_{1,i} &= k_i \\ \overline{k}_{2,i} &= k_i^{in} \\ \overline{k}_{3,i} &= k_i^{out} \\ \overline{k}_{4,i} &= \frac{s_i}{\overline{w}} \\ \overline{k}_{5,i} &= \frac{s_i^{in}}{\overline{w}} \\ \overline{k}_{6,i} &= \frac{s_i^{out}}{\overline{w}} \end{aligned} \quad (9)$$

and both $\overline{k}_{\gamma,L}$ and $\overline{k}_{\gamma,R}$ are found using $P(\overline{k})$ or $\hat{P}(\overline{k})$ as described above.

Since different metrics can be used to identify the three types of vertices, more than one metric can be used simultaneously, which is convenient when analysing small networks, such as the cases where $ws = 50$ in Section SIII of the Supporting Information. After a careful consideration of possible combinations, these were reduced to six:

- Exclusivist criterion C_1: vertices are only classified if the class is the same according to all metrics. In this case, vertices classified do not usually reach N (or 100%), which is indicated by a black line in Figure 3.
- Inclusivist criterion C_2: a vertex has the class given by any of the metrics. Therefore, a vertex may belong to more than one class, and the total number of memberhips may exceed N (or 100%), which is indicated by a black line in Figure 3.
- Exclusivist cascade C_3: vertices are only classified as hubs if they are hubs according to all metrics. Intermediary are the vertices classified either as intermediary or hubs with respect to all metrics. The remaining vertices are regarded as peripheral.
- Inclusivist cascade C_4: vertices are hubs if they are classified as such according to any of the metrics. The remaining vertices are intermediary if they belong to this category for any of the metrics. Peripheral vertices are those which are classified as such with respect to all metrics.
- Exclusivist externals C_5: vertices are hubs if they are classified as such according to all the metrics. The remaining vertices are peripheral if they are peripheral or hubs for all metrics. The remaining nodes are intermediary.
- Inclusivist externals C_6: hubs are vertices classified as hubs according to any metric. The remaining vertices are peripheral if they are classified as such according to any metric. The rest of the vertices are intermediary.

Using Equations (8), these *compound criteria* C_δ, with δ integer in the interval $1 \leq \delta \leq 6$, can be specified as:

$$
\begin{aligned}
C_1 &= \{c_{1,\phi} = \{i \mid i \in e_{\gamma,\phi},\ \forall\, \gamma\}\} \\
C_2 &= \{c_{2,\phi} = \{i \mid \exists\ \gamma : i \in e_{\gamma,\phi}\}\} \\
C_3 &= \{c_{3,\phi} = \{i \mid i \in e_{\gamma,\phi'},\ \forall\, \gamma,\ \forall\, \phi' \geq \phi\}\} \\
C_4 &= \{c_{4,\phi} = \{i \mid i \in e_{\gamma,\phi'},\ \forall\, \gamma,\ \forall\, \phi' \leq \phi\}\} \\
C_5 &= \{c_{5,\phi} = \{i \mid i \in e_{\gamma,\phi'},\ \forall\, \gamma, \\
&\qquad \forall\, (\phi'+1)\%4 \leq (\phi+1)\%4\}\} \\
C_6 &= \{c_{6,\phi} = \{i \mid i \in e_{\gamma,\phi'},\ \forall\, \gamma, \\
&\qquad \forall\, (\phi'+1)\%4 \geq (\phi+1)\%4\}\}
\end{aligned}
\tag{10}
$$

Notice that the exclusivist cascade is the same sectioning of an inclusivist cascade from periphery to hubs, but with inverted order of sectors. The simplification of all possible compound possibilities to the small set listed above might be formalized in strict mathematical terms, but this was considered out of the scope for current interests.

D. Principal Component Analysis of topological metrics

Principal Component Analysis (PCA) is a well documented technique[34] and was employed in the investigation to acquire knowledge about: 1) which metrics contribute to each principal component and in what proportion; 2) how much of the dispersion is concentrated in each component; 3) expected values and dispersions for these quantities over various networks. This enables a characterization of the human interaction networks in terms of the relative importance of networks measures ad the way they combine.

Let $\mathbf{X} = \{X[i,j]\}$ be a matrix where each element is the value of of each metric j at vertex i . Let $\mu_X[j] = \frac{\sum_i X[i,j]}{I}$ be the mean of metric j over all I vertices, $\sigma_X[j] = \sqrt{\frac{\sum_i (X[i,j]-\mu_X[j])^2}{I}}$ the standard deviation of metric j, and $\mathbf{X}' = \{X'[i,j]\} = \left\{\frac{X[i,j]-\mu_X[j]}{\sigma_X[j]}\right\}$ the matrix with the *z-score* of each metric. Let $\mathbf{V} = \{V[j,k]\}$ be the matrix $J \times J$ of eigenvectors of the covariance matrix $\mathbf{C}$ of $\mathbf{X}'$, one eigenvector per column. Each eigenvector combines the original measures into one principal component, therefore $V'[j,k] = 100\frac{|V[j,k]|}{\sum_{j'}|V[j',k]|}$ is the percentage of the principal component k that is directly proportional to the measure j. Let $\mathbf{D} = \{D[k]\}$ be the eigenvalues associated to the eigenvectors $\mathbf{V}$, then $D'[k] = 100\frac{D[k]}{\sum_{k'} D[k']}$ is the percentage of total dispersion of the system that the principal component k is responsible for. We consider, in general, the three greatest eigenvalues and the respective eigenvectors in percentages: $\{(D'[k],\ V'[j,k])\}$. These usually sum up between 60 and 95% of the dispersion and reveal patterns for a first analysis. In particular, given L snapshots l of the interaction network, we are interested in the mean $\mu_{V'}[j,k]$ and the standard deviation $\sigma_{V'}[j,k]$ of the contribution of metric j to the principal component k, and the mean $\mu_{D'}[k]$ and the standard deviation $\sigma_{D'}[k]$ of the contribution of the component k to the dispersion of the system:

$$
\begin{aligned}
\mu_{V'}[j,k] &= \frac{\sum_{l=1}^{L} V'[j,k,l]}{L} \\
\sigma_{V'}[j,k] &= \sqrt{\frac{\sum_{l=1}^{L} (\mu_{V'} - V'[j,k,l])^2}{L}} \\
\mu_{D'}[k] &= \frac{\sum_{l=1}^{L} D'[k,l]}{L} \\
\sigma_{D'}[k] &= \sqrt{\frac{\sum_{l=1}^{L} (\mu_{D'} - D'[k,l])^2}{L}}
\end{aligned}
\tag{11}
$$

The covariance matrix $\mathbf{C}$ is the correlation matrix because $\mathbf{X}'$ is normalized. Therefore, $\mathbf{C}$ is also directly observed as a first clue for patterns by the most simple associations: low absolute values indicate low correlation (and a possible independence); high values indicate positive correlation; negative values with a high absolute value indicate negative correlation.

E. Evolution and audiovisualization of the networks

The evolution of the networks was observed within sequences of snapshots. In each sequence, a fixed number of messages, i.e. the window size ws, was used for all snapshots. The snapshots are disjoint in the message timeline, and were used to perform both PCA with topological metrics and Erdös sectioning. Figures and tables were usually inspected in $ws = \{50, 100, 200, 400, 500, 800, 1000, 2000, 2500, 5000, 10000\}$ messages. Variations in the number of vertices, edges and other network characteristics, within the same window size ws, are exposed in Section SIII of the Supporting Information document.

Network structures were mapped to video animations, sound and musical structures, image galleries and online gadgets developed for this research[35–37]. Such *audiovisualizations* were crucial in the initial steps and to guide the research into the most important features of network evolution.

IV. RESULTS AND DISCUSSION

A. Activity along time

Regular patterns of activity were observed along time in the scales of seconds, minutes, hours, days and months. Histograms in each of the time scales were computed as were circular average and dispersion values. Values and dedicated expositions are provided in Tables II-VI.

For example, uniform activity is found with respect to seconds, minutes and days of the months, weekend days

TABLE II. The rescaled circular mean θ'_μ and the circular dispersion $\delta(z)$, described in Section III A, for different timescales. This example table was constructed using all LAD messages, and the results are the same for other lists, as shown in Section SI A of the Supporting Information document. The most uniform distribution of activity was found in seconds and minutes. Hours of the day exhibited the most concentrated activity (lowest $\delta(z)$), with mean between 2 p.m. and 3 p.m. ($\theta' = -9.61$). Weekdays, month days and months have mean near zero (i.e. near the beginning of the week, month and year) and high dispersion. Notice that θ'_u has the dimensional unit of the corresponding time period while $\delta(z)$ is dimensionless.

scale	mean θ'_μ	dispersion $\delta(z)$
seconds	–//–	9070.17
minutes	–//–	205489.40
hours	-9.61	4.36
weekdays	-0.03	29.28
month days	-2.65	2657.77
months	-0.56	44.00

TABLE III. Activity percentages along the hours of the day. Nearly identical distributions were observed on other social systems as shown in Section SI B 1 of the Supporting Information document. Highest activity was observed between noon and 6pm (with 1/3 of total day activity), followed by the time period between 6pm and midnight. Around 2/3 of the activity takes place from noon to midnight but the activity peak occurs between 11 a.m. and 12 p.m. This table shows results for the activity in CPP.

<table>
<tr><th></th><th>1h</th><th>2h</th><th>3h</th><th>4h</th><th>6h</th><th>12h</th></tr>
<tr><td>0h</td><td>3.66</td><td rowspan="2">6.42</td><td rowspan="2">8.20</td><td rowspan="3">9.30</td><td rowspan="4">10.67</td><td rowspan="10">33.76</td></tr>
<tr><td>1h</td><td>2.76</td></tr>
<tr><td>2h</td><td>1.79</td><td rowspan="2">2.88</td></tr>
<tr><td>3h</td><td>1.10</td><td rowspan="2">2.47</td></tr>
<tr><td>4h</td><td>0.68</td><td rowspan="2">1.37</td><td rowspan="3">3.44</td></tr>
<tr><td>5h</td><td>0.69</td></tr>
<tr><td>6h</td><td>0.83</td><td rowspan="2">2.07</td><td rowspan="3">4.35</td><td rowspan="5">23.09</td></tr>
<tr><td>7h</td><td>1.24</td></tr>
<tr><td>8h</td><td>2.28</td><td rowspan="2">6.80</td></tr>
<tr><td>9h</td><td>4.52</td><td rowspan="3">18.75</td><td rowspan="3">21.03</td></tr>
<tr><td>10h</td><td>6.62</td><td rowspan="2">14.23</td></tr>
<tr><td>11h</td><td>7.61</td></tr>
<tr><td>12h</td><td>6.44</td><td rowspan="2">12.48</td><td rowspan="3">18.95</td><td rowspan="3">25.05</td><td rowspan="4">37.63</td><td rowspan="10">66.24</td></tr>
<tr><td>13h</td><td>6.04</td></tr>
<tr><td>14h</td><td>6.47</td><td rowspan="2">12.57</td></tr>
<tr><td>15h</td><td>6.10</td><td rowspan="3">18.68</td></tr>
<tr><td>16h</td><td>6.22</td><td rowspan="2">12.58</td><td rowspan="3">23.60</td></tr>
<tr><td>17h</td><td>6.36</td></tr>
<tr><td>18h</td><td>6.01</td><td rowspan="2">11.02</td><td rowspan="3">15.88</td><td rowspan="5">28.61</td></tr>
<tr><td>19h</td><td>5.02</td></tr>
<tr><td>20h</td><td>4.85</td><td rowspan="2">9.23</td></tr>
<tr><td>21h</td><td>4.38</td><td rowspan="3">12.73</td><td rowspan="3">17.59</td></tr>
<tr><td>22h</td><td>4.06</td><td rowspan="2">8.36</td></tr>
<tr><td>23h</td><td>4.30</td></tr>
</table>

TABLE IV. Activity percentages along weekdays. Higher activity was observed during workweek days, with a decrease of activity on weekend days of at least one third and at most two thirds.

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
LAU	15.71	15.81	15.88	16.43	15.14	**10.13**	**10.91**
LAD	14.92	17.75	17.01	15.41	14.21	**10.40**	**10.31**
MET	17.53	17.54	16.43	17.06	17.46	**7.92**	**6.06**
CPP	17.06	17.43	17.61	17.13	16.30	**6.81**	**7.67**

exhibit about half the activity of regular weekdays, and there is a peak of activity between 11am and 12pm.

In the scales of seconds and minutes, activity is uniform, with the messages being slightly more evenly distributed in all lists than in simulations with the uniform distribution[38]. In the networks, $\frac{min(incidence)}{max(incidence)} \in (0.784, .794)$ while simulations reach these values but have on average more discrepant higher and lower peaks $\xi = \frac{min(incidence')}{max(incidence')} \Rightarrow \mu_\xi = 0.7741$ and $\sigma_\xi = 0.02619$. Therefore, the incidence of messages at each second of a minute and at each minute of an hour was considered uniform. In these cases, the circular dispersion is maximized and the mean has little meaning as indicated in Table II. As for the hours of the day, an abrupt peak is found between 11am and 12pm with the most active period being the afternoon, with one third of total daily activity, and two thirds of activity is allocated in the second 12h of each day. Days of the week revealed a decrease between one third and two thirds of activity on weekends. Days of the month were regarded as homogeneous with an inconclusive slight tendency of the first week to be more active. Months of the year revealed patterns matching usual work and academic calendars. The time period examined here was not sufficient for the analysis of activity along the years. These patterns are exemplified in Tables III-VI.

B. Stable sizes of Erdös sectors

The distribution of vertices in the hub, intermediary, periphery Erdös sectors is remarkably stable along time if the snapshots hold 200 or more messages, as evident in Figure 3 of current document and in Section SIII of the Supporting Information document. Activity is highly concentrated on the hubs, while a very large number of peripheral vertices contribute to only a fraction of the activity. This is expected for a system with a scale-free profile and confirmed by Table VII of the distribution of activity among participants.

Typically, $[3\% - 12\%]$ of the vertices are hubs, $[15\% - 45\%]$ are intermediary and $[44\% - 81\%]$ are peripheral, which is consistent with other studies[39]. These results hold for the total, in and out degrees and strengths. Stable sizes are also observed for 100 or less messages if clas-

TABLE V. Activity along the days of the month cycle. Nearly identical distributions are found in all systems as indicated in Section SI B 3 of the Supporting Information. Although slightly higher activity rates are found in the beginning of the month, the most important feature seems to be the homogeneity made explicit by the high circular dispersion in Table II. This specific example and empirical table corresponds to the activity of the MET email list.

	1 day	5	10	15 days
1	3.05	18.25	35.24	50.96
2	3.38			
3	3.62			
4	4.25			
5	3.94			
6	3.73	16.98		
7	3.17			
8	3.26			
9	3.56			
10	3.26			
11	3.81	15.73	31.98	
12	2.91			
13	3.30			
14	2.75			
15	2.95			
16	3.36	16.25		49.04
17	3.16			
18	3.44			
19	3.36			
20	2.93			
21	3.20	15.79	32.78	
22	3.11			
23	3.60			
24	2.74			
25	3.13			
26	3.13	16.99		
27	3.07			
28	3.61			
29	3.60			
30	3.57			

TABLE VI. Activity percentages on months along the year. Activity is usually concentrated in Jun-Aug and/or in Dec-Mar, potentially due to academic calendars, vacations and end-of-year holidays. This table corresponds to activity in LAU. Similar results are posted in Section SI B 4 of the Supporting Information document.

	m.	b.	t.	q.	s.
Jan	10.22	19.56	28.24	35.09	49.16
Fev	9.34				
Mar	8.67	15.53			
Apr	6.86		20.93		
Mai	7.28	14.07		30.36	
Jun	6.80				
Jul	8.97	16.29	24.47		50.84
Ago	7.32				
Set	8.18	16.25		34.55	
Out	8.06		26.36		
Nov	7.64	18.30			
Dez	10.66				

sification of the three sectors is performed with one of the compound criteria established in Section III C. The networks often hold this basic structure with as few as 10-50 messages, i.e. concentration of activity and the abundance of low-activity participants take place even with very few messages, which is highlighted in Section SIII of the Supporting Information document. A minimum window size for the observation of more general properties might be inferred by monitoring both the giant component and the degeneration of the Erdös sectors.

In order to support hypotheses about the generality of these findings, we list the Erdös sector sizes of 12 networks from Facebook, Twitter and Participabr in Table S30 of the Supporting Information document. The fractions of hubs, intermediary and periphery nodes are essentially the same as for the email list networks but with exceptions and a greater variability.

TABLE VII. Distribution of activity among participants. The first column presents the percentage of messages sent by the most active participant. The column for the first quartile (Q_1) exposes the minimum percentage of participants responsible for at least 25% of total messages with the actual number in parenthesis. Similarly, the column for the first three quartiles Q_3 gives the minimum percentage of participants responsible for 75% of total messages. The last decile D_{-1} column exposes the maximum percentage of participants responsible for 10% of messages.

list	hub	Q_1	Q_3	D_{-1}
LAU	2.78	1.19 (26.35%)	13.12 (75.17%)	67.32 (-10.02%)
LAD	4.00	1.03 (26.64%)	11.91 (75.18%)	71.14 (-10.03%)
MET	11.14	1.02 (34.07%)	8.54 (75.64%)	80.49 (-10.02%)
CPP	14.41	0.29 (33.24%)	4.18 (75.46%)	83.65 (-10.04%)

C. Stability of principal components and the prevalence of symmetry over clusterization

The principal components of the participants are very stable in the topological space, i.e. in PCA space of network measures. Table VIII exemplifies the formation of principal components by exposing averages over non-overlapped activity snapshots of a network. The most important result of this application of PCA, the stability of principal components, is underpinned by the very small dispersion of the contribution of each metric to each principal component.

The first principal component is an average of centrality metrics: degrees, strengths and betweenness centrality. On one hand, the similar relevance of all centrality metrics is not surprising since they are highly correlated, e.g. degree and strength have Spearman correlation coefficient $\in [0.95, 1]$ and Pearson coefficient $\in [0.85, 1)$ for window sizes greater than a thousand messages. On the other hand, each of these metrics is related to a dif-

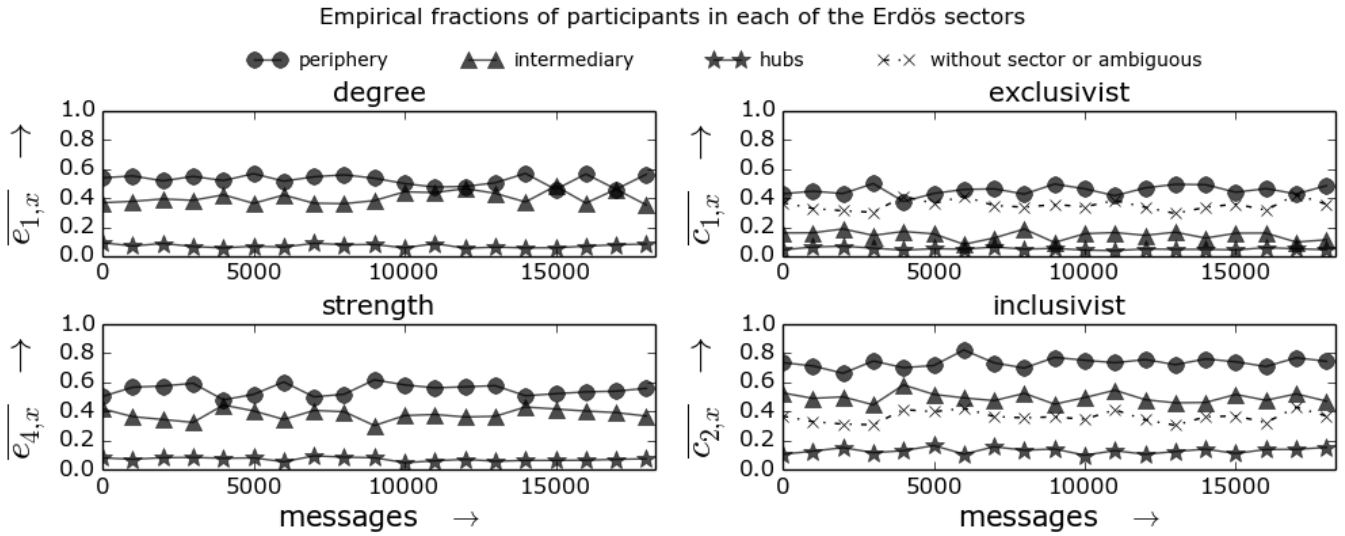

FIG. 3. Stability of Erdös sector sizes. Fractions of participants derived from degree and strength criteria, E_1 and E_4 in Section III C, are both on the left. Fractions derived from the exclusivist C_1 and the inclusivist C_2 compound criteria are exposed in the plots to the right. The ordinates $\overline{e_{n,x}} = \frac{|e_{n,x}|}{N}$ denote the fraction of participants in sector x through criterion E_n and, similarly, $\overline{c_{n,x}} = \frac{|c_{n,x}|}{N}$ denotes the fraction of participants in sector x through criterion C_n. Sections SIII and SIV of the Supporting Information document hold a systematic collection of such timeline figures with all simple and compound criteria specified in Section III C, and present these results for networks from Facebook, Twitter and Participabr.

TABLE VIII. Loadings for the 14 metrics into the principal components for the MET list, $ws = 1000$ messages in 20 disjoint positioning. The clustering coefficient (cc) appears as the first metric in the Table, followed by 7 centrality metrics and 6 symmetry-related metrics. Note that the centrality measurements, including degrees, strength and betweenness centrality, are the most important contributors for the first principal component, while the second component is dominated by symmetry metrics. The clustering coefficient is only relevant for the third principal component. The three components have in average more than 85% of the variance.

	PC1		PC2		PC3	
	μ	σ	μ	σ	μ	σ
cc	0.89	0.59	1.93	1.33	**21.22**	2.97
s	**11.71**	0.57	2.97	0.82	2.45	0.72
s^{in}	**11.68**	0.58	2.37	0.91	3.08	0.78
s^{out}	**11.49**	0.61	3.63	0.79	1.61	0.88
k	**11.93**	0.54	2.58	0.70	0.52	0.44
k^{in}	**11.93**	0.52	1.19	0.88	1.41	0.71
k^{out}	**11.57**	0.61	4.34	0.70	0.98	0.66
bt	**11.37**	0.55	2.44	0.84	1.37	0.77
asy	3.14	0.98	**18.52**	1.97	2.46	1.69
μ^{asy}	3.32	0.99	**18.23**	2.01	2.80	1.82
σ^{asy}	4.91	0.59	2.44	1.47	**26.84**	3.06
dis	2.94	0.88	**18.50**	1.92	3.06	1.98
μ^{dis}	2.55	0.89	**18.12**	1.85	1.57	1.32
σ^{dis}	0.57	0.33	2.74	1.63	**30.61**	2.66
λ	49.56	1.16	27.14	0.54	13.25	0.95

ferent participation characteristic, and their equal relevance for variability, as measured by dispersion, is noticeable. Also, this suggests that these centrality metrics

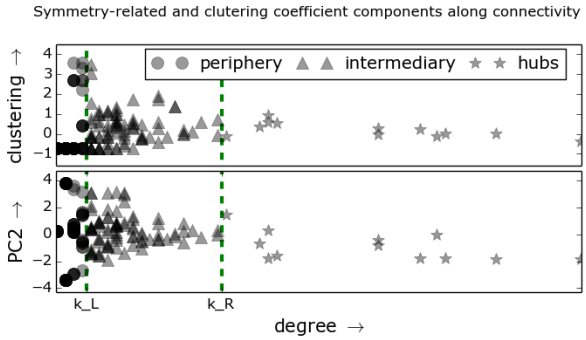

FIG. 4. The first plot exposes the well-known pattern of degree versus clustering coefficient, characterized by the higher clustering coefficient of lower degree vertices. The second plot shows the greater dispersion of the symmetry-related ordinates dominant in PC2. This greater dispersion suggests that symmetry-related metrics are more powerful, for characterizing interaction networks than the clustering coefficient, especially for hubs and intermediary vertices. This figure was reflects a snapshot of the LAU list with $ws = 1000$.

are equally adequate for characterizing the networks and the participants.

Evident in Table VIII and Figure 4, dispersion is greater in symmetry-related metrics than in clustering coefficient. As expected by basic complex network theory, peripheral vertices have low values of centrality measures and greater dispersion with respect to clustering coefficient. This reflects in the relevance of the symmetry-related metrics. We conclude that the symmetry met-

rics are more powerful, with respect to dispersion in the topological measures space, in characterizing interaction networks and their participants, than the clustering coefficient, especially for hubs and intermediary vertices. Interestingly, the clustering coefficient is always combined with the standard deviation of the asymmetry and disequilibrium of edges σ^{asy} and σ^{dis}.

Similar results are presented in Sections SII and SIV of the Supporting Information document for other email lists and other interaction networks. Larger variability was found among the latter networks, which motivated the use of interaction networks derived from email lists for benchmarks.

D. Types from Erdös sectors

Assigning a type to a participant raises important issues about the scientific cannon for human types and the potential for stigmatization and prejudice. The type of the Erdös sector to which a participant belongs can be regarded as implying a social type to the corresponding participant. In this case, the type of a participant changes both along time and as different networks are considered, despite the stability of the network, and therefore the potential for prejudice of such participant typology is attenuated[12]. In other words, an individual is a hub in a number of networks and peripheral in other networks, and even within the same network he/she most probably changes type along time[35].

The importance of this issue can be grasped by the consideration of static types derived from quantitative criteria. For example, in email lists with a small number of participants, the number of threads has a negative correlation with the number of participants. When the number of participants exceeds a threshold, the number of threads has a positive correlation with the number of participants. This finding is illustrated in Figure 5 and can also be observed in Table I. The attribution of types to individuals, derived from this analysis of the systems, has more potential for prejudice because the derived participant type is static. The types, in this latter case, fail to acknowledge that human individuals are not immutable entities.

Beyond the results discussed above and in previous sections, our main observations regarding the Erdös sectors and the implicit participant types, consistent with the literature[7], are that 1) hubs and intermediary participants usually have intermittent activity. Stable activity was found only in smaller communities. For instance, the MET list have stable hubs while all LAU, LAD and CPP present intermittent hubs. 2) Network structure seems to be most influenced by the activity of intermediary participants as they have less extreme roles than hubs and peripheral participants and can therefore connect to the sectors and other participants in a more selective and explicit manner.

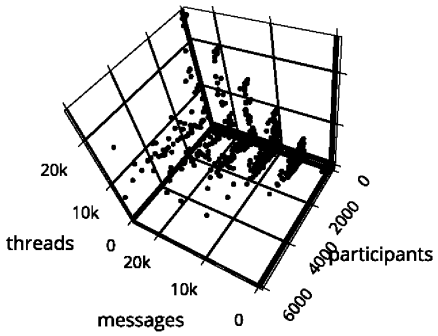

FIG. 5. A scatter plot of number of messages M versus number of participants N versus number of threads Γ for 140 email lists. Highest Γ are associated with low N. The correlation between N and Γ is negative for low values of N but positive otherwise. This negative correlation between N and Γ can also be observed in Table I. Accordingly, for $M = 20000$ messages, this inflection of correlation was found around $N = 1500$ and CPP, LAU, LAD, MET lists present smaller networks.

V. CONCLUSIONS

The temporal stability reported in this article is very counterintuitive. Some of the observed characteristics were expected from current scientific knowledge, such as the fat tail structure and the preponderance of centrality measures, but the very small standard deviations of principal components formation (see Sections III D and IV C), the presence of the Erdös sectors even in networks with few participants (see Sections III C and IV B), and the recurrent activity patterns along different timescales (see Sections III A and IV A), goes a step further in characterizing scale-free networks in the context of the interaction of human individuals. Furthermore, the importance of symmetry-related metrics, which surpassed that of clustering coefficient, with respect to dispersion of the system in the topological measures space, might add to the current understanding of key-differences between digraphs and undirected graphs in complex networks. Noteworthy is the also very stable fraction participants in each Erdös sector when the network reaches more than 200 participants. Benchmarks were derived from email list networks and the supplied analysis of networks from Facebook, Twitter and Participabr might ease hypothesizing about the generality of these characteristics. Selected tables and figures sum dozens of pages and are gathered in the Supporting Information document.

Further work should expand the analysis to include more types of networks and more metrics. The data and software needed to attain these results should also receive dedicated and in-depth documentation as they enable a greater level of transparency and work share, which is adequate for both benchmarking and specifically for the study of systems constituted by human individuals (see

Section II). The derived typology of hub, intermediary and peripheral participants has been applied for semantic web and participatory democracy efforts, and these developments might be enhanced to yield scientific knowledge[25]. Also, we plan to further explore and publish of the visualization and sonification processes used for this research[35,40] and the extreme linguistic differences found in each of the Erdös sectors[41].

ACKNOWLEDGMENTS

Financial support was obtained from CNPq (140860/2013-4, project 870336/1997-5), United Nations Development Program (contract: 2013/000566; project BRA/12/018) and FAPESP. The authors are grateful to the American Jewish Committee for maintaining an online copy of the Adorno book used on the epigraph[12], to Gmane creators and maintainers for the public email list data, to the communities of the email lists and other groups used in the analysis, and to the Presidency of the Brazilian Republic for keeping Participabr code and data open. We are also grateful to developers and users of Python scientific tools.

[1] J. L. Moreno, "Who shall survive?: A new approach to the problem of human interrelations." The Journal of Social Psychology **6**, 388–393 (1935).
[2] M. Newman, *Networks: an introduction* (Oxford University Press, 2010).
[3] B. Latour, "Reassembling the social. an introduction to actor-network-theory," Journal of Economic Sociology **14**, 73–87 (2013).
[4] C. Bird, A. Gourley, P. Devanbu, M. Gertz, and A. Swaminathan, "Mining email social networks," in *Proceedings of the 2006 international workshop on Mining software repositories* (ACM, 2006) pp. 137–143.
[5] A. Vázquez, J. G. Oliveira, Z. Dezsö, K.-I. Goh, I. Kondor, and A.-L. Barabási, "Modeling bursts and heavy tails in human dynamics," Physical Review E **73**, 036127 (2006).
[6] B. Ball and M. E. Newman, "Friendship networks and social status," arXiv preprint arXiv:1205.6822 (2012).
[7] G. Palla, A.-L. Barabási, and T. Vicsek, "Quantifying social group evolution," Nature **446**, 664–667 (2007).
[8] E. A. Leicht, G. Clarkson, K. Shedden, and M. E. Newman, "Large-scale structure of time evolving citation networks," The European Physical Journal B **59**, 75–83 (2007).
[9] B. Travençolo and L. d. F. Costa, "Accessibility in complex networks," Physics Letters A **373**, 89–95 (2008).
[10] M. E. Newman, "Modularity and community structure in networks," Proceedings of the National Academy of Sciences **103**, 8577–8582 (2006).
[11] N. L. Quenk, *Essentials of Myers-Briggs type indicator assessment*, Vol. 66 (Wiley. com, 2009).
[12] T. W. Adorno, E. Frenkel-Brunswik, D. J. Levinson, and R. N. Sanford, "The authoritarian personality." (1950).
[13] K. Gergen and M. Gergen, *Historical social psychology* (Psychology Press, 2014).
[14] R. Albert and A.-L. Barabási, "Topology of evolving networks: local events and universality," Physical review letters **85**, 5234 (2000).
[15] K. Marek-Spartz, P. Chesley, and H. Sande, "Construction of the gmane corpus for examining the diffusion of lexical innovations," (2012).
[16] J.-P. Onnela, S. Arbesman, M. C. González, A.-L. Barabási, and N. A. Christakis, "Geographic constraints on social network groups," PLoS one **6**, e16939 (2011).
[17] V. Palchykov, K. Kaski, J. Kertész, A.-L. Barabási, and R. I. Dunbar, "Sex differences in intimate relationships," Scientific reports **2** (2012).
[18] R. Fabbri, "Python package to observe time stability in the gmane database," (2015), https://pypi.python.org/pypi/gmane.
[19] Wikipedia, "Gmane — Wikipedia, the free encyclopedi," (2013), online; accessed 27-October-2013.
[20] Gmane.linux.audio.users is list ID in Gmane.
[21] Gmane.linux.audio.devel is list ID in Gmane.
[22] Gmane.comp.gcc.libstdc++.devel is list ID in Gmane.
[23] Gmane.politics.organizations.metareciclagem is list ID in Gmane.
[24] R. Fabbri, "Python package to analyze the gmane database," (2015), https://pypi.python.org/pypi/gmane.
[25] R. Fabbri, "Content extraction through api from the Brazilian Federal Portal of Social Participation and its tools to a social participation cloud," Tech. Rep. (United Nations Development Programme and Brazilian Presidency of the Republic, 2014) https://github.com/ttm/pnud5/blob/master/latex/produto.pdf?raw=true.
[26] R. Fabbri, "Data from Participa.br, Cidade Democrtica and AA, in XML/RDF and Turtle/RDF," (2014), http://datahub.io/organization/socialparticipation.
[27] M. Woelfle, P. Olliaro, and M. H. Todd, "Open science is a research accelerator," Nature Chemistry **3**, 745–748 (2011).
[28] R. Fabbri, "What are you and i? [anthropological physics fundamentals]," (2015), https://www.academia.edu/10356773/What_are_you_and_I_anthropological_physics_fundamentals_.
[29] D. C. Antunes, R. Fabbri, and M. M. Pisani, "Anthropological physics and social psychology in the critical research of networks," CSDC'15 online conference, Conference on Complex Systems, https://www.youtube.com/watch?v=oeOKYc3-nbM), year=2015,.
[30] E. A. Leicht and M. E. Newman, "Community structure in directed networks," Physical review letters **100**, 118703 (2008).
[31] M. Newman, "Community detection and graph partitioning," arXiv preprint arXiv:1305.4974 (2013).
[32] U. Brandes, "A faster algorithm for betweenness centrality*," Journal of Mathematical Sociology **25**, 163–177 (2001).
[33] M. O. Jackson, "Social and economic networks: Models and analysis," (2013), https://class.coursera.org/networksonline-001.
[34] I. Jolliffe, *Principal component analysis* (Wiley Online Library, 2005).
[35] R. Fabbri, "Video visualizations of email interaction network evolution," (2013-5), https://www.youtube.com/playlist?list=PLf_EtaMqu3jVodaqDjN7yaSgsQx2Xna3d.
[36] R. Fabbri, "Image gallery of email interaction networks." (2013), http://hera.ethymos.com.br:1080/redes/python/autoRede/gmane.linux.audio.devel_3000-4200-280/.
[37] R. Fabbri, "Online gadget for making email interaction network images, gml files and measurements," (2013), http://hera.ethymos.com.br:1080/redes/python/autoRede/escolheRedes.php.
[38] Numpy version 1.8.2, "random.randint" function, was used for simulations, algorithms in https://pypi.python.org/pypi/gmane.
[39] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," Physics reports **424**, 175–308 (2006).
[40] R. Fabbri, "Versinus: a visualization method for graphs in evolution," arXiv preprint arXiv:1412.7311 (2013), http://arxiv.org/abs/1412.7311.
[41] R. Fabbri, "A connective differentiation of textual production in interaction networks," (2013), http://arxiv.org/abs/1412.7309.