Temporal stability in human interaction networks: sector sizes, topological prominence and time activity

Renato Fabbri,[1, a)] Ricardo Fabbri,[b)] Deborah C. Antunes,[c)] Marilia M. Pisani,[d)] Leonardo Paulo 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: 4 October 2015)

In this study we report on stable aspects 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 in various scales. Notably, the activity in 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 centrality measurements (degree, strength and betweenness), followed by symmetry-related metrics and then clustering coefficient. The activity of participants followed the expected scale-free trace, thus yielding three classes of vertices by comparison against the Erdös-Rényi model, namely hubs, intermediary and periphery. The relative size of these three sectors did not vary with time and was essentially the same for all email lists. Typically, 3-12% of the vertices are hubs, 15-45% are intermediary and the remainder are peripheral vertices. Similar results for the distribution of participants in the three categories and for the relative importance of the topological metrics were obtained for 12 additional networks from Facebook, Twitter and Participabr. Consistent with expectations from the literature, these properties may be general for human interaction networks, which has important implications in establishing a typology based on quantitative criteria.

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

'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

Studies on human interaction networks have started long before modern computers, dating back to 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 in email lists was the most convenient for deriving results and for benchmarking, but networks from Facebook, Twitter and Participabr were also considered. Using a timeline of activity snapshots with a constant number of contiguous messages, we found remarkable stability for several of the network properties. Namely, 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 sectors are unshaken. This is not an intuitive result, given that participants transition in network structure constantly. Because these properties were shared by networks from various sources, and are consistent with the complex networks literature, we advocate that the con-

[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

clusions might be valid for general classes of interaction networks. In particular, this allows us to bridge the gap between data analysis and social sciences in the discussion of types of networks and participants. Noteworthy, typologies are the canon of scientific literature for classification of human agents, with pragmatic standards[11] and critical paradigms[12,13].

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

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 the frequency of interactions, which can either be 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 are web browsing and library loans[5]), and different linguistic traces were related to weak and strong ties[15].

Unreciprocated edges often exceed 50% in the analyzed networks, which matches empirical evidence from the literature[8] and 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 this was not considered in the present article 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 20×10^3 email lists 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 with metadata of its messages, including sent time, place, sender name, and sender email address. The GMANE usage in scientific research is reported in studies of isolated lists and of lexical innovations[4,15].

We analyzed many email lists (and data from Twitter, Facebook and Participabr) and selected four of them in order to make a thorough analysis, from which general properties can be inferred. These lists are:

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

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

- Development list for the standard C++ library[22], with computer programmers from different countries. English is the most used language. Abbreviated as CPP from now on.

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

The first 20,000 messages of each list were considered, with 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 additional 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.

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[34]. Data from Participabr was used from the linked data/semantic web RDF triples reported in[35] and available in[36]. Script 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[37] and lightens ethical and moral issues of researching systems constituted of human individuals[32,33].

III. METHODS

A. Time activity statistics

Messages were counted over time as histograms in terms of seconds, minutes, hours, days of the week, days of the month, and months of the year. Most standard measures of location and dispersion, such as usual mean and standard deviation, hold little meaning in a compact Riemannian manifold. Equivalent measures were taken using circular statistics, in which each measurement t (data point) is represented as a unity 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 and the highest 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 I 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

Interaction networks can be modeled both weighted or unweighted, both directed or undirected[4,24,25]. Networks in this paper are directed and weighted, the most informative of trivial 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 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 human social information 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].

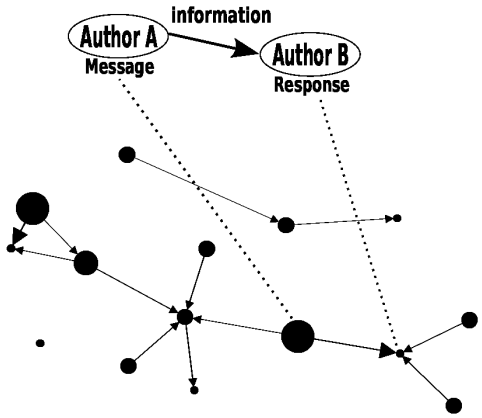

FIG. 1. The formation of interaction networks from exchanged messages. Each vertex represents a participant. A reply message from author B to a message from author A is regarded as evidence that B received information from A and yields a directed edge. Multiple messages add "weight" to a directed edge.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

1. Topological metrics

The topology of the networks was studied using 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: 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. The standard clustering coefficient for undirected graphs was used.
- Betweenness centrality bt_i: fraction of geodesics that contain vertex i. The betweenness centrality index considered directions and weight, as specified in[27].

In order to capture symmetries in the activity of participants, the following metrics were introduced for a vertex i:

- Asymmetry: $asy_i = \frac{k_i^{in} - k_i^{out}}{k_i}$.
- Mean of asymmetry of edges: $\mu_i^{asy} = \frac{\sum_{j \in J_i} e_{ji} - e_{ij}}{|J_i| = k_i}$, where e_{xy} is 1 if there is an edge from x to y, and 0 otherwise. J_i is the set of neighbors of vertex i, and $|J_i| = k_i$ is the number of neighbors of vertex i.
- Standard deviation of asymmetry of edges: $\sigma_i^{asy} = \sqrt{\frac{\sum_{j \in J_i} [\mu_i^{asy} - (e_{ji} - e_{ij})]^2}{k_i}}$.
- Disequilibrium: $dis_i = \frac{s_i^{in} - s_i^{out}}{s_i}$.
- Mean of disequilibrium of edges: $\mu_i^{dis} = \frac{\sum_{j \in J_i} \frac{w_{ji} - w_{ij}}{s_i}}{k_i}$, 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: $\sigma_i^{dis} = \sqrt{\frac{\sum_{j \in J_i} [\mu_i^{dis} - \frac{(w_{ji} - w_{ij})}{s_i}]^2}{k_i}}$.

These measures are used both for the Erdös sectioning (described in Secion III D) and for performing PCA (as described in Section III C).

C. Principal Component Analysis of topological metrics

Principal Component Analysis (PCA) is a well documented technique[26] and was used to acquire knowledge about: 1) which metrics contribute to each principal component and in which proportion; 2) how much of the dispersion is concentrated in each component; 3) expected values and dispersions for these quantities over various networks.

Let $\mathbf{X} = \{X[i,j]\}$ be a matrix of all vertices i and the respective values for each metric j, $\mu_X[j] = \frac{\sum_i X[i,j]}{I}$ 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. Be $\mathbf{D} = \{D[k]\}$ 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 between 60 and 95% of the dispersion and reveal patterns for a sound 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:

$$\begin{aligned} \mu_{V'}[j,k] &= \frac{\sum_{l=1}^{L} V'[j,k,l]}{L} \\ \sigma_{V'}[j,k] &= \sqrt{\frac{\sum_{l=1}^{L} (\mu_{V'} - V'[j,k,l])^2}{L}} \\ \mu_{D'}[k] &= \frac{\sum_{l=1}^{L} D'[k,l]}{L} \\ \sigma_{D'}[k] &= \sqrt{\frac{\sum_{l=1}^{L} (\mu_{D'} - D'[k,l])^2}{L}} \end{aligned} \tag{3}$$

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.

D. Erdös sectioning

In a scale-free network, the peripheral, intermediary and hubs sectors can be derived from a comparison against an Erdös-Rényi network with the same number of edges and vertices[28], as depicted in Figure 2. We shall refer to this procedure as *Erdös sectioning*, with the resulting sectors being referred to as *Erdös sectors* or *primitive sectors*.

The degree distribution $\widetilde{P}(k)$ of an ideal 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:

$$\widetilde{P}(k) < P(k) \Rightarrow \text{k is intermediary degree} \tag{4}$$

If $\mathcal{N}_f(N, z)$ is directed and has no self-loops, the probability 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

$$P(k) = \binom{2(N-1)}{k} p_e^k (1-p_e)^{2(N-1)-k} \tag{5}$$

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 at pure chance (e.g. without preferential attachment and as in the Erdös Rényi model) are correctly associated to the hubs sector. Vertices with very few connections, which are way more abundant than expected by pure chance, are assigned to the periphery. Vertices with degree values predicted as the most abundant if connections are created by pure chance, near the average, and less frequent in the real network, are classified as intermediary.

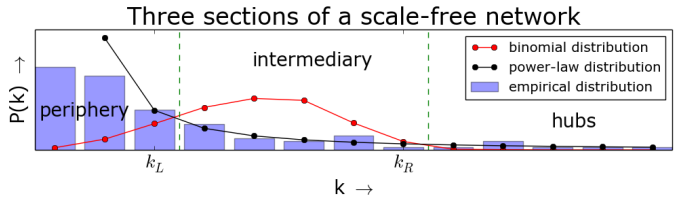

FIG. 2. Classification of vertices by comparing degree distributions. The binomial distribution of the Erdös Rényi network exhibit more intermediary vertices, while a scale-free network, associated with the power-law distribution, has more peripheral and hub vertices. The sector borders are defined with 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 hubs 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 throughout exposition of the method is exposed in Section III D.

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 $\Delta_i \cap \Delta_j = \emptyset \; \forall \; i \neq j$:

$$\Delta_i = \Bigg\{ k \;\Big|\; \overline{\Delta}_{i-1} < k \leq j \text{ and } \Bigg[N - \sum_{k=0}^{\overline{\Delta}_{i-1}} \eta_k < \eta \text{ or } \Bigg[\sum_{k=\overline{\Delta}_{i-1}+1}^{j} \eta_k \geq \eta \text{ and } \Bigg(\sum_{k=\overline{\Delta}_{i-1}+1}^{j-1} \eta_k < \eta \text{ or } j = \overline{\Delta}_{i-1} + 1 \Bigg) \Bigg] \Bigg] \Bigg\} \tag{6}$$

where η_k is the number of vertices with degree k, while $\overline{\Delta}_{i-1} = max(\Delta_{i-1})$, and $\overline{\Delta}_0 = -1$. Equation 4 can now be written in the form:

$$\sum_{x=min(\Delta_i)}^{max(\Delta_i)} \widetilde{P}(x) < \sum_{x=min(\Delta_i)}^{max(\Delta_i)} P(x) \Rightarrow \\ \Rightarrow \Delta_i \text{ holds intermediary degree values.} \tag{7}$$

If 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 for comparison, with $\overline{w} = 2\frac{z}{\sum_i s_i}$ the average weight of an edge and s_i the strength of vertex i. For in and out degrees (k^{in}, k^{out}) comparison of the real network should be made with

$$\hat{P}(k^{way}) = \binom{N-1}{k^{way}} p_e^k (1-p_e)^{N-1-k^{way}}, \tag{8}$$

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

$$\begin{aligned} e_{\gamma,1} &= \{\, i \mid \overline{k}_{\gamma,L} \geq \overline{k}_{\gamma,i}\} \\ e_{\gamma,2} &= \{\, i \mid \overline{k}_{\gamma,L} < \overline{k}_{\gamma,i} \leq \overline{k}_{\gamma,R}\} \\ e_{\gamma,3} &= \{\, i \mid \quad\; \overline{k}_{\gamma,i} < \overline{k}_{\gamma,R}\}, \end{aligned} \tag{9}$$

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

$$\begin{aligned}
\overline{k}_{1,i} &= k_i \\
\overline{k}_{2,i} &= k_i^{in} \\
\overline{k}_{3,i} &= k_i^{out} \\
\overline{k}_{4,i} &= \frac{s_i}{\overline{w}} \\
\overline{k}_{5,i} &= \frac{s_i^{in}}{\overline{w}} \\
\overline{k}_{6,i} &= \frac{s_i^{out}}{\overline{w}}
\end{aligned} \tag{10}$$

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, compound criteria were defined for conveniently analysing networks with a low number of messages, such as in Section III of the Supporting Information. For example, a very stringent criterion can be used, according to which a vertex is only regarded as pertaining to a sector if it is so for all the metrics. 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 100%, which is indicated by a black line in Figure 4.
- 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 members may exceed 100%, which is indicated by a black line in Figure 4.
- 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 classified as intermediary if they belong to this category for any of the metrics. Peripheral vertices will then be those which were not classified as hub or intermediary with any of the metrics.
- Exclusivist externals C_5: vertices are only hubs if they are classified as such according to all the metrics. The remaining vertices are classified as peripheral if they fall into the periphery or hub classes for any metric. The rest of the nodes are classified as intermediary.
- Inclusivist externals C_6: hubs are vertices classified as hubs according to any metric. The remaining vertices will be peripheral if they are classified as such according to any metric. The rest of the vertices will be intermediary vertices.

Using equations 9, these compound criteria C_δ, with δ integer in the interval $1 \leq \delta \leq 6$, can be described as:

$$\begin{aligned}
C_1 &= \{c_{1,\phi} = \{i \mid i \ \in e_{\gamma,\phi},\ \forall\, \gamma\}\} \\
C_2 &= \{c_{2,\phi} = \{i \mid \exists\ \gamma : i \in e_{\gamma,\phi}\}\} \\
C_3 &= \{c_{3,\phi} = \{i \mid i \ \in e_{\gamma,\phi'},\ \forall\, \gamma,\ \forall\, \phi' \geq \phi\}\} \\
C_4 &= \{c_{4,\phi} = \{i \mid i \ \in e_{\gamma,\phi'},\ \forall\, \gamma,\ \forall\, \phi' \leq \phi\}\} \\
C_5 &= \{c_{5,\phi} = \{i \mid i \ \in e_{\gamma,\phi'},\ \forall\, \gamma, \\
&\qquad \forall\, (\phi'+1)\%4 \leq (\phi+1)\%4\}\} \\
C_6 &= \{c_{6,\phi} = \{i \mid i \ \in e_{\gamma,\phi'},\ \forall\, \gamma, \\
&\qquad \forall\, (\phi'+1)\%4 \geq (\phi+1)\%4\}\}
\end{aligned} \tag{11}$$

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.

E. Evolution and visualization of the networks

The evolution of the networks was observed within sequences of snapshots. with a fixed number of messages, Each of these sequences had a fixed number of messages, the window size ws, which was considered with different shifts in the message timeline to obtain snapshots. In each snapshot was performed both the PCA with topological metrics and the Erdös sectioning. The values of ws employed were 50, 100, 200, 400, 500, 800, 1000, 2000, 2500, 5000 and 10000. Variations in the number of vertices, edges and other network characteristics are exposed in Section III of the Supporting Information document.

Networks were visualized with animations, sonifications, image galleries and online gadgets developed specifically for this research[29–31]. Such *audiovisualizations* were crucial to guide the research into the most important features of network evolution. Furthermore, the size of the three Erdös sectors could be visualized in a timeline fashion. Visualization of network structure was especially useful in the initial inspection of data and of the structures derived from the email lists.

IV. RESULTS AND DISCUSSION

A. Activity along time

The observed activity along time, in terms of seconds, minutes, hours, days and months, is practically the same for all lists. Histograms in each time scale were computed as were circular average and dispersion values. We

TABLE II. The rescaled circular mean θ'_μ and the circular dispersion $\delta(z)$ described in Section III A. This typical table was constructed using all LAD list messages, and the results are the same for other lists, as shown in Section I A of the Supporting Information document. The most uniform distribution of activity was found in seconds and minutes, where the mean has little meaning. Hours of the day exhibited the most concentrated activity (lowest $\delta(z)$), with mean between 14h and 15h ($\theta' = -9.61$). Weekdays, month days and months have mean near zero (i.e. near the beginning of the week, month and year) and high dispersion.

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

provide detailed values in Table II-VI because they can be used for characterizing nodes (participants) in these and other networks, and networks themselves. For example, they may serve for identification of outliers in a community.

In the scale of seconds and minutes, activity obeys a homogeneous pattern, with the messages being slightly more evenly distributed in all lists than in simulations using 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 appeared around 11am with the most active period being the afternoon. 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 being 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.

TABLE III. Activity percentages along the hours of the day for the CPP list. Nearly identical distributions are found on other lists as shown in Section I B 1 of the Supporting Information document. Higher activity was observed between noon and 6pm, followed by the time period between 6pm and midnight. Around 2/3 of the whole activity takes place from noon to midnight. Nevertheless, the activity peak occurs around midday, with a slight skew toward one hour before noon.

<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="2">9.30</td><td rowspan="4">10.67</td><td rowspan="8">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="2">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><td rowspan="3">21.03</td></tr>
<tr><td>9h</td><td>4.52</td><td rowspan="2">18.75</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="2">18.95</td><td rowspan="3">25.05</td><td rowspan="4">37.63</td><td rowspan="8">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="2">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="2">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 the days of the week for the four email lists. Higher activity was observed during weekdays, with a decrease of activity on weekends of at least one third and two thirds in extreme cases.

	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

B. Stability of principal components and the prevalence of symmetry over clusterization for dispersion

The contribution of each metric to the variance is very similar for all the networks and along time. In applying PCA to the snapshots, the contribution of each metric to the principal components presents very small standard deviation. Table VII exemplifies the formation of principal components for the MET email list. Similar results are presented in Sections II and IV of the Supporting Information document for the other email lists (benchmarks) and other interaction networks, with the consideration of strategic combinations of metrics.

TABLE V. Activity in the days along the month for the MET list. Nearly identical distributions are found on other lists as indicated in Section I 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.

	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 of the months along the year from LAU list. Activity is usually concentrated in Jun-Aug and/or in Dec-Mar (see Section I B 4 of the Supporting Information). These observations fit academic calendars, vacations and end-of-year holidays.

	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				

TABLE VII. 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 80% 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

The first principal component is an average of centrality metrics: degrees, strengths and betweenness centrality. Therefore, all of these centrality measurements are equally important for characterizing the networks. On one hand, the relevance of all centrality metrics is not surprising since they are highly correlated, e.g.the degree and strength have Spearman correlation coefficient $\in [0.95, 1]$ and Pearson coefficient $\in [0.85, 1)$ for $ws > 1000$. On the other hand, each of these metrics is related to a different participation characteristic, and their equal relevance is noticeable. The clustering coefficient is presented in almost perfect orthogonality to centrality metrics.

Dispersion was more prevalent in symmetry-related metrics than for the clustering coefficient, as indicated in Table VII. This is also illustrated in Figure 3, where each vertex is colored according to the sector they belong to. As expected, peripheral vertices have very low values in the first component (centrality related) and greater dispersion in the third component (clustering related). The PCA plot in the third system of Figure 3, where all metrics are considered, reflects the relevance of the symmetry-related metrics for the variance. We conclude that the symmetry metrics is more useful, with respect to the dispersion in the topological space, in characterizing interaction networks (and their participants) than the clustering coefficient, especially for hubs and intermediary vertices.

The relative importance of the topological metrics

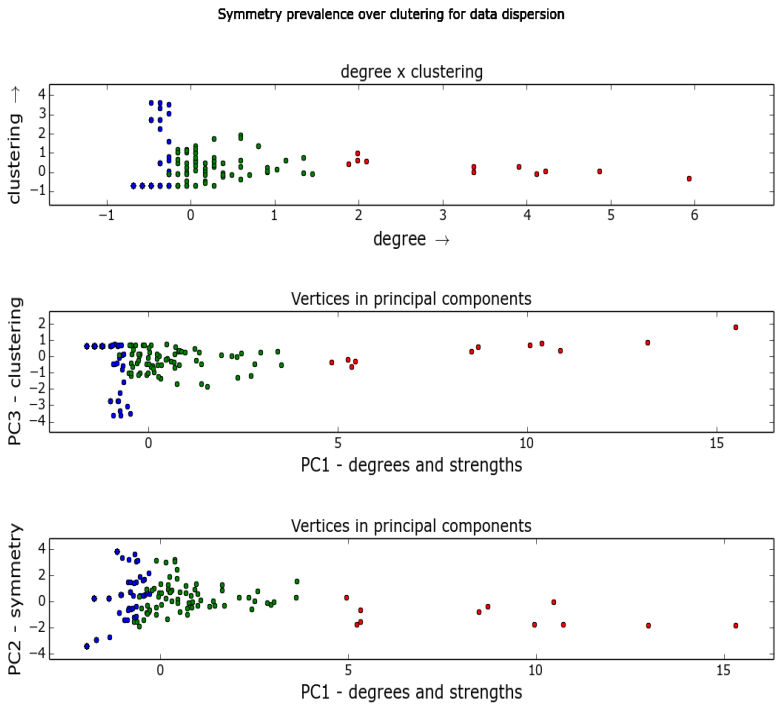

FIG. 3. The first plot shows degree versus clustering coefficient. This typical pattern is well known with high clustering more incident in vertices with lower degrees. The second plot is analogous but the abscissas is an average of centrality metrics. The third plot exhibits the greater dispersion of the symmetry-related ordinates. This greater dispersion suggests that symmetry-related metrics are more powerful, with respect to dispersion, for characterizing interaction networks than the clustering coefficient, especially for hubs and intermediary vertices. This figure was obtained with a snapshot of the LAU list in a window size of $ws = 1000$ messages. Similar structures were observed in all window sizes $ws \in [500, 10000]$, in networks derived from email lists, and in networks from Facebook, Twitter and Participabr, which suggests a common relationship between the metrics of degrees, strengths and betweenness centrality, the symmetry-related metrics and clustering coefficient.

was also observed for the additional 12 networks from Facebook, Twitter and Participabr. With the exception of two of these networks, the overall behavior was maintained in that centrality measurements were found prevalent in the first principal component, followed by symmetry-related metrics on the second principal component and then clustering coefficient on the third principal component. The results are given in Tables S31, S32, S33, S34 of the Supporting Information document. Larger variability was found among such networks, which motivated the use of interaction networks derived from email lists for benchmarks.

C. Scalable fat-tail structure: constancy of membership fractions in each Erdös sector

The distribution of vertices in the hubs, intermediary, periphery Erdös sectors is remarkably stable along time, provided that a sufficiently large sample of 200 or more messages is considered. Moreover, the same distribution applies to the networks of all email lists analyzed, as demonstrated in Figure 4 and in Section III 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, as confirmed by the data in Table VIII for the distribution of activity among participants.

Typically, $\approx [3-12]\%$ of the vertices are hubs, $\approx [15-45]\%$ are intermediary and $\approx [44-81]\%$ are peripheral, which is consistent with the literature[39]. These results hold for the total, in and out degrees and strengths. Stable distributions can also be obtained for 100 or less messages if classification of the three sectors is performed with one of the compound criteria established in Section III D. The networks hold their 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 III 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 obtained the Erdös sectors of 12 networks from Facebook, Twitter and Participabr. The results are given in Table S30 of the Supporting Information, which indicate that the percentages of hubs, intermediary and periphery nodes are essentially the same as for the email lists.

TABLE VIII. 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 ($1Q$) shows the minimum percentage of participants responsible for at least 25% of total messages. Similarly, the column for the first three quartiles $1-3Q$ gives the minimum percentage of participants responsible for 75% of total messages. The last decile $-10D$ column exposes the maximum percentage of participants responsible for 10% of messages.

list	hub	$1Q$	$1-3Q$	$-10D$
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%)

D. Types from Erdös sectors

A sector to which a vertex belongs can be regarded as yielding a type to the corresponding participant. Assigning a type to a participant inevitably raises an important question regarding the possible stigmatization. We take the view that the participation typology inherent in the Erdös sectors is not stigmatizing because the type of an individual changes along time and as different networks are considered[12]. That is to say, an individual is a hub in a number of networks and peripheral in other networks, and even within a network he/she probably changes type along time. Indeed, we did observe often transitions of participants from one sector to another. The typology proposed here bridges exact and human sciences and may be enriched with concepts from other typologies, such as Meyer-Briggs, Pavlov or the authoritarian types of the F-Scale[12].

We analyzed the time evolution of the networks using visualization tools developed for this research[40,41] and inspected the raw data to infer the main characteristics of each type. Our main observations may be summarized as follows:

- Core hubs usually have intermittent activity. Very stable activity was found on MET hubs, which is consistent with the literature where greater stability occurs in smaller communities[7].
- Typically, the activity of hubs is trivial: they interact as much as possible, in every occasion with everyone. The activity of peripheral vertices also follows a simple pattern: they interact very rarely, in very few occasions. Therefore, intermediary vertices seem responsible for the network structure. Intermediary vertices may exhibit preferential communication to peripheral, intermediary, or hub vertices; can be marked by stable communication partners; can involve stable or intermittent patterns of activity.
- Some of the most active participants receive many responses with relative few messages sent, and rarely are top hubs. These seem as authorities and contrast with participants that respond much more than receive responses.
- The most obvious community structure, as observed by a high clustering coefficient, i.e. members known each other often, is found mostly in peripheral and intermediary sectors.

With regard to the networks as the whole objects of analysis, we were able to observe a negative correlation between the number of threads and the number of participants. When the number of participants exceeds a threshold, the number of threads displays a positive correlation with the number of participants. This finding is illustrated in Figure 5 and can also be observed in Table I. Obviously, network types can be derived from such results, which was not attempted here but left for the reader and future work.

V. CONCLUSIONS

The most important result from the analysis of time evolution of the four email lists is certainly the time-independence observed not only for the activity but also for the properties of the networks themselves. For example, the relative fractions of participants classified as hubs, intermediary and peripheral vertices remained practically constant along time across all email lists studied. Furthermore, the PCA analysis of the topological metrics characterizing the networks also indicated that the contribution of each metric did not vary in time. Centrality metrics were found to be the most relevant to characterize the network topology, followed by symmetry-related metrics, which were more relevant, with respect to variance, than clustering.

A systematic study of the activity of participants belonging to the three distinct Erdös sectors indicated simple patterns for hubs and peripheral vertices, while the network structure was governed by the intermediary vertices. These properties were shared by all email lists and were time-independent, being consistent with the literature. Moreover, both the distribution of Erdös sectors and the contribution from the metrics to the PCA were found to apply to networks from Facebook, Twitter and Participabr. We may therefore consider the classification of agents into Erdös sectors as a first step leading to a human typology which bridges exact sciences, with

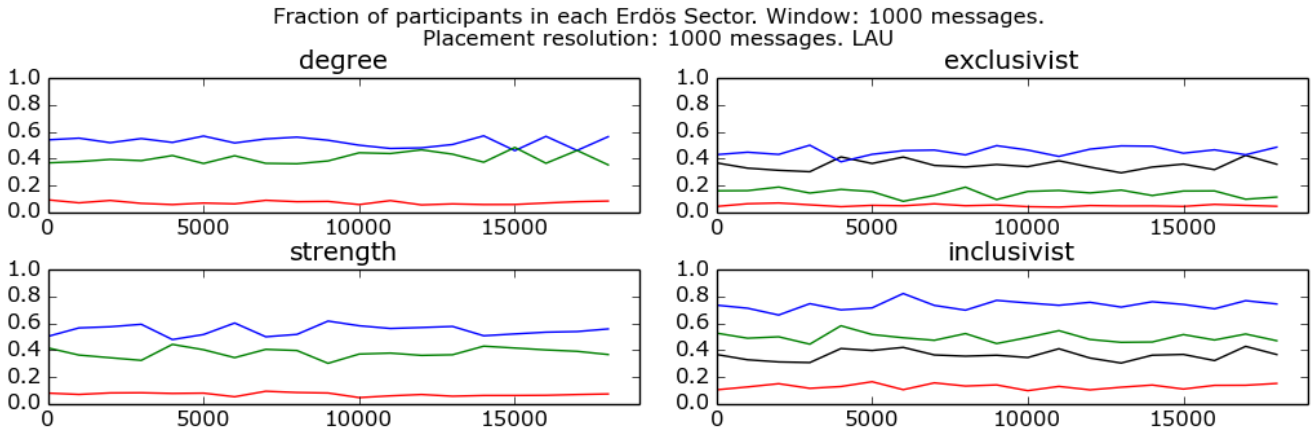

FIG. 4. Fractions of agents in each Erdös sector, where the fractions for hubs, intermediary and peripheral vertices are represented in red, green and blue, respectively. We used two simple criteria, namely degree and strength, for the graphics on the left. For the graphics on the right we employed the Exclusivist and Inclusivist compound criteria, with black lines representing the fraction of vertices without class and with more than one class, respectively. See Section III of Supporting Information for a collection of such timeline figures with all simple and compound criteria and metrics. Table S30, also from Supporting Information, presents these fractions of agents in snapshots of networks from Facebook, Twitter and Participabr.

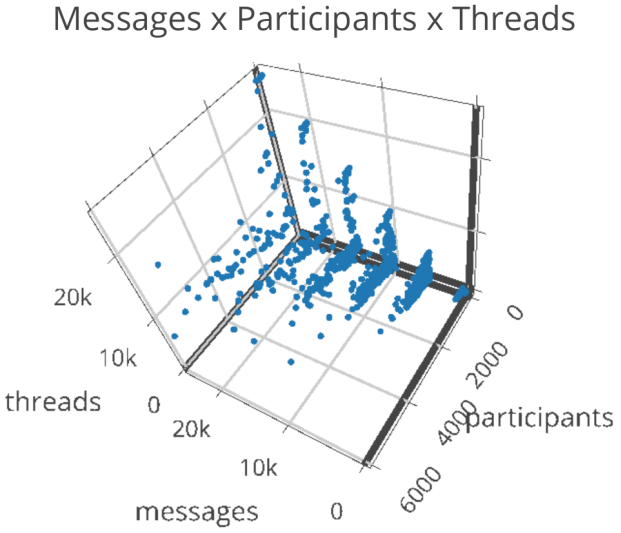

FIG. 5. A scatter plot of number of messages (M) versus number of participants (N) versus number of threads (Γ) for 140 email lists. Highest number of threads are found in lists with few participants. The correlation between N and Γ is negative for low values of N but positive otherwise. This negative correlation between N and Γ can also be observed in Table I. For $M = 20000$ messages, positive correlation of N and Γ is present mostly above 1500 participants. All LAU, LAD, MET lists present smaller networks.

quantitative procedures for the classification, and human sciences, where there is a legacy in the observation of human types.

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, and to the communities of the email lists and other groups used in the analysis, and to the Brazilian Presidency of the 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]E. A. Leicht and M. E. Newman, "Community structure in directed networks," Physical review letters **100**, 118703 (2008).
[25]M. Newman, "Community detection and graph partitioning," arXiv preprint arXiv:1305.4974 (2013).
[26]I. Jolliffe, *Principal component analysis* (Wiley Online Library, 2005).
[27]U. Brandes, "A faster algorithm for betweenness centrality*," Journal of Mathematical Sociology **25**, 163–177 (2001).
[28]M. O. Jackson, "Social and economic networks: Models and analysis," (2013), https://class.coursera.org/networksonline-001.
[29]R. Fabbri, "Video visualizations of email interaction network evolution," (2013-5), https://www.youtube.com/playlist?list=PLf_EtaMqu3jVodaqDjN7yaSgsQx2Xna3d, https://www.youtube.com/playlist?list=PLf_EtaMqu3jWYQiJZYhVlJVngb7vsf6na, https://www.youtube.com/playlist?list=PLf_EtaMqu3jVb7CTt59t3ZnrmXuG0N3cO, https://www.youtube.com/playlist?list=PLf_EtaMqu3jVFS_AJZm_HuO9pywnSWaNF, https://www.youtube.com/playlist?list=PLf_EtaMqu3jU-1j4jiIUiyMqyVSzIYeh6, https://www.youtube.com/playlist?list=PLf_EtaMqu3jUZpAX3cKPC5J0t3q836CLy, https://www.youtube.com/playlist?list=PLf_EtaMqu3jUYO_XfJdqQELdbFnpqYEfb.
[30]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/.
[31]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.
[32]R. Fabbri, "What are you and i? [anthropological physics fundamentals]," (2015), https://www.academia.edu/10356773/What_are_you_and_I_anthropological_physics_fundamentals_.
[33]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,.
[34]R. Fabbri, "Python package to analyze the gmane database," (2015), https://pypi.python.org/pypi/gmane.
[35]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.
[36]R. Fabbri, "Data from Participa.br, Cidade Democrtica and AA, in XML/RDF and Turtle/RDF," (2014), http://datahub.io/organization/socialparticipation.
[37]M. Woelfle, P. Olliaro, and M. H. Todd, "Open science is a research accelerator," Nature Chemistry **3**, 745–748 (2011).
[38]Numpy version 1.6.1, "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, "A connective differentiation of textual production in interaction networks," (2013), http://arxiv.org/abs/1412.7309.
[41]R. Fabbri, "Versinus: a visualization method for graphs in evolution," arXiv preprint arXiv:1412.7311 (2013), http://arxiv.org/abs/1412.7311.