\documentclass[%  
 aip,  
 jmp,%  
 amsmath,amssymb,  
%preprint,%  
 reprint,%  
%author-year,%  
%author-numerical,%  
]{revtex4-1}  
  
\usepackage{graphicx}% Include figure files  
\usepackage{grffile}  
\usepackage{dcolumn}% Align table columns on decimal point  
\usepackage{bm}% bold math  
%\usepackage[mathlines]{lineno}% Enable numbering of text and display math  
%\linenumbers\relax % Commence numbering lines  
\usepackage{multirow}  
\usepackage{color} % for the notes  
\usepackage{etex}  
\reserveinserts{58}  
\usepackage{morefloats}  
\usepackage{hyperref}  
  
\maxdeadcycles=1000  
  
\begin{document}  
  
\preprint{XXXXX (preprint)}  
  
\maxdeadcycles=1000

\begin{document}

\preprint{XXXXX (preprint)}

\title{ On the topology of networks deriving from email lists}% Force line breaks with \\

\author{Renato Fabbri}%  
 \homepage{http://ifsc.usp.br/~fabbri/}  
 \email{fabbri@usp.br}  
 \affiliation{   
S\~ao Carlos Institute of Physics, University of S\~ao Paulo (IFSC/USP)%\\This line break forced with \textbackslash\textbackslash  
}%  
%  
%\author{Camilo Akimushkin Valencia}  
% \homepage{http://ifsc.usp.br/}  
% \email{akimushkin@ifsc.usp.br}  
% \altaffiliation[Also at ]{IFSC-USP}%Lines break automatically or can be forced with \\  
%  
  
\author{Vilson Vieira da Silva Junior}  
 \homepage{http://automata.cc/}  
 \email{vilson@void.cc}  
 \altaffiliation[Also at ]{IFSC-USP}%Lines break automatically or can be forced with \\  
  
\author{Ricardo Fabbri}  
 \homepage{http://www.lems.brown.edu/~rfabbri/}  
 \email{rfabbri@iprj.uerj.br}  
 \altaffiliation{  
Instituto Polit\'ecnico, Universidade Estadual do Rio de Janeiro (IPRJ)  
}%Lines break automatically or can be forced with \\  
  
\author{Deborah Christina Antunes}  
 \homepage{http://lattes.cnpq.br/1065956470701739}  
 \email{deborahantunes@gmail.com}  
 \altaffiliation{  
Curso de Psicologia, Universidade Federal do Cer\'a (UFC)  
}%Lines break automatically or can be forced with \\  
  
\author{Marilia Mello Pisani}  
 \homepage{http://lattes.cnpq.br/6738980149860322}  
 \email{marilia.m.pisani@gmail.com}  
 \altaffiliation{  
Centro de Ciências Naturais e Humanas, Universidade Federal do ABC (CCNH/UFABC)  
}%Lines break automatically or can be forced with \\  
  
  
%\author{Luciano da Fontoura Costa}  
% \homepage{http://cyvision.ifsc.usp.br/~luciano/}  
% \email{ldfcosta@gmail.com}  
% \altaffiliation[Also at ]{IFSC-USP}%Lines break automatically or can be forced with \\  
%  
%\author{Osvaldo N. Oliveira Jr.}  
% \homepage{www.polimeros.ifsc.usp.br/professors/professor.php?id=4}  
% \email{chu@ifsc.usp.br}  
% \altaffiliation[Also at ]{IFSC-USP}%Lines break automatically or can be forced with \\  
  
  
\date{\today}% It is always \today, today,  
 % but any date may be explicitly specified

\begin{abstract}

This article reports a minimal and general characterization of interaction networks evolution. Such a task involves a selection of aspects to investigate, which lead to: 1) activity distribution in time and among participants, 2) a sound and stable classification of vertex: peripheral, intermediary and hub sectors, 3) composition of basic measures into components with greater dispersion. While time patterns of activity are not obvious, participant activity follow concentrations expected by scale-free networks. Comparison with ideal Erd\"os-R\'enyi network with the same number of edges and vertexes revealed as a sound criterion for distinguishing sectors on the networks. Principal components in basic measures spaces revealed interesting and regular patterns of independence and dispersion. This includes a ranking of measures that most contribute to dispersion: 1) degree and strength measures, 2) symmetry related quantization, and 3) clusterization. Results suggested typologies for these networks and participants. Further work include considerations of text production, psychoanalysis inspired typologies, participatory democracy exploitation of observed properties, and better visualization support for network evolution.

\end{abstract}

\pacs{89.75.Fb,05.65.+b,89.65.-s}% PACS, the Physics and Astronomy

\keywords{complex networks, social network analysis, pattern recognition, statistics}%Use showkeys class option if keyword

\maketitle

\section{\label{sec:into}Introduction}

\subsection{Relatedwork}

Networks deriving from email lists have been analyzed in various pieces of work, focusing mainly on the evolution and topology of such networks (certo?). In a study of evolution of interaction networks ~\cite{barabasiEvo}, two distinct topologies were found to emerge, depending on the frequency of interactions, which exhibited a generalized power law or an exponential connectivity distribution~\cite{barabasiTopologicalEv} (certo?). In e-mail networks, free-scale properties were verified~\cite{bird} (certo?), and different linguistic traces were related to weak and strong ties~\cite{GMANE2} (o que quer dizer com weak and strong ties?). Such results are in accordance with phenomena observed in this work and linguistic characterization is being described in another document~\cite{rcText}. In networks of e-mail lists, unreciprocated edges often exceed 50\%, which matches empirical evidence reported in~\cite{newmanEvolving} (certo?). Although no correlation of topological characteristics and geographical position was found in ~\cite{barabasiGeo}, geographical incidences should be present in further refinement of the analysis (o que se quer dizer com geographical incidences should be present?).  
Então as redes oriundas de listas de emails em geral são scale-free? Mas não sempre?

The size of a community in an e-mail list is related to the appearance of strong hubs (o que é um strong hub?), according to the seminal Nature Letter by Palla, Bara{\'a}si and Vicsek~\cite{barabasiEvo}. .(Esta é uma conclusão importantíssima, mas não podemos baseá-la apenas no estudo de 4 redes, e só com a MET sendo diferente. Para chegar a uma conclusão robusta precisaremos analisar muito mais redes, ainda que elas não sejam dissecadas quanto às outras propriedades como as 4 listas no artigo).

Studies on network evolution often consider solely network growth, in which there is a monotonic increase in the number of events considered~\cite{barabasiEvo}. The evolution considered in this study is characterized by a constant number of messages (o que quer dizer?), which is also present in the literature, but was less explored to date.

Quais são os principais resultados da literatura sobre evolução de redes de e-mails?

Controllability of these networks is also an uncovered issue. These have unintuitive properties and might bring into forefront crucial differences between email interaction networks and interaction networks in Facebook or Twitter~\cite{barabasiControlCapacity,barabasiControlCentrality,barabasiControllability}.  
  
Gender related behavior has been notified (quer dizer notificado? Se for, não está correto) in mobile phone datasets~\cite{barabasiSex}, which can be further investigated to hold in email lists and in evolving terms, as a community oriented, non-private interactions are drawn from public emails groups with hundreds of participants.   
  
Considered years altogether, hundreds to thousands of participants post on a list, more rarely dozens or tenths of thousands (não entendi esta afirmação). The most active lists usually reaches a few thousands of participants. Authors have not checked each list (more than 20 thousand public email groups~\cite{GMANE}), and this might lead to a deeper insights in community-related network evolution.

\section{Description of the email lists analyzed}

Four email lists were selected and regarded as a representative set of medium to large email lists. They were obtained from the GMANE email archive~\cite{GMANE}, which consists of more than 20,000 email lists and more than 130,000,000 messages~\cite{GMANEwikipedia}. These lists cover a variety of topics, mostly technology-related. The archive can be described as a corpus with metadata of its messages, including send time, place, sender name, sender email address etc. GMANE usage in scientific research is reported in studies of isolated lists and of lexical innovations~\cite{GMANE2,bird}. The scripts for gathering and processing GMANE email messages are given in Appendix~\ref{scripts}.  
We selected the four lists below for their diversity, thus allowing us to probe general properties of email lists.

\begin{itemize}  
 \item Linux Audio Users list\footnote{gmane.linux.audio.users is list ID in GMANE archive.}. Dominated by participants with hybrid artistic and technological interests. Participants come from different countries, and English is the language used the most. Abbreviated as LAU from now on.  
 \item Linux Audio Developers list\footnote{gmane.linux.audio.devel is list ID in GMANE archive.}. Participants come from different countries, and English is the language used the most. More technical and less active version of LAU. Abbreviated LAD from now on.  
 \item Development list for the standard C++ library\footnote{gmane.comp.gcc.libstdc++.devel is list ID in GMANE archive.}. Dominated by specialized computer programmers, with participants coming from different countries, and English is the language used the most. Abbreviated as CPP from now on.  
 \item List of the MetaReciclagem project\footnote{gmane.politics.organizations.metareciclagem is list ID in GMANE archive.}. Dominated by Brazilian activists and digital culture interests. Participants are mostly Brazilians, and Portuguese is the most used language, although Spanish and English usage is also commonplace. Abbreviated MET from now on.  
\end{itemize}

The first 20,000 messages of each list were considered, with total timespan, authors, threads and missing messages being given in table~\ref{geralListas}.

\begin{table}

\centering

\begin{tabular}{|l|c|c|c|c|c|}\hline

list& $date\_1$ & $date\_{M}$ & $N$ & $\Gamma$ & $\overline{M}$ \\\hline

LAU & Jun/29/2003 & Jul/23/2005 & 1183 & 3373 & 5 \\

LAD & Jun/30/2003 & Oct/07/2009 & 1268 & 3113 & 4 \\

MET & Ago/01/2005 & Mar/07/2008 & 492 & 4607 & 23 \\

CPP & Mar/13/2002 & Aug/25/2009 & 1052 & 4506 & 7 \\ \hline

\end{tabular}

\caption{Columns $date\_1$ and $date\_M$ have dates of first and last messages from the 20,000 messages considered. $N$ is the number of participants (number of different email addresses). $\Gamma$ is the number of threads (count of messages without antecedent). $\overline{M}$ is messages missing in the 20,000 collection, $100\frac{23}{20000}=0.115$ percent in the worst case. MET notably has the fewer participants and the larger number of threads. This relation holds for the two pairs of lists considered: as the number of participants increase, the number of threads decreases.(verificar? Veja 2a. hipótese abaixo)}

\label{geralListas}

\end{table}

**Assuntos para discutir em seguida**

**How different are the natures of the email lists?**

First hypothesis – The more technical networks can be distinguished from the less technical ones. Show results, preferably with a 2D graph with clusters associated with the two types of network. If necessary, use other lists even if they are not studied in the present article. It is just to prove the point.

Second hypothesis – The lower the number of participants the higher the number of threads. Is this true? Is it not just a coincidence for the 4 lists analyzed (see Table 1). Is it possible that the higher number of threads is associated with more technical networks?

\subsection{Temporal characterization of network activity}

The incidence of messages at each second in a minute and in each minute in an hour is compatible with uniform distribution tests. This should be confirmed with an analytical study since messages were slightly more evenly distributed in all lists: for both seconds and minutes $\frac{max(incidence)}{min(incidence)} \in (1.26,1.275]$, while simulations reach these values, but are in average they are more discrepant $\xi=\frac{max(incidence')}{min(incidence')} \Rightarrow \mu\_\xi=1.2918 \text{ and } \sigma\_\xi=0.04619$.

Higher activity was observed between noon and 6 pm, followed by the time period between 6 pm and midnight. Therefore, participants work in the evening as well. Around 2/3 of the whole activity takes place from noon to midnight. See Table SI in the Supporting Information. Nevertheless, the activity peak occurs around midday, with a slight skew toward one hour before noon. Higher activity was observed during weekdays, as expected, especially for the more technical CPP and MET (see Table S2). No clear pattern is seen with regard to the weeks in a month, as indicated in Table S3. Activity is concentrated in Jun-Aug for MET and LAD, and from Dec-Mar for CPP, LAU and LAD (see Table S4). These observations fit academic calendars, vacations and end-of-year holidays.What is the reason for this behavior?

\subsection{**Networks deriving from the lists**}

Interaction networks were derived from the lists, which are directed and weighted, considered as more informative among all the possibilities (directed non-weighted (qual é mais usado? Non-weighted or unweighted?), and undirected weighted and undirected unweighted) ~\cite{bird,newmanCommunityDirected,newmanCommunity2013}. The networks were constructed as follows: a direct response from participant B to participant A forms an edge from A to B, as information went from A to B. The reasoning is: if B wrote a response to A, he read what A wrote and formulated a response, so B assimilated information from A, thus $A \rightarrow B$. Inverting edge direction yields the status network, as B read the message and considered what A wrote worth of responding, giving status to A, thus $B\rightarrow A$. This article uses the information network and depicted in Figure~\ref{formationNetwork}. Edges in both directions are allowed. Each time an interaction occurs, one unit is added to the edge weight. Self-loops were regarded as non-informative and discarded.

\begin{figure}[hb]  
 \centering  
 \includegraphics[width=0.5\textwidth]{figs/criaRede\_}  
 \caption{Formation of interaction network from email messages. Each vertex represents a participant. If participant B replies participant A, that is regarded as evidence that B received information from A. Multiple messages add ``weight'' to directed edge. Further details are in Section~\ref{intNet}.}  
 \label{formationNetwork}  
\end{figure}  
  
  
Previous messages on the thread create directed edges from their author to the observed message's author. Edges can be created from all antecedent messages on the message-response thread. Here, only immediate predecessors are linked to a new message's author, both for simplicity and for the valid objection that in adding two edges, $x\rightarrow y$ and $y\rightarrow z$, there is also a connection between $x\rightarrow z$. Potential interpretations for this weaker connection are usually common sense, such as: double length, half weight or with one more ``obstacle'' (não entendi a explicação?).

This suggests the adoption of other centrality measures that account for the connectivity with all nodes, such as betweenness centrality and accessibility~\cite{luMeasures,access}. Precisaremos acertar a colocação desta afirmação importante, pois está sendo mencionada uma medida, mas as medidas de redes ainda não foram introduzidas.

\subsection{Metrics for the network topology}

The topology of the networks was characterized with a small selection of the most standard measurements for each vertex:

\begin{itemize}  
 \item Degree $d\_i$: number edges linked to node $i$.  
 \item In-degree $d\_i^{in}$: number of edges ending at node $i$.  
 \item Out-degree $d\_i^{out}$: number of edges departing from node $i$.  
 \item Strength $s$: sum of weights of all edges linked to node $i$.  
 \item In-strength $s\_i^{in}$: sum of weights of all edges ending at node $i$.  
 \item Out-strength $s\_i^{out}$: sum of weights of all edges departing from node $i$.  
 \item Clustering coefficient $cc\_i$: fraction of pairs of neighbors of $i$ that are linked.

The standard clustering coefficient for undirected graphs was used.  
 \item Betweenness centrality $bt\_i$: fraction of geodesics that contain the node $i$. Betweenness centrality index considered directions and weight, as specified in~\cite{faster}.  
\end{itemize}  
  
In order to capture asymmetries in the activity of participants, the following metrics were introduced (see subsection~\ref{prevalence}):  
  
\begin{itemize}  
 \item asymmetry of note $i$: $asy\_i=\frac{d\_i^{in}-d\_i^{out}}{d\_i}$.  
 \item mean of asymmetry of edges: $\mu\_i^{asy}=\frac{\sum\_{j\in J\_i} e\_{ji}-e\_{ij}}{|J\_i|}$. Where $e\_{xy}$ is 1 if there is and edge from $x$ to $y$, $0$ otherwise. $|J\_i|$ is the number of neighbors of vertex $i$.  
 \item standard deviation of asymmetry of edges: $\sigma\_i^{asy}=\sqrt{\frac{\sum\_{j\in J\_i}[\mu\_{asy} -(e\_{ji}-e\_{ij}) ]^2 }{|J\_i|} }$  
 \item disequilibrium: $dis\_i=\frac{s\_i^{in}-s\_i^{out}}{s\_i}$.  
 \item mean of disequilibrium of edges: $\mu\_i^{dis}=\frac{\sum\_{j \in J\_i}\frac{w\_{ji}-w\_{ij}}{s\_i}}{|J\_i|}$, where $w\_{xy}$ is weight of edge $x\rightarrow y$ and zero if there is no such edge.  
 \item standard deviation of disequilibrium of edges: $\sigma\_i^{dis}=\sqrt{\frac{\sum\_{j\in J\_i}[\mu\_{dis}-\frac{(w\_{ji}-w\_{ij})}{s\_i}]^2}{|J\_i|}}$  
\end{itemize}

\subsection{Evolution of networks}

Evolution of the networks was observed within a fixed number of messages (window size: $ws$) that shifts in the message timeline. The window sizes used were $ws=$50, 100, 200, 400, 500, 800, 1000, 2000, 2500, 5000 and 10000. Within a same $ws$, the number of vertices (certo?) and edges vary in time, as do other network characteristics. Further work should deepen inspection of measure interdependence, this article holds to measures in Section~\ref{measures}. (não entendi o que quer dizer com “this article holds to measures in …”  
  
 The evolution of the networks was visualized with animations, image galleries and online gadgets ~\cite{animacoes,galGMANE,appGMANE}. Mapping of various topological measures to glyphs and layouts are being further explored as a parallel research. Furthermore, stable aspects of measures prominence along time are captured through mean and standard deviation (see Section~\ref{measures}). Constant sector sizes along time are observed in a timeline fashion in Appendix~\ref{figures}.

\subsection{Topology of the networks}

The networks obtained from the 4 lists exhibit free-scale and small world properties with regard to the degree of connectivity, as expected for a social network.This is shown in Figure ???.(Acho interessante mostrar que as redes são scale-free)

Using only the topological measurements mentioned above, we could not identify clusters using principal component analysis (PCA) projections, which would contain distinct types of participants in the email lists. Therefore, one is not able to separate hubs from less active participants based on the topological measurements. Nevertheless, because the activity distribution for the participants obeys a power law, one may define a clear criterion to identify hubs, intermediary and peripheral hubs. This is done by comparing the actual distribution with that predicted by the Ërdos-Rényi model, thus allowing us to define peripheral, intermediary and hub  
sectors~\cite{3setores}, as depicted in Figure~\ref{fig:setores}.  
The degree distribution $\widetilde{P}(k)$ of an ideal  
scale-free network $\mathcal{N}\_f$ with $N$ vertexes and $z$ edges, has less  
average degree vertices than the distribution $P(k)$ of an Erd\"os-R\'enyi  
random graph with the same number of vertices and edges:

\begin{equation}\label{criterio}  
 \widetilde{P}(k)<P(k) \Rightarrow \text{k is intermediary degree}  
\end{equation}  
  
If $\mathcal{N}\_f$ is directed and has self-loops, the probability  
for the presence of an unknown edge is $p=\frac{z}{N(N-1)}$, where $N(N-1)$ is the maximum number of edges for a network with $N$ vertices, directed edges and without selfloops.  
A vertex in the ideal Erd\"os-R\'enyi digraph with the same number of vertices and edges, and thus the same probability $p$ for the presence of an edge, will have degree $k$ with probability:  
  
\begin{equation}  
 P(k)=\binom{2(N-1)}{k}p^k(1-p)^{2(N-1)-k}  
\end{equation}  
  
The lower degree fat tail represents the border vertices, i.e. the peripheral sector. The higher degree fat tail is the hub sector. The arguments behind this classification are: 1) vertices so connected that they are virtually inexistent in networks connected at pure chance, especially without preferential attachment, 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. Degree values predicted as the most abundant if connections are created by pure chance, near the average, and less frequent in free-scale phenomena, are classified as intermediary vertices.

To ensure statistical validity, bins can be chosen to span the average of $\eta$ gaps between measure values (não entendi?). Thus, each bin, starting at degree $k\_i$, spans $\Delta\_i=\frac{\sum\_{i=1}^{\eta}(k\_{i+1}-k\_i)}{\eta}$ values, with borders $k\_i$ and $k\_i+\Delta\_{i}$. This changes equation~\ref{criterio} to:  
  
\begin{equation}\label{criterio2}  
 \sum\_{x=k\_i}^{k\_i+\Delta\_i} \widetilde{P}(x) < \sum\_{x=k\_i}^{k\_i+\Delta\_i} P(x) \Rightarrow \text{i is intermediary}  
\end{equation}  
  
If instead 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, where $\overline{w}$ is the average weight of an edge and $s\_i$ is the vertex strength. For in and out degrees and strengths, comparisons should be made with $\kappa\_i=2k\_i^{in}$, $\kappa\_i=2k\_i^{out}$, $\kappa\_i=2\frac{s\_i^{in}}{\overline{w}}$ and $\kappa\_i=2\frac{s\_i^{out}}{\overline{w}}$. Results of these segmentations are discussed in subsection~\ref{subsec:pih}.  
  
Since different metrics can be used in the segmentation to identify the three types of vertices, various criteria can be defined, e.g. with a very stringent criterion according to which a vertex will only be classified as hub if it is so for all the metrics. After a careful inspection of possible combinations, these were reduced to six:

\begin{itemize}  
\item Exclusivist criteria: vertices are only classified if the class is the same according to all metrics. In this case, the total number of vertices classified (usually) does not reach 100\%, which is indicated by a black line in Appendix~\ref{figures}.

\item Inclusivist criteria: a vertex has the class given by any of the metrics. Therefore, a vertex can belong to more than one class, and the percentages of members may add to more than 100%, which is indicated by a black line in Appendix~\ref{figures}.

\item Exclusivist cascade: hubs 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.  
 \item Inclusivist cascade: vertices are hubs if they are so classified 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 never classified as hub or intermediary with any of the metrics.

\item Exclusivist externals: 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 by any metric. The rest of the nodes are classified as intermediary.

\item Inclusivist externals: 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.  
\end{itemize}

These compound criteria, and their possible reduction, can be formalized in strict mathematical terms, but this was considered out of the scope of the present article. Important here is to notice that the compound criteria can be used to examine network sections in the case of a low number of messages, as will be shown later.

For the networks deriving from the email lists studied here, if degree distributions are used for defining the sectors, hubs can reach $10\%$ of all vertices. If instead 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, with $\overline{w}$ being the average weight of an edge and $s\_i$ the vertex strength. Then, for the networks here hubs account for approximately $5\%$ of all vertices, i.e. classification based on strength yields half the number of hubs as the plain degree. These distributions include in and out degrees and strengths, for which the comparison should use $\kappa\_i=2k\_i^{in}$, $\kappa\_i=2k\_i^{out}$, $\kappa\_i=2\frac{s\_i^{in}}{\overline{w}}$ and $\kappa\_i=2\frac{s\_i^{out}}{\overline{w}}$. There is a concentration of hub activity and of vertices with few connections, as indicated in Table~\ref{autores}.

The topology of the four networks investigated here was remarkably stable over time, in terms of the composition of hubs, intermediary and peripheral vertices, provided that a sample of at least 200 messages are taken as the window of analysis. This is shown in Figures ???and applies to all the 6 criteria of segmentation described above. Especially with 1000 or more messages, stable fractions of $\approx 5\%$ of hubs, $\approx [15-20]\%$ of intermediary and $\approx [75-80]\%$ peripheral vertices were obtained. Significantly, when the compound criterion ???was used, the stability increased and a window of only 200??? was already sufficient for the analysis. In fact, a reasonable window size for observation can be inferred by monitoring the giant component size and the degeneration of the hub, intermediary and peripheral sections. This degeneration is critical in the span of 50-100 messages. Upon using the compound criteria, such as exclusive cascade of Figure~\ref{fig:cpp250\_} for example, the network seems to hold basic structure even with as few as 20-50 messages. This indicates that concentration of activity and low-activity participants occurs even with very few messages.  
  
  
For the networks analysed, differences of using this smoothing process were not significant. There were between 20 and 1200 participants, so each participant were between $5\%$ and $0.08\%$ of all participants. The bottom line is: for network sizes considered, if connectivity intensity would most probably not exist in an Ed\"os Renyi network, than it is not an intermediary intensity. As peripheral vertex are abundant, this statistic discussion has no relevance. (não entendi?)  
  
With regard to the activity of the three types of participants, Table~\ref{autores} shows a concentration of hub activity and of nodes with few connections (peripheral?). Como interpretaressesresultados??.

\begin{table}

\caption{Distribution of activity among agents. The first column brings the percentage of messages sent by the most active participant. The column for the first quartile ($1Q$) exhibits minimum percentage of participants responsible for at least 25\% of total messages. Similarly, the column for the first three quartiles $1-3Q$ exhibits minimum percentage of participants responsible for 75\% of total messages. The last decile $10D$ column has maximum percentage of participants responsible for 10\% of activity (messages).}

\begin{center}

\begin{tabular}{ | l || c | c | c | c | }

\hline

list& hub & $ 1Q $ & $ 1-3Q $ & $10D$ \\ \hline

CPP & 14.41 & 0.19 (27.8\%) & 4.09 (75.13\%) & 83.65 (-10.04\%) \\

MET & 11.14 & 0.81 (30.61\%) & 8.33 (75,11\%) & 80.49 (-10.02\%) \\

LAU & 2.78 & 1.10 (25.16\%) & 13.02 (75,04\%) & 67.37 (-10.03\%) \\

LAD & 4.00 & 0.95 (25.50\%) & 11.83 (75,07\%) & 71.13 (-10.03\%) \\\hline

\end{tabular}

\end{center}

\label{autores}

\end{table}

\subsection{Relative importance of the topological metrics}

Acho que o conteúdo do parágrafo abaixo precisa ser precedido por uma explicação do que é Principal Component, e de que tipo de análise está sendo feito.

The principal component exhibit ponderation of centrality measures: degrees, strengths and betweenness centrality. Clustering coefficient is presented in almost perfect orthogonality. Dispersion is more prevalent in symmetry related measures than clustering coefficient. This holds for all network snapshots observed, even with as few messages as to degenerate structure. Symmetric and asymmetric edges have been reported as bounded to different roles played by participants and relations~\cite{newmanEvolving}. Principal components formation from original measures can be observed in Tables~\ref{compPCA0},~\ref{compPCA} and~\ref{compPCA2}. Individual vertexes relation with top two principal components can be seen in Figures~\ref{PCA} and~\ref{PCA2}. This peculiar first component that consists of the averaged sum of degree, strength and betweenness measures was verified to be incident in virtually all networks with 500 or more messages and most smaller networks (degeneration of basic structure is critical with $ws \approx 50-100$ messages). This composition of principal component suggests that all six degree and strength measures are equally important for system characterization, although it is known that they do not relate to the same participation characteristics.  
  
As expected, degree and strength are highly correlated, with Spearman correlation coefficient $\in [0.95,1]$ and Pearson coefficient $\in [0.85,1)$ for $ws>1000$. A high degree is associated with low clustering coefficient, as shown in figure~\ref{clust}.

\begin{figure}

\centering

\includegraphics[width=\columnwidth]{figs/ev0pr11CC}

\caption{Clustering coefficient versus node degree for the network obtained with the LAU list, $ws = 1000$ email messages. The general layout is in accordance with the literature (o queissoquerdizer? O queestá de acordo com a literatura?).With connected vertexes with low clusterization and high clusterization being gradually more incident as number of connections is lowered. Nãoentendi a sentença?.}

\label{clust}

\end{figure}

Figure~\ref{PCA} shows PCA plots obtained with the metrics degree, strength, betweenness and clustering for the list ???. The nodes classified into hubs, intermediary and peripheral nodes, according to the criterion ???, are distinguished by different colors. As already mentioned, it is not possible to establish clusters that would indicate a clear distinction among the different types of nodes. Nevertheless, one may note that hubs have variable values of PC1, including high values, whereas peripheral nodes have much lower PC1 values. This is understandable because, as indicated below, PC1 is mostly associated with connectivity (degree and/or strength). The intermediary nodes, as expected, fall in between in terms of their placements. The loads for building the PCA plots are given in Table~\ref{compPCA}, from which it is clear that PC1 is related to degree, strength and betweenness centrality. Therefore, the placement of the nodes in the PCA plot is primarily defined by centrality measurements (isso me pareceóbvioporque as categorias dos nósforamdefinidas com base emgrau, etc. Então, não sei se chega a ser um resultado, pois é consequência trivial da definição feita para as categorais).In fact, all the six degree and strength metrics are equally important for the placement.The second component PC2 is mostly associated with the clustering coefficient, from which one may conclude that hubs have low clustering (porquedesteresultado? Issogeralmenteacontece?), as observed in figure~\ref{clust}. The behavior illustrated in Figure~\ref{PCA} also applies to the other lists (?????) and for windows of analysis down to ??? messages.

\begin{figure}

\centering

\includegraphics[width=\columnwidth]{figs/ev0pr3PCA}

\caption{PCA of in and out degree and strength, betweenness centrality and clustering coefficient, as specified in subsection~\ref{measures}. On table~\ref{compPCA} is composition of principle components. Principle component is a pondered sum of degree and strength measures and betweenness centrality. Second component is mostly clustering coefficient. Similarity to plot in figure~\ref{clust} was verified to be regular, which suggests similar relation is held by degree and strength measures to clustering coefficient. Also, at least in this context, betweenness centrality is similar to a degree or strength measure.}

\label{PCA}

\end{figure}

When the asymmetry metrics are included, the PCA plot in figure~\ref{PCA2} shows a larger dispersion in the placement of nodes. Now PC2 is no longer related to the clustering coefficient, but rather to the asymmetry metrics. PC1 is still dominated by centrality metrics. These results are summarized in table~\ref{compPCA2}.

\begin{figure}

\centering

\includegraphics[width=\columnwidth]{figs/ev0pr1PCA}

\caption{Degree and strength, clustering coefficient, betweenness centrality and symmetry related measures are used for this scatter plot of principal components. Compositions of first three components are in table~\ref{compPCA2} and measure details in subsection~\ref{measures}. Most importantly, clustering coefficient is only relevant for third component, being second component representative of symmetry measurements of vertex interactions.}

\label{PCA2}

\end{figure}

\begin{table}

\centering

\begin{tabular}{|l|c|c| c|c| c|c|}\hline

& \multicolumn{2}{c|}{PC1} & \multicolumn{2}{c|}{PC2} & \multicolumn{2}{c|}{PC3} \\\hline

& $\mu$ & $\sigma$ & $\mu$ & $\sigma$ & $\mu$ & $\sigma$ \\\hline

$d$ & {\bf 48.02} & 1.39 & 2.82 & 1.74 &48.09 & 0.32 \\

$cc$ & 4.12 & 2.94 & {\bf 90.45} & 3.98 &3.98 & 0.77 \\

$bt$ & {\bf 47.87} & 1.55 & 6.74 & 4.08 & 47.93 & 0.46 \\ \hline

$\lambda$ & 64.67 & 0.52 & 33.26 & 0.23 &2.08 & 0.40 \\ \hline

\end{tabular}

\caption{Principal components composition in the simplest case: with degree, clustering coefficient and betweenness centrality. LAU list, $ws=1000$ messages in 20 disjoint positioning was used for statistics. The first component is associated with degree and betweenness centrality metrics. The second component is mostly clustering coefficient. First and second components sum more than 95\% of total dispersion.}

\label{compPCA0}

\end{table}

\begin{table}

\centering

\begin{tabular}{|l|c|c| c|c| c|c|}\hline

& \multicolumn{2}{c|}{PC1} & \multicolumn{2}{c|}{PC2} & \multicolumn{2}{c|}{PC3} \\\hline

& $\mu$ & $\sigma$ & $\mu$ & $\sigma$ & $\mu$ & $\sigma$ \\\hline

$d$ & {\bf 14.58} & 0.14 &0.43 & 0.35 & 1.51 & 1.08 \\

$d^{in}$ & {\bf 14.12} & 0.14 & 1.71 & 1.22 & 17.80 & 6.20 \\

$d^{out}$ & {\bf 13.95} & 0.12 &2.80 & 1.83 & 21.15 & 5.62 \\

$s$ & {\bf 14.48} & 0.13 &0.78 & 0.65 & 5.51 & 4.71 \\

$s^{in}$ & {\bf 14.10} & 0.14 & 2.17 & 1.28 & 17.32 & 6.11 \\

$s^{out}$ & {\bf 14.05} & 0.13 &2.08 & 1.14 & 19.31 & 4.86 \\ \hline

$cc$ & 0.99 & 0.70 & {\bf 83.38} & 4.83 &2.75 & 1.62 \\

$bt$ & {\bf 13.73} & 0.19 &6.65 & 1.31 & 14.66 & 10.14 \\ \hline

$\lambda$ & 81.80 & 0.83 & 12.53 &0.09 & 3.24 & 0.62 \\ \hline

\end{tabular}

\caption{Principal components' composition in percentages. LAU list, $ws=1000$ messages in 20 disjoint positioning was used for statistics. First component is a weighted sum of degree and strength metrics and betweenness centrality. The second component is mostly related to the clustering coefficient. The first and second components contribute with more than 90\% of dispersion.}

\label{compPCA}

\end{table}

\begin{table}

\centering

\begin{tabular}{|l|c|c| c|c| c|c|}\hline

& \multicolumn{2}{c|}{PC1} & \multicolumn{2}{c|}{PC2} & \multicolumn{2}{c|}{PC3} \\\hline

& $\mu$ & $\sigma$ & $\mu$ & $\sigma$ & $\mu$ & $\sigma$ \\\hline

$d$ & {\bf 11.51} & 0.42 &2.00 & 0.76 & 2.39 & 0.49 \\

$d^{in}$ & {\bf 11.45} & 0.34 &2.86 & 0.91 & 1.68 & 0.67 \\

$d^{out}$ & {\bf 10.68} & 0.60 & {\bf 7.43} & 1.00 & 3.00 & 1.02 \\

$s$ & {\bf 11.37} & 0.42 &1.75 & 0.71 & 4.31 & 0.63 \\

$s^{in}$ & {\bf 11.33} & 0.35 &2.39 & 1.10 & 3.69 & 0.86 \\

$s^{out}$ & {\bf 10.74} & 0.55 & {\bf 6.14} & 1.05 & 4.75 & 0.98 \\ \hline

$cc$ & 0.91 & 0.64 &2.68 & 1.67 & {\bf 22.27} & 6.43 \\

$bt$ & {\bf 10.87} & 0.38 &1.17 & 0.93 & 4.03 & 1.42 \\ \hline

$asy$ & 3.99 & 1.45 & {\bf 18.13} & 1.67 &2.55 & 1.77 \\

$\mu\_{asy}$ & 4.15 & 1.40 & {\bf 17.07} & 1.78 &2.49 & 1.67 \\

$\sigma\_{asy}$ & 1.21 & 0.67 & {\bf 17.49} & 0.79 &3.29 & 2.33 \\

$dis$ & 5.78 & 0.51 &1.94 & 1.28 & {\bf 24.75} & 3.73 \\

$\mu\_{dis}$ & 0.79 & 0.49 & {\bf 14.00} & 1.14 &3.73 & 3.13 \\

$\sigma\_{dis}$ & 5.18 & 0.72 &4.93 & 2.48 & {\bf 17.04} & 4.78 \\ \hline

$\lambda$ & 51.09 & 1.07 & 20.04 & 1.31 &9.23 & 6.63 \\ \hline

\end{tabular}

\caption{Distribution of components, added measures of symmetry described in subsection~\ref{measures}. LAU list, $ws=1000$ messages in 20 disjoint positioning was used for statistics. In this case, clusterization is pushed to third component, with disequilibrium measures. Second component is primarily symmetry measures, but also some out degree and strength contribution. Betweenness again has a role similar to degree, but weaker. Clusterization component combines with disequilibrium, while asymmetry is related to out degree and strength. Three components have in average 80.36\% of dispersion.}

\label{compPCA2}

\end{table}

\subsection{Temporal evolution of activity}\label{evolNet}

The results shown in (figures, tables??) Appendix~\ref{figures} allow us to infer the following conclusions related to the temporal evolution of activity:

\begin{itemize}  
 \item Core hubs usually have intermittent activity. Very stable activity was found on MET hubs, which motivated its integration to this work. Literature reports greater stability of participation in smaller communities~\cite{barabasiEvo}, which is the reason why smaller numbers of participants in MET was considered a direct cause of stable hubs activity.  
 \item Typically, hub activity is trivial: they interact as much as possible, in every occasion with everyone. Peripheral vertex activity also follows a simple pattern: they will interact very rarely, in very few occasions. Intermediary vertices seem responsible for the network structure.  
 \item The network operation is dictated by the behavior of intermediary vertices behavior. These can exhibit preferential communication to peripheral or hub vertices.  
 \item Some of the most active participants receive many responses with relative few messages sent, and are never top hubs. These seem as authorities and contrast with participants that respond much more than receive responses.  
 \item The most obvious community structure, as observed by hung clustering coefficient (não entendi?), is found only in peripheral and intermediary sectors.  
\end{itemize}  
  
  
This ``primitive typology'', characterized by peripheral, intermediary and hub types, can be further scrutinized using concepts involved in other typologies, such Meyer-Briggs, Pavlov or F-Scale. This has no intention of being a direct result from numeric analysis, it is a refinement the description of found structure and classes considered. (não entendi o comentário)

%\begin{table}[hi]

% \centering

% \begin{tabular}{|l |c|c|c|c|}\hline

% & $\mu\_{exc}$ & $\mu\_{inc}$ & $\mu\_{N}$ & $\mu\_z$ \\ \hline

% CPP

% \end{tabular}

% \caption{Fraction of vertex in each section with respect to different window sizes.}

% \label{tab:structEv}

%\end{table}

%

\section{Concluding remarks and future work}\label{concluding}

Further work should observe textual production of network sectors. Resulting knowledge purposes to network and participants tipologization, and both topological and textual analysis should foster characterization of interaction networks and participation incidences.

Regarding topological aspects,further work should inspect other measures incident in each sector, along with common characteristics. Observance of attributes with greater contribution to principal components of LDA should reveal best chances to present these three sections as clusters. Another possibility, specially for a brute-force characterization of such sections, is to remove vertexes with degree close to $k\_1$ or $k\_2$ depicted in figure~\ref{fig:setores}.

Observed networks was coherent with literature in different aspects, such as concentration of activity, and clusterization versus connectivity patterns. Even so, verification of results in other virtual environment, such as Facebook and LinkedIn, might help understanding how general are this structures and what are convenient uses.

\begin{acknowledgments}

Renato Fabbri is grateful to CNPq (project 870336/1997-5) and

the Postgraduate Committee of the IFSC/USP. This author is also grateful for

the American Jewish Committee for maintaining an online copy of the Adorno book

used on the epigraph~\cite{adorno}. Authors thanks Gmane creators and keepers, specifically: Gmane is run by Lars MagneIngebrigtsen, and the administrators are Tom Koelman, Jason R. Mastaler, Steinar Bang, Jon Ericson, Wolfgang Schnerring, Sebastian D.B. Krause, Nicolas Bareil, Raymond Scholz, and Adam Sjøgren. Authors thank refered email lists communities and welcome feedback as core contribution to this, and similar, research.

\end{acknowledgments}

\appendix

\section{Data description}

Messages are downloaded from GMANE database by RSS in the mbox email text format.

They are requested one by one to avoid reaching maximum size of the requests accepted by

GMANE API.

Every message has about 30 fields, from which the following are crucial

for the present work:

\begin{itemize}

\item ``From'' field, as it specifies the sender of the message, in the usual format of ``First\\_name Last\\_Name $<email>$''.

\item ``Date'' field, which is given with the resolution of a second.

\item ``Message-ID'', important to state antecedent/consequent relation between messages and therefore from an author to a replier.

\item ``References'', has the ID of the message it is an answer for, if any, and earlier messages in the thread.

\end{itemize}

Field ``In-Reply-To'' has only the ID of the message it replies and can be sometimes

a shortcut or an alternative to ``References''. Also, the textual content of the messages,

accessed through ``payload'' method of the mbox message object, is of central interest and

the authors dedicated an article to include the textual content of the messages to the analysis.

Basic constructs for obtaining all results in this article are described in~\ref{ap:os}. Scripts, written in Python programming language, are publicly available at~\cite{scriptsFim} and very briefly specified below.

\subsection{Third party libraries and software}

Programming resources used were mainly Python and part of the

common scientific bundle for the language. More specifically,

scripts where written for 2.7.3 version of Python,

with the following third party libraries: Numpy, Pylab/Matplotlib, NetworkX, IGraph.

Behind the scenes, Graphviz is accessed via PyGraphviz to make network drawings.

\subsection{Own scripts}\label{ap:os}

All results were obtained with scripts writen in the Python programming language. These are kept in a public repository for backup and sharing with research community~\cite{scriptsFim}. Core scripts, for deriving structures and results exhibited in this article, are in the LEIAME file.

\section{Figures of vertex classification fractions as the network evolves}\label{figures}

Two lists are exhibited in this section, CPP and LAD. These structures are very similar in all

four lists and laying extensively all figures is redundant. Window sizes of $ws =$ 10000, 5000,

1000, 500, 250, 100 and 50 messages were used.

Texto para o Apêndice que traz as figuras de dependência temporal

Evolution of network is observed within a fixed number of messages (or window size $ws$) that shifts in the message timeline.

All 50, 100, 200, 400, 500, 800, 1000, 2000, 2500, 5000, 10000 number of messages were used as $ws$. Within a same $ws$, the number vertex and edges vary in time, as do other network characteristics. Further work should deepen inspection of measure interdependence, this article holds to measures exposed in subsection~\ref{measures}.

\begin{figure\*}[hb]

\centering

\includegraphics[width=\textwidth]{figs/CPP/10000}

\caption{Distribution of vertices with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.}

\label{fig:cpp10000}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/CPP/10000\_2}

\caption{Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection~\ref{benchmarks}.}

\label{fig:cpp10000\_}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/CPP/5000}

\caption{Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.}

\label{fig:cpp5000}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/CPP/5000\_2}

\caption{Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection~\ref{benchmarks}.}

\label{fig:cpp5000\_}

\end{figure\*}

%%%%

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/CPP/1000}

\caption{Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.}

\label{fig:cpp1000}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/CPP/1000\_2}

\caption{Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection~\ref{benchmarks}.}

\label{fig:cpp1000\_}

\end{figure\*}

%%%

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/CPP/500}

\caption{Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.}

\label{fig:cpp500}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/CPP/500\_2}

\caption{Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection~\ref{benchmarks}.}

\label{fig:cpp500\_}

\end{figure\*}

%%%%%%%%%

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/CPP/250}

\caption{Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.}

\label{fig:cpp250}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/CPP/250\_2}

\caption{Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection~\ref{benchmarks}.}

\label{fig:cpp250\_}

\end{figure\*}

%%%%%%%

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/CPP/100}

\caption{Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.}

\label{fig:cpp100}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/CPP/100\_2}

\caption{Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection~\ref{benchmarks}.}

\label{fig:cpp100\_}

\end{figure\*}

%%%%%%%%%%

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/CPP/50}

\caption{Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.}

\label{fig:cpp50}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/CPP/50\_2}

\caption{Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection~\ref{benchmarks}.}

\label{fig:cpp50\_}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/LAD/10000}

\caption{Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.}

\label{fig:lad10000}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/LAD/10000\_2}

\caption{Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection~\ref{benchmarks}.}

\label{fig:lad10000\_}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/LAD/5000}

\caption{Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.}

\label{fig:lad5000}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/LAD/5000\_2}

\caption{Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection~\ref{benchmarks}.}

\label{fig:lad5000\_}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/LAD/1000}

\caption{Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.}

\label{fig:lad1000}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/LAD/1000\_2}

\caption{Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection~\ref{benchmarks}.}

\label{fig:lad1000\_}

\end{figure\*}

%%%

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/LAD/500}

\caption{Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.}

\label{fig:lad500}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/LAD/500\_2}

\caption{Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection~\ref{benchmarks}.}

\label{fig:lad500\_}

\end{figure\*}

%%%%%%%%%

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/LAD/250}

\caption{Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.}

\label{fig:lad250}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/LAD/250\_2}

\caption{Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection~\ref{benchmarks}.}

\label{fig:lad250\_}

\end{figure\*}

%%%%%%%

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/LAD/100}

\caption{Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.}

\label{fig:lad100}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/LAD/100\_2}

\caption{Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection~\ref{benchmarks}.}

\label{fig:lad100\_}

\end{figure\*}

%%%%%%%%%%

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/LAD/50}

\caption{Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.}

\label{fig:lad50}

\end{figure\*}

\begin{figure\*}[hbtp]

\centering

\includegraphics[width=\textwidth]{figs/LAD/50\_2}

\caption{Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection~\ref{benchmarks}.}

\label{fig:lad50\_}

\end{figure\*}

%

%\begin{figure\*}

% \centering

% \includegraphics[width=\textwidth]{pcm}

% \caption{Pulse Code Modulation (PCM) audio: an analogical signal is represented by 25 samples with 4 bits each.}

% \label{fig:PCM}

%\end{figure\*}

%

\nocite{\*}

\bibliography{evsn}% Produces the bibliography via BibTeX.

\end{document}

%

% \*\*\*\*\*\* End of file aipsamp.tex \*\*\*\*\*\*

\subsection{Visualization of network evolution}

In refining hypotheses, visualization of the network was crucial. Animations, image galleries and an online gadgets were made~\cite{animacoes,galGMANE,appGMANE}. Mapping of various topological measures to glyphs and layouts is being explored as a parallel research. This article is dedicated to preliminary data analysis, observation of primary messages and the classification of network sectors in peripheral, intermediary and hubs, as a primitive typology, described in next subsection.

Supporting Information

Table S1 abaixo

\begin{table\*}

%\tiny

\caption{Hours of the day and percentage of activity ($\frac{\text{counted messages}}{\text{total messages}}$) in each hour, 6 hours and 12 hours. Maximum activity rates are in bold. In hour columns, minimum activity is also bold. The less active period of the day is around 4-6h. Maximum activity is between 10-13h. Afternoon is most active in 6h division of the day. The noon has $\approx \frac{2}{3}$ of 24h activity. }

\begin{center}

\begin{tabular}{ |l|| c|c|c| c|c|c| c|c|c| c|c|c|}

\hline

& \multicolumn{3}{c|}{CPP} & \multicolumn{3}{c|}{MET} & \multicolumn{3}{c|}{LAU} & \multicolumn{3}{c|}{LAD} \\ \hline

& 1h & 6h & 12h & 1h & 6h & 12h & 1h & 6h & 12h & 1h & 6h & 12h \\ \hline\hline

0h & 3.66 & \multirow{6}{\*}{10.67} & \multirow{12}{\*}{33.76} & 2.87 & \multirow{6}{\*}{7.15} & \multirow{12}{\*}{29.33} & 3.58 & \multirow{6}{\*}{10.14} & \multirow{12}{\*}{36.88} & 4.00 & \multirow{6}{\*}{10.77} & \multirow{12}{\*}{33.13} \\

1h & 2.76 &&& 1.77 &&& 2.22 &&& 2.52 && \\

2h & 1.79 &&& 1.04 &&& 1.63 &&& 1.79 && \\

3h & 1.10 &&& 0.64 &&& 1.06 &&& 1.06 && \\

4h & {\bf 0.68} &&& 0.47 &&& 0.84 &&& 0.75 && \\

5h & 0.69 &&& {\bf 0.38} &&& {\bf 0.82} &&& {\bf 0.66} && \\\cline{3-3}\cline{6-6}\cline{9-9}\cline{12-12}

6h & 0.83 & \multirow{6}{\*}{23.09} && 0.72 & \multirow{6}{\*}{22.18} && 1.17 & \multirow{6}{\*}{26.74} && 0.85 & \multirow{6}{\*}{22.36} & \\

7h & 1.24 &&& 1.33 &&& 2.37 &&& 1.56 && \\

8h & 2.28 &&& 2.67 &&& 3.54 &&& 2.96 && \\

9h & 4.52 &&& 4.40 &&& 6.04 &&& 4.68 && \\

10h & 6.62 &&&6.29 &&& {\bf 6.83} &&& 5.93 && \\

11h & {\bf 7.61} &&&6.78 &&& 6.79 &&& 6.40 && \\\hline

12h & 6.44 & \multirow{6}{\*}{\bf 37.63} & \multirow{12}{\*}{\bf 66.24} & {\bf 7.33} & \multirow{6}{\*}{\bf 42.22} & \multirow{12}{\*}{ \bf 70.66} & 6.11 & \multirow{6}{\*}{\bf 35.65} & \multirow{12}{\*}{ \bf 63.12} & {\bf 6.41} & \multirow{6}{\*}{\bf 37.25} & \multirow{12}{\*}{\bf 66.87} \\

13h & 6.04 &&&7.08 &&& 6.26 &&& 6.12 && \\

14h & 6.47 &&&7.09 &&& 6.38 &&& 6.33 && \\

15h & 6.10 &&&7.14 &&& 5.93 &&& 5.98 && \\

16h & 6.22 &&&6.68 &&& 5.52 &&& 6.40 && \\

17h & 6.36 &&&6.89 &&& 5.46 &&& 6.02 && \\\cline{3-3}\cline{6-6}\cline{9-9}\cline{12-12}

18h & 6.01 & \multirow{6}{\*}{28.61} && 5.99 & \multirow{6}{\*}{28.44} && 5.24 & \multirow{6}{\*}{27.46} && 5.99 & \multirow{6}{\*}{29.63} & \\

19h & 5.02 &&&5.23 &&& 4.52 &&& 5.03 && \\

20h & 4.85 &&&4.98 &&& 4.55 &&& 4.63 && \\

21h & 4.38 &&&4.37 &&& 4.42 &&& 4.59 && \\

22h & 4.06 &&&4.24 &&& 4.51 &&& 4.88 && \\

23h & 4.30 &&&3.64 &&& 4.23 &&& 4.53 && \\\hline

\end{tabular}

\end{center}

\label{dia}

\end{table\*}

Table S2 abaixo

\begin{table}[h]

\caption{Concentration of activity on days along the week. Weekend days are at least $\frac{1}{3}$ less active and can reach $\frac{1}{3}$ of activity. MET concentrates activity in weekdays the most, leaving only 13.98\% of total activity to Saturday and Sunday. LAU is the one that less concentrates activity in weekdays, reaching 20.94\% of total activity in weekends. These might suggest professional relation of CPP and MET participants to the topics of interest, or a hobby relation of LAU and LAD participants.Esta é uma informação importante que deve ser conectada com a possibilidade de distinguir diferentes tipos de listas}

\begin{center}

\begin{tabular}{ | l | c | c | c | c | c | c | c |}

\hline

& Mon & Tue & Wed & Thu & Fri & Sat &Sun \\ \hline

CPP & 17.06 & 17.43 & 17.61 & 17.13 & 16.30 & 6.81 & 7.67 \\ \hline

MET & 17.53 & 17.54 & 16.43 & 17.06 & 17.46 & 7.92 & 6.06 \\ \hline

LAU & 15.71 & 15.80 & 15.88 & 16.43 & 15.13 & 10.13 & 10.91 \\ \hline

LAD & 14.91 & 17.73 & 17.01 & 15.40 & 14.25 & 10.39 & 10.30 \\\hline

\end{tabular}

\end{center}

\label{semana}

\end{table}

Table S3 abaixo

\begin{table\*}

%\tiny

\caption{Activity along the days of the month. As can be noted, there is no clear pattern. One might point a slight tendency for the first two weeks to be more active, although this table does not present statistical significance for such an assumption. For the scope of this study, differences of activity along the month are assumed to be non existent.}

\begin{center}

\begin{tabular}{ |l|| c|c|c| c|c|c| c|c|c| c|c|c|}

\hline

& \multicolumn{3}{c|}{CPP} & \multicolumn{3}{c|}{MET} & \multicolumn{3}{c|}{LAU} & \multicolumn{3}{c|}{LAD} \\ \hline

day & 1 day & 7 days & 14 days & 1 day & 7 days & 14 days & 1 day & 7 days & 14 days & 1 day & 7 days & 14 days \\ \hline\hline

1 & 3.19 & \multirow{7}{\*}{23.05} & \multirow{14}{\*}{45.63} & 3.01 & \multirow{7}{\*}{25.16} & \multirow{14}{\*}{48.08} & 3.34 & \multirow{7}{\*}{23.06} & \multirow{14}{\*}{47.31} & 3.22 & \multirow{7}{\*}{21.96} & \multirow{14}{\*}{46.70} \\

2 & 3.07 &&& 3.38 &&& 3.38 &&& 3.42 && \\

3 & 3.20 &&& 3.55 &&& 3.20 &&& 2.87 && \\

4 & 3.63 &&& 4.34 &&& 3.52 &&& 2.91 && \\

5 & 2.85 &&& 3.93 &&& 2.68 &&& 3.30 && \\

6 & 3.67 &&& 3.76 &&& 3.18 &&& 3.52 && \\

7 & 3.45 &&& 3.18 &&& 3.77 &&& 2.27 && \\\cline{3-3}\cline{6-6}\cline{9-9}\cline{12-12}

8 & 3.12 & \multirow{7}{\*}{22.57} && 3.36 & \multirow{7}{\*}{22.92} && 3.62 & \multirow{7}{\*}{24.25} && 3.72 & \multirow{7}{\*}{24.73} & \\

9 & 2.57 &&& 3.44 &&& 3.82 &&& 3.97 && \\

10 & 2.92 &&& 3.17 &&& 3.06 &&& 3.77 && \\

11 & 3.54 &&& 3.88 &&& 3.11 &&& 3.27 && \\

12 & 3.23 &&&2.94 &&& 3.40 &&& 2.75 && \\

13 & 3.39 &&&3.29 &&& 3.55 &&& 3.34 && \\

14 & 3.81 &&&2.83 &&& 3.69 &&& 3.93 && \\\hline

15 & 3.35 & \multirow{7}{\*}{23.02} & \multirow{14}{\*}{46.31} &2.72 & \multirow{7}{\*}{21.87} & \multirow{14}{\*}{ 43.56} & 3.23 & \multirow{7}{\*}{22.84} & \multirow{14}{\*}{ 44.01 } & 3.37 & \multirow{7}{\*}{22.82} & \multirow{14}{\*}{46.00} \\

16 & 3.77 &&&2.96 &&& 2.94 &&& 3.37 && \\

17 & 3.45 &&&3.01 &&& 3.02 &&& 2.95 && \\

18 & 3.47 &&&3.39 &&& 3.63 &&& 3.22 && \\

19 & 2.90 &&&3.42 &&& 3.16 &&& 3.59 && \\

20 & 2.80 &&&3.09 &&& 3.25 &&& 3.21 && \\

21 & 3.29 &&&3.27 &&& 3.61 &&& 3.13 && \\\cline{3-3}\cline{6-6}\cline{9-9}\cline{12-12}

22 & 2.88 & \multirow{7}{\*}{23.29} &&2.92 & \multirow{7}{\*}{21.69} && 3.80 & \multirow{7}{\*}{21.17} && 3.07 & \multirow{7}{\*}{23.18} & \\

23 & 4.01 &&&3.27 &&& 3.03 &&& 3.06 && \\

24 & 3.13 &&&2.92 &&& 2.31 &&& 2.72 && \\

25 & 3.57 &&&2.83 &&& 2.38 &&& 3.16 && \\

26 & 3.27 &&&2.97 &&& 3.49 &&& 3.57 && \\

27 & 3.27 &&&3.41 &&& 2.92 &&& 3.92 && \\

28 & 3.17 &&&3.36 &&& 3.26 &&& 3.69 && \\\hline

29 & 3.68 & \multirow{3}{\*}{8.06} & \multirow{3}{\*}{8.06} & 2.93 & \multirow{3}{\*}{8.36} & \multirow{3}{\*}{8.36} & 3.34 & \multirow{3}{\*}{8.68} & \multirow{3}{\*}{8.68} & 3.15 & \multirow{3}{\*}{7.30} & \multirow{3}{\*}{7.30} \\

30 & 2.76 &&&3.14 &&& 3.75 &&& 2.71 && \\

31 & 1.63 &&&2.29 &&& 1.60 &&& 1.45 && \\\hline

\end{tabular}

\end{center}

\label{mes}

\end{table\*}

Table S4 abaixo

\begin{table\*}[t]

\scriptsize

\caption{Activity along the year, in months, trimesters, quadrimesters and semesters. Engagement in list participation seem to concentrate in two periods: middle of the year (Jun-Aug, lists MET and LAD), and transition from years (Dec-Mar, lists CPP, LAU and LAD). Messages were considered as to complete 12-month slots, so every month has the same time of occurrences.}

\begin{center}

\begin{tabular}{ |l|| c|c|c|c|c| c|c|c|c|c| c|c|c|c|c| c|c|c|c|c|}

\hline

& \multicolumn{5}{c|}{CPP} & \multicolumn{5}{c|}{MET} & \multicolumn{5}{c|}{LAU} & \multicolumn{5}{c|}{LAD} \\ \hline

& m. & b. & t. & q. & s. & m. & b. & t. & q. & s. & m. & b. & t. & q. & s. & m. & b. & t. & q. & s. \\ \hline\hline

Jan & 8.70 & \multirow{2}{\*}{17.00} & \multirow{3}{\*}{\bf 27.23} & \multirow{4}{\*}{\bf 36.48} & \multirow{6}{\*}{\bf 54.26} & 4.88 & \multirow{2}{\*}{11.01} & \multirow{3}{\*}{16.90} & \multirow{4}{\*}{23.32} & \multirow{6}{\*}{47.74} & 10.22 & \multirow{2}{\*}{\bf 19.56} & \multirow{3}{\*}{\bf 28.23} & \multirow{4}{\*}{\bf 35.09} & \multirow{6}{\*}{49.17} & 11.23 & \multirow{2}{\*}{18.49} & \multirow{3}{\*}{26.43} & \multirow{4}{\*}{36.04} & \multirow{6}{\*}{\bf 57.95} \\

Fev& 8.29 &&&&& 6.13 &&&&& 9.34 &&&&& 7.26 &&&& \\\cline{3-3}\cline{8-8}\cline{13-13}\cline{18-18}

Mar & {\bf 10.23} & \multirow{2}{\*}{\bf 19.49} &&&& 5.89 & \multirow{2}{\*}{12.31} &&&& 8.67 & \multirow{2}{\*}{15.52} &&&& 7.94 & \multirow{2}{\*}{17.55} &&& \\\cline{4-4}\cline{9-9}\cline{14-14}\cline{19-19}

Apr & 9.26 && \multirow{3}{\*}{27.03} &&& 6.42 && \multirow{3}{\*}{30.84} &&& 6.85 && \multirow{3}{\*}{20.94} &&& 9.61 && \multirow{3}{\*}{\bf 31.51} && \\\cline{3-3}\cline{5-5}\cline{8-8}\cline{10-10}\cline{13-13}\cline{15-15}\cline{18-18}\cline{20-20}

Mai & 9.41 & \multirow{2}{\*}{17.78} && \multirow{4}{\*}{33.46} && 10.46 & \multirow{2}{\*}{\bf 24.42} && \multirow{4}{\*}{\bf 47.83} && 7.27 & \multirow{2}{\*}{14.09} && \multirow{4}{\*}{30.37} && 8.94 & \multirow{2}{\*}{\bf 21.91} && \multirow{4}{\*}{\bf 37.56} & \\

Jun & 8.37 &&&&& {\bf 13.96} &&&&& 6.81 &&&&& {\bf 12.97} &&&& \\\cline{3-3}\cline{4-4}\cline{6-6}\cline{8-9}\cline{11-11}\cline{13-14}\cline{16-16}\cline{18-19}\cline{21-21}

Jul & 8.70 & \multirow{2}{\*}{15.68} & \multirow{3}{\*}{22.94} && \multirow{6}{\*}{45.73} & 13.23 & \multirow{2}{\*}{23.41} & \multirow{3}{\*}{\bf 31.16} && \multirow{6}{\*}{\bf 52.26} & 8.96 & \multirow{2}{\*}{16.28} & \multirow{3}{\*}{24.47} && \multirow{6}{\*}{\bf 50.82} & 9.02 & \multirow{2}{\*}{15.65} & \multirow{3}{\*}{22.29} && \multirow{6}{\*}{42.05} \\

Ago & 6.98 &&&&& 10.28 &&&&& 7.31 &&&&& 6.63 &&&& \\\cline{3-3}\cline{5-5}\cline{8-8}\cline{10-10}\cline{13-13}\cline{15-15}\cline{18-18}\cline{20-20}

Set & 7.26 & \multirow{2}{\*}{15.36} && \multirow{4}{\*}{30.06} && 7.75 & \multirow{2}{\*}{16.80} && \multirow{4}{\*}{28.86} && 8.18 & \multirow{2}{\*}{16.24} && \multirow{4}{\*}{34.54} && 6.63 & \multirow{2}{\*}{12.38} && \multirow{4}{\*}{26.40} & \\\cline{4-4}\cline{9-9}\cline{14-14}\cline{19-19}

Oct & 8.10 && \multirow{3}{\*}{22.80} &&& 9.05 && \multirow{3}{\*}{21.10} &&& 8.06 && \multirow{3}{\*}{26.36} &&& 5.74 && \multirow{3}{\*}{19.77} && \\\cline{3-3}\cline{8-8}\cline{13-13}\cline{18-18}

Nov & 7.86 & \multirow{2}{\*}{14.69} &&&& 7.46 & \multirow{2}{\*}{12.06} &&&& 7.63 & \multirow{2}{\*}{18.30} &&&& 7.63 & \multirow{2}{\*}{14.02} &&& \\

Dec & 6.81 &&&&& 4.59 &&&&& {\bf 10.66} &&&&& 6.39 &&&& \\\hline

\end{tabular}

\end{center}

\label{ano}

\end{table\*}

A seção discutindo tipologias talvez tenha que vir após os resultados da topologia e evolução das redes. Ainda não encontrei o melhor lugar para ela, e por isso estou deixando aqui no fim.

\subsection{Typological deepening}

There are other ways to split a network. To point a common example, the center of the network is defined as all the nodes whose maximum distance to any other node is the radius\footnote{Radius is the minimum maximum distance to all nodes. Equivalently, the radius is the minimum eccentricity.}.   
In the same framework, the periphery (as opposed to the center) consists of the nodes whose maximum distance to any node is the diameter\footnote{Diameter is the maximum geodesic on the network.}. Accordingly, the intermediary sector can be defined as the nodes that are not in the center or in the periphery. Interestingly, in the email networks analyzed, with these criteria, the center can often be a factor of 4 times larger than the periphery and the intermediary group often exceed 93\% of the nodes~\cite{networkx}.  
  
Models of human dynamics can be used to predict and classify activity. In this case, agent activity is commonly considered a Poisson process, as a consequence of the randomly distributed events in time. Even so, evidence-based models suggests that human activity patterns follow non-Poisson statistics, characterized by a long tail of inactivity with of bursts of rapidly occurring events~\cite{barabasiHumanDyn,barabasiPhone}. Emails are reported as having a heavy tailed distribution with $\alpha=1$, together with web browsing and library loans~\cite{barabasiHumanDyn}.  
  
Typologies can also be conveniently adapted from psychiatric, psychological and psychoanalytic theories.   
Concerning empirical research,  
Theodor Adorno was a core conceiver of an one-of-a-kind typology that resulted from observing authoritarian  
personality traces\footnote{Some of them related to Nazism adoption, antisemitism and potential fascists.}, sometimes depicted as an authoritarian syndrome.  
%Influenced by Social and Psychoanalytic Theories, Adorno et all applied a questionnaire to individuals, from which they reached a position in the the ``F Scale'', to verify etnocentric, conservatory and antidemocratic trends~\cite{adorno}. From psychoanalitic interviews and the F Scale, they derived a typology, which gathers prejudice-inclined traces in personality. Both, low and high scores are considered with prejudicial traces. This typology has nine authoritarian types, the six types with high score in the F Scale: surface resentment, conventional, authoritarian, rebel and psychopath, crank, manipulative; and three of the five types with low score in the F Scale: rigid, protesting, impulsive, easygoing, genuine liberal. Each side of the dipole has a rank of intensity that increases as the order written above.  
 Other typologies include Jung's extroversion-introversion trait with four modes of orientation. This four modes are divided in two perceiving functions (sensation and intuition) and two judging functions (thinking and feeling)~\cite{jung}. Myers-Briggs Type Indicator extrapolated Jungian theories into a questionnaire and added perceiving and judging as a fourth dipole~\cite{myers}. Even plain Freudian criteria, such as neurosis, psychosis, perversity and denegation, can be used directly for such categorization, as they have verbal and behavioral typical traces~\cite{freud,freud2}.  
  
It was considered central to benefit from key human typologies, both by adding descriptions to a type and by further characterizing classes in the terms encountered.