Stability in human interaction networks: primitive typology of vertex, prominence of measures and time activity statistics

Renato Fabbri,[1, a)] Vilson V. da Silva Jr.,[b)] Ricardo Fabbri,[c)] Deborah C. Antunes,[d)] and Marilia M. Pisani[e)]
São Carlos Institute of Physics, University of São Paulo (IFSC/USP)

(Dated: 15 April 2015)

This article reports interaction networks stability by means of three quantitative criteria: activity distribution in time and among participants; a sound classification of vertices in peripheral, intermediary and hub sectors; the combination of basic measures into principal components with greater variance. We analyzed the temporal activity and topology evolution of networks in four email lists by considering window sizes from 50 to 10,000 messages, which were made to slide to generate snapshots of the network along a timeline. Activity in terms of seconds and minutes exhibit an uniform pattern, while hours, days and months exhibit stable concentrations. Participant activity follows the expected distribution of scale-free networks. We compare these networks to Erdös-Rényi networks in order to assign members to three distinct sectors, namely hubs, intermediary and periphery. Approximately 5% of the vertices are hubs, 15-35% are intermediary and the remainder belongs periphery. The metrics that most contribute to data dispersion were found to be centrality-related (degree, strength and betweenness), followed by symmetry-related, and then clustering coefficient. The results also embrace a sketch of a physics-based typology of agents, with proper social and psychological speculations. Because the network properties reported did not depend on the email list and were stable over time, and because observed structure is in accordance with expectations driven from literature for human interaction networks, we believe that the properties observed are also present in general human interaction networks. Current unfoldings include governance and accountability proposals and implementations, anthropological physics experiments, audiovisual reconnaissance, and the report of quantitative differences of textual production as connectivity of agents changes.

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

'The reason for the persistent plausibility of the typological approach, however, is not a static biological one, but just the opposite: dynamic and social. The fact that human society has been up to now divided into classes affects more than the external relations of men. The marks of social repression are left within the individual soul.' - *Adorno et al, 1969, p. 747*

I. INTRODUCTION

Human interaction networks have been studied long before modern computers. Antecedents can be found in nineteenth century works, while the foundation of the social network analysis field is generally attributed to the psychiatrist Jacob Moreno[1]. More recently, the continuous increase of stored data related to human interaction has stimulated research, and overall characterization is scattered in literature.

Reports on complex networks are usually focused on a particular measure or functionality, such as accessibility measure or detecting communities[2,3]. Although there are research on networks arisen from specific phenomenon, such as social or protein networks, there are few attempts for a general characterization and even less endeavors that take in account the temporal dynamic.

The present work is aimed at finding common and stable characteristics among (email) interaction networks. This is achieved through observations along time, which imply network evolution, a field that has received attention from the research community for more than a decade[4,5]. The evolution is represented herein by a timeline of activity snapshots with a constant number of contiguous messages. These snapshots yield a succession of networks. This approach, albeit coherent and intuitive, seems not to be explored to date[6].

We consider only directed, weighted and human interaction networks. Undirected and unweighted representations of such networks are also found in the literature and can be obtained by simplification[7]. Meaningful measures depend heavily on the model and system characteristics[8,9], therefore the most basic (and ubiquitous, and

[a)] http://ifsc.usp.br/~fabbri/; Electronic mail: fabbri@usp.br
[b)] http://automata.cc/; Electronic mail: vilson@void.cc; Also at IFSC-USP
[c)] http://www.lems.brown.edu/~rfabbri/; Electronic mail: rfabbri@iprj.uerj.br; Instituto Politécnico, Universidade Estadual do Rio de Janeiro (IPRJ)
[d)] http://lattes.cnpq.br/1065956470701739; Electronic mail: deborahantunes@gmail.com; Curso de Psicologia, Universidade Federal do Cerá (UFC)
[e)] http://lattes.cnpq.br/6738980149860322; Electronic mail: marilia.m.pisani@gmail.com; Centro de Cincias Naturais e Humanas, Universidade Federal do ABC (CCNH/UFABC)

fundamental) metrics of centrality, clustering and symmetry are analyzed altogether in principal components formation.

Typologies are the canon of scientific literature for classification of human agents[10]. Consequently, it was considered central to benefit from human typologies. This is a complex matter carefully addressed in methodology, results and appendix.

Although all networks considered originated from email lists, coherence with literature suggests that results hold for a more general class of interaction networks, such as observed in online platforms (e.g. LinkedIn, Facebook, Twitter). Indeed, text mining routines and typologies of online participants benefit from the results here presented[11,12].

Next subsection exposes related work. Section II details data while Section III is dedicated to the characterization methods. Section IV holds results and discussions, leading to Section V for conclusions and further work. Appendix A 2 develops simple graph results used in the methods, Appendix C exposes access to data and scripts, and Appendix B further introduces the typological approach. Detailed tables and figures are bundled in Support Information, and are cited upon need.

A. Related work

Works on network evolution often consider solely network growth, in which there is a monotonic increase in the number of events considered[4]. Exceptions are reported in this section, with emphasis on those more closely related to the present article.

The evolution of human interaction networks was addressed with a community focus, where the direction of edges was not taken into account[4]. In another article, 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[13]. In email list networks, scale-free properties were reported with $\alpha = 1$[14] (as are web browsing and library loans[15]), and different linguistic traces were related to weak and strong ties[7].

Unreciprocated edges often exceed 50% in the networks analyzed, which matches empirical evidence from literature[5]. No correlation of topological characteristics and geographical coordinates was found[16], therefore geographical positions were discarded in the present article. On the other hand, gender related behavior in mobile phone datasets has been reported[17] and was not considered in this article as the email messages and addresses have no gender related metadata[18].

There are research reports based on the results presented in this article, which are exposed in Section V.

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 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. A relation holds for all lists carefully considered: as the number of participants increases, the number of threads decreases. This underpins a typology sketch of networks, as exposed in Section IV D.

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

II. DATA DESCRIPTION: EMAIL LISTS AND MESSAGES

Email list messages were obtained from the GMANE email archive[18], which consists of more than 20,000 email lists and more than 130,000,000 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[7,14].

Although many lists were analyzed, four email lists were selected for thorough observation, making it easier to infer general properties. These lists, picked as representing both a diverse set and usual lists, are:

- Linux Audio Users list[20], with participants holding hybrid artistic and technological interests, from different countries. English is the language used the most. Abbreviated as LAU from now on.
- Linux Audio Developers list[21], with participants are from different countries, and English is the language used the most. A more technical and less active version of LAU. Abbreviated LAD from now on.
- Development list for the standard C++ library[22], with computer programmers from different countries. English is the language used the most. Abbreviated as CPP from now on.
- List of the MetaReciclagem project[23], with Brazilian activists holding digital culture interests. Portuguese is the most used language, although Spanish and English are also incident. Abbreviated 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.

III. CHARACTERIZATION METHODS

The email lists and the networks generated from them were characterized by: 1) statistics of activity along time, from seconds to years; 2) sectioning of the networks in hubs, intermediary and periphery vertices; 3) topological measures prominence regarding dispersion; 4) iterative visualization and data mining; 5) typological speculation about networks and participants. Each of these procedures are described below.

Distribution of activity among participants received little dedicated attention, but is considered through almost all results of this article, as degree and strength are highly correlated to activity.

A. Time activity statistics

Messages were counted along time with respect to seconds, minutes, hours, days of the week, days of the month, and months of the year. This resulted in histograms from which obvious patterns can be observed. Furthermore, the ration $\frac{b_h}{b_l}$ of the highest and lowest incidences on the histograms served as a hint of how the distribution relates to the uniform distribution.

Measures of mean and dispersion were taken using circular statistics. For such, each *measure* (data point) is represented as a complex number of unit magnitude $z = e^{i\theta} = \cos(\theta) + i\sin(\theta)$, where $\theta = measure\frac{2\pi}{period}$. Moments m_n, lengths of moments R_n, mean angle θ_μ, and rescaled mean angle θ'_μ are defined as:

$$\begin{aligned} m_n &= \frac{1}{N}\sum_{i=1}^{N} z_i^n \\ R_n &= |m_n| \\ \theta_\mu &= Arg(m_1) \\ \theta'_\mu &= \frac{period}{2\pi}\theta_\mu \end{aligned} \tag{1}$$

The 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 $\delta(z)$:

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

Although very simple, such measures enables both comparison of the email networks exposed in this article, and the deliverance of a benchmark to which other interaction networks can be compared.

B. Interaction network

Interaction networks can be modeled both weighted or unweighted, both directed or undirected[14,24,25]. Networks in this article are directed and weighted, the more informative of trivial possibilities, i.e. we did not observe directed unweighted, undirected weighted, and undirected unweighted representations of the interaction networks. The 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 $A \to B$. 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 article uses the information network as described above and depicted in Figure 1. Edges in both directions are allowed. Each time an interaction occurs, one is added to the edge weight. Self-loops were regarded as non-informative and discarded. These human social interaction networks are reported in the literature as exhibiting scale-free and small world properties, as expected for (some) social networks[1,14].

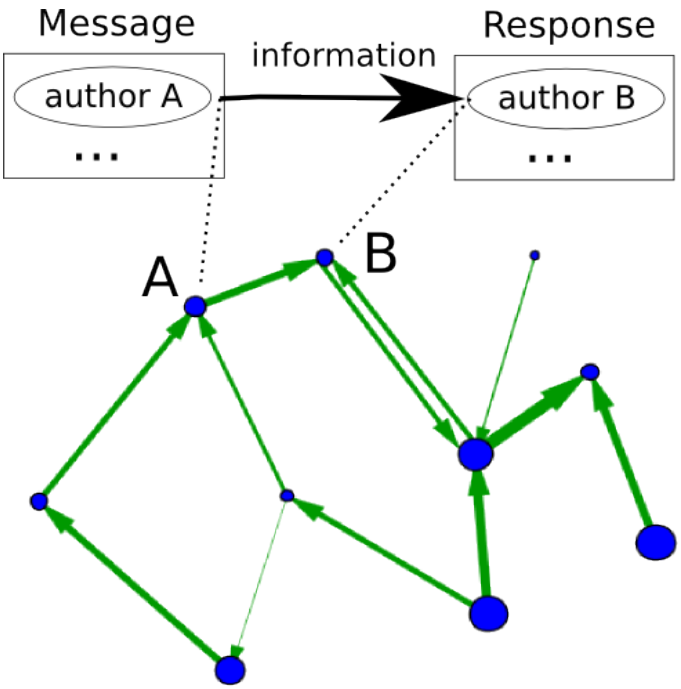

FIG. 1. Formation of interaction network from email messages. Each vertex represents a participant. A reply message from participant B to a message from participant A is regarded as evidence that B received information from A and yields a directed edge. Multiple messages add "weight" to a directed edge.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

Edges can be created from all antecedent message authors on the message-response thread to each message

author. We only linked the immediate antecedent to the new message author, both for simplicity and for the valid objection that in adding two edges, $x \to y$ and $y \to z$, there is also a weaker connection between x and z. Potential interpretations for this weaker connection are: double length, half weight or with one more "obstacles". This suggests pertinence of centrality measures that account for the connectivity with all nodes, such as betweenness centrality and accessibility[2,26].

C. Erdös sectioning

In scale-free networks, the periphery, intermediary and hubs sectors can be derived from comparison with an Erdös-Rényi network with the same number of edges and vertices[27], as depicted in Figure 2. As a shorthand, this procedure is from now on called *Erdös sectioning*, and the resulting sectors the *Erdös sectors* or *primitive sectors*.

The degree distribution $\widetilde{P}(k)$ of an ideal scale-free network $\mathcal{N}_f$ 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:

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

If $\mathcal{N}_f$ is directed and has no self-loops, the probability of an edge between two arbitrary vertices is $p_e = \frac{z}{N(N-1)}$ (see Appendix A 2). 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:

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

The lower degree fat tail represents 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: 1) vertices so connected that they are virtually inexistent in networks connected at pure chance (e.g. 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. Vertices with degree values predicted as the most abundant if connections are created by pure chance, near the average, and less frequent in scale-free phenomena, are classified as intermediary.

To ensure statistical validity of the histograms, bins can be chosen to contain at least η vertices of the real network. Thus, each bin, starting at degree k_i, spans $\Delta_i = [k_i, k_j]$ degree values, where j is the smallest integer with which there are at least η vertices with degree larger than or equal k_i, and less than or equal k_j. This changes equation 3 to:

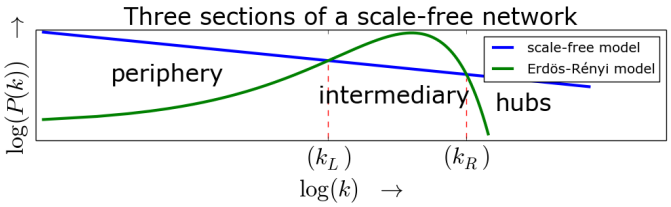

FIG. 2. Degree distribution of scale-free and Erdös-Rényi ideal networks. The latter has more intermediary vertices, while the former has more peripheral and hub vertices. Sector borders are given by the two intersections $k_{\Leftarrow}$ and $k_{\Rightarrow}$ of the connectivity distributions. Characteristic degrees are in compact intervals of degree: $[0, k_{\Leftarrow}]$, $(k_{\Leftarrow}, k_{\Rightarrow}]$, $(k_{\Rightarrow}, k_{max}]$ for the Erdös sectors (periphery, intermediary and hubs).

$$\sum_{x=k_i}^{k_j} \widetilde{P}(x) < \sum_{x=k_i}^{k_j} P(x) \Rightarrow \text{i is intermediary} \qquad (5)$$

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:

$$\hat{P}(k^{way}) = \binom{N-1}{k^{way}} p_e^k (1-p_e)^{N-1-k^{way}} \qquad (6)$$

where *way* can be *in* or *out*. In and out strengths (s^{in}, s^{out}) are divided by $\overline{w}$ and compared also using $\hat{P}$. Notice 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 total degree.

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

$$\begin{aligned} e_{\gamma,1} &= \{\, i \mid \overline{k}_{\gamma,L} \ge \overline{k}_{\gamma,i} \} \\ e_{\gamma,2} &= \{\, i \mid \overline{k}_{\gamma,L} < \overline{k}_{\gamma,i} \le \overline{k}_{\gamma,R} \} \\ e_{\gamma,3} &= \{\, i \mid \overline{k}_{\gamma,i} < \overline{k}_{\gamma,R} \} \end{aligned} \qquad (7)$$

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

$$\begin{aligned}
\overline{k}_{1,i} &= k_i \\
\overline{k}_{2,i} &= k_i^{in} \\
\overline{k}_{3,i} &= k_i^{out} \\
\overline{k}_{4,i} &= \frac{s_i}{\overline{w}} \\
\overline{k}_{5,i} &= \frac{s_i^{in}}{\overline{w}} \\
\overline{k}_{6,i} &= \frac{s_i^{out}}{\overline{w}}
\end{aligned} \tag{8}$$

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 can be defined. 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 (usually) does not reach 100%, which is indicated by a black line in the figures of Appendix E.
- 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 total members may add more than 100%, which is indicated by a black line in the figures of Appendix E.
- 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 so 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 by 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 7, these compound criteria C_δ, with δ integer in the interval $1 < \delta < 6$ can be described as:

$$\begin{aligned}
C_1 &= \{c_{1,\phi} = \{i \mid i \in e_{\gamma,\phi},\ \forall\, \gamma\}\} \\
C_2 &= \{c_{2,\phi} = \{i \mid \exists\ \gamma : i \in e_{\gamma,\phi}\}\} \\
C_3 &= \{c_{3,\phi} = \{i \mid i \in e_{\gamma,\phi'},\ \forall\, \gamma,\ \forall\, \phi' \geq \phi\}\} \\
C_4 &= \{c_{4,\phi} = \{i \mid i \in e_{\gamma,\phi'},\ \forall\, \gamma,\ \forall\, \phi' \leq \phi\}\} \\
C_5 &= \{c_{5,\phi} = \{i \mid i \in e_{\gamma,\phi'},\ \forall\, \gamma, \\
&\qquad \forall\, (\phi'+1)\%4 \leq (\phi+1)\%4\}\} \\
C_6 &= \{c_{6,\phi} = \{i \mid i \in e_{\gamma,\phi'},\ \forall\, \gamma, \\
&\qquad \forall\, (\phi'+1)\%4 \geq (\phi+1)\%4\}\}
\end{aligned} \tag{9}$$

The simplification of all the compound possibilities to the small set listed above can be formalized in strict mathematical terms, but this was considered out of the scope for current interests. Notice that the exclusivist cascade is the same sectioning that an inclusivist cascade from periphery to hubs (with inverted order of sectors precedence). Compound criteria can be used to examine network sections in the case of a low number of messages, such as in the last figures of Appendix E.

Results from applying the Erdös sectioning, i.e. the achievement of well defined sectors by comparison of the real network with the Erdös Rényi model, are reported in Section IV B and Appendix E.

D. Topology metrics for Principal Component Analysis

The topology of the networks was observed by Principal Component Analysis (PCA[28]) outcomes with a small selection of the most basic and fundamental measurements for each vertex:

- 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 the vertex i. Betweenness centrality index considered directions and weight, as specified in[29].

In order to capture symmetries in the activity of participants, the following metrics were introduced for a vertex i (see Section IV C):

- 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 and edge from x to y, 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: $\sigma_i^{asy} = \sqrt{\frac{\sum_{j \in J_i} [\mu_{asy} - (e_{ji} - e_{ij})]^2}{k_i}}$.
- 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: $\sigma_i^{dis} = \sqrt{\frac{\sum_{j \in J_i} [\mu_{dis} - \frac{(w_{ji} - w_{ij})}{s_i}]^2}{k_i}}$.

E. Evolution of the networks

The evolution of the networks was observed within a fixed number of messages, the window size ws. This same number of contiguous messages is considered with different shifts in the message timeline to obtain snapshots. Each snapshot was used to both do the Erdös sectioning (see Section III C) and to perform PCA with topological measures (see Section III D). Furthermore, this approach to evolution was used for visualization (see Section III F) and for typological speculations (see Section III G). The ws used were 50, 100, 200, 400, 500, 800, 1000, 2000, 2500, 5000 and 10000. Within a same ws, the number of vertices and edges vary in time, as do other network characteristics.

F. Visualization of network evolution

The evolution of the networks was visualized with animations, image galleries and online gadgets made for this research[30–32]. Such visualizations were crucial to guide research into the most important features of network evolution, and prompted us to capture the prominence of topological metrics along time using mean and standard deviations (see Section III D and Appendix D 1), in addition to the size of the three sectors in a timeline fashion (Appendix E). Visualization of network structure was specially useful as part of the email lists data mining, from which parts of relevant structures and results were driven. See Appendix I A for further directions about visualization and text mining concerning the results herein presented, with dedicated articles.

G. Typology speculation

Qualitative, typological speculations were the result of all the methodology. More specifically, the Erdös sectors was regarded as yielding a *primitive typology* of human agents in social contexts with at least dozens of participants. Visualization and data mining were efficient in fueling speculations about the qualities of each type. The stability found in principal components of topological measures suggested that this primitive typology is general for (practically) all human contexts. Also, the first author is a developer with cultural interests and has been directly and indirectly part of these communities for more than a decade. This gave further safety about assumptions although care was taken not to regard this acquaintanceship as primary source or as definitive evidence. The interested reader should see Appendix B for further context, as it might be considered audacious to bridge from a physics-based, quantitative classification to a qualitative and speculative approach.

IV. RESULTS AND DISCUSSION

Remarkable features from the analysis of the four email lists are:

- The activity along time is practically the same for all lists (Section IV A) and suggests stable patterns.
- The fraction of participants in each Erdös sector is stable and can be observed even with very few messages (Section IV B).
- The measures combine into principal components in the same way for all lists and all snapshots. Furthermore, symmetry related measures, defined in this article, present more dispersion than the usual clustering coefficient (Section IV C).
- Typology speculations are immediate from results (Section IV D).

The reminder of this section contains a detailed discussion of these results and Support Information holds tables and figures from which such results can be observed.

A. Constancy and discrepancy of activity along time

One remarkable feature from the analysis is that the activity along time is practically the same for all lists. Circular mean and standard deviation for each time scale are given in Table II.

1. Seconds and minutes

Messages were slightly more evenly distributed in all lists than in simulations using uniform distribution[33]: for

TABLE II. The rescaled circular mean θ'_μ, the standard deviation $S(z)$, the variance $Var(z)$, the circular dispersion $\delta(z)$ and the relation of maximum and minimum incidence at each time unit $\frac{max(incidence}{min(incidence}$. The Section III A exposes the theoretical background considered of directional (or circular) statistics.

scale	θ'_μ	$S(z)$	$Var(z)$	$\delta(z)$	$\frac{max(incidence)}{min(incidence)}$
seconds	–//–	3.31	1.00	28205.46	1.26
minutes	–//–	3.18	0.99	12275.59	1.27
hours	-9.39	1.48	0.67	3.91	11.18
weekdays	-0.17	1.83	0.81	12.66	2.59
month days	-0.31	3.14	0.99	9541.67	1.93
months	0.15	2.34	0.93	115.49	1.50

both seconds and minutes $\frac{max(incidence)}{min(incidence)} \in (1.26, 1.275]$ in the lists, while Simulations reach these values but have in average more discrepant higher and lower peaks $\xi = \frac{max(incidence')}{min(incidence')} \Rightarrow \mu_\xi = 1.2918$ and $\sigma_\xi = 0.04619$. Therefore, the incidence of messages at each second of a minute and at each minute of an hour was considered uniform, i.e. no trend was detected and the pattern is the uniformity. Circular dispersion is maximized and mean has little meaning.

2. Hours of the day

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. See Table VIII for the fraction of activity in 1h, 6h and 12h periods along the day.

3. Days of the week

Higher activity was observed during weekdays, specially for the CPP and MET lists (see Table IX). The decrease of activity on weekends reaches at least one third, and two thirds in extreme cases.

4. Days along the month

Variation of activity in the days along the month is less prominent, one cannot point much more than a - probably not statistically relevant - tendency of first and second weeks to be more active. The most important trait seems to be homogeneity. Last days of the month (29, 30 and 31) are not present in every month, and observed activity is proportional to incidence rates. See Table X for further insights.

TABLE III. .

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

5. Months and larger divisions of the year

Activity is concentrated in Jun-Aug for MET and LAD, and in Dec-Mar for CPP, LAU and LAD (see Table XI). These observations fit academic calendars, vacations and end-of-year holidays.

6. Activity long years

Literature reports relation of number of participants and hubs stability to community longevity[4]. Also, activity should follow trends such as economic, climate and technological. Periodicity should be observed to some extent. Nevertheless, the time period examined was not sufficient for the analysis of activity along the years (see Section V).

B. Scalable fat-tail structure: constancy of membership fractions in the Erdös sectors

There is a concentration of hub activity and of vertex with few connections. Table VII exposes this expected distribution of activity among participants in a scale-free context.

The distribution of vertices in the hubs, intermediary, periphery Erdös sectors defined in Section III C 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 the four email lists, to which are dedicated the various figures in Appendix E. Typically, $\approx 5\%$ of the vertices are found to be hubs, $\approx [15-35]\%$ are intermediary and $\approx [60-80]\%$ are peripheral, which is consistent with the literature[34]. These results hold for total, in and out degrees and strengths. Stable distributions can also be obtained for 100 of less messages if classification of the three sectors is performed with one of the compound criteria established in Section III C. The networks hold their basic structure with as few as 10-50 messages, concentration of activity and the abundance of low-activity participants take place even with very few messages, which is highlighted in the last figures of Appendix E. A minimum window size for observation of more general properties might be inferred by monitoring the giant component and the degeneration of the hub, intermediary and peripheral sections.

For the histograms used in the classification process (see Section III C), the use of at least η vertices for each bin did not yield significant differences. That was understood as a consequence of the observation scale (see Appendix A 2).

C. Variance prevalence of centrality over symmetry and symmetry over clusterization

The topology was analyzed using standard, well-established metrics of centrality and clustering. We also introduced symmetry measures because there is evidence that they are relevant in social contexts[5]. The contribution from the distinct metrics to network topology is very similar for all the networks, and did not vary with time. This stability in network behavior is remarkable, as can be noticed by the small standard deviations of the contribution of each metric to principal components (from usual Principal Component Analysis[28]) along time of Tables IV-VI.

The principal component exhibited a pondered sum of centrality measures: degrees, strengths and betweenness centrality. This composition of the 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. Clustering coefficient is presented in almost perfect orthogonality. Dispersion was more prevalent in symmetry related measures than clustering coefficient.

Degree and strength are highly correlated, with Spearman correlation coefficient $\in [0.95, 1]$ and Pearson coefficient $\in [0.85, 1)$ for $ws > 1000$. The typical PCA plot for the two first components exhibited in Table V is shown in the second system of Figure 4, where each vertex is colored according to the sector they belong. As expected, peripheral vertices have very low values in the first component (centrality related) and greater dispersion in the second component (clustering related). The PCA plot in the third system of Figure 4, where all metrics are considered, reflects symmetry metrics relevance for the variance. This can be observed in Table VI where the clustering coefficient is only relevant for the third principal component (with contributions from out-degree and out-strength). We concluded that the symmetry-related measurements can be more meaningful in characterizing interaction networks (and their participants) than the clustering coefficient, especially for hubs and intermediary vertices.

D. Primitive typology from Erdös sectors

Consider the sector to which a participant belongs as her/his type. General considerations of this typology are:

- A participant belongs to many networks (his family, family of a friend, the email list to which he belongs, work friends, etc.).
- A participant might belong to all three sectors at the same time. Actually, it is reasonable to assume that almost every person belongs to all three sectors for some of the networks.
- It is given by construction that a participant is of just one type at a network in a given snapshot (exception for the compound inclusive and exclusive criterion given in Section III C).
- The participant often transitions from one sector to another within a network.
- Reported stability of network structure arises from activity with continuous change of the participant types.
- Type often changes with the scale of the activity window. For example, two snapshots of 200 and 10000 messages will probably reveal participants with more than one type.

This ephemerality of the human type is in accordance with canonical literature on psychological types. The observance of often transitions between types is even though of as an argument against prejudice[41]. Visualizations and raw data manipulations suggests further typological peculiarities. These are initial observations, which inspired this article and other ongoing research[11,35]:

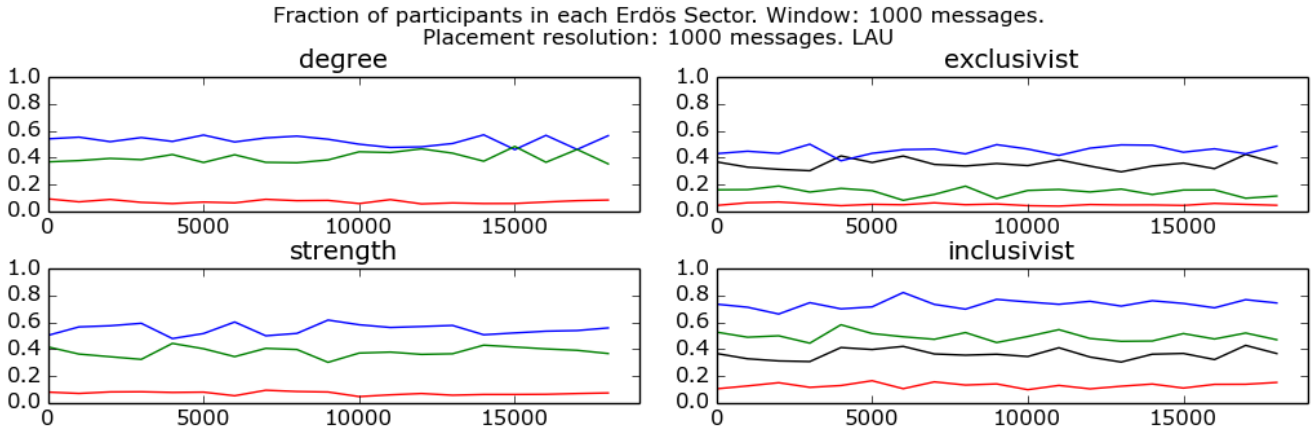

FIG. 3. Fraction of agents in each Erdös sector. Hubs, intermediary and periphery fractions are represented in red, green and blue. In this figure, simple criteria employed using degree and strength. Exclusivist and inclusivist compound criteria have black lines for fraction of vertices without class and with more than one class, respectively. See Appendix E for a collection of such timeline figures with all simple and compound criteria and selected measures.

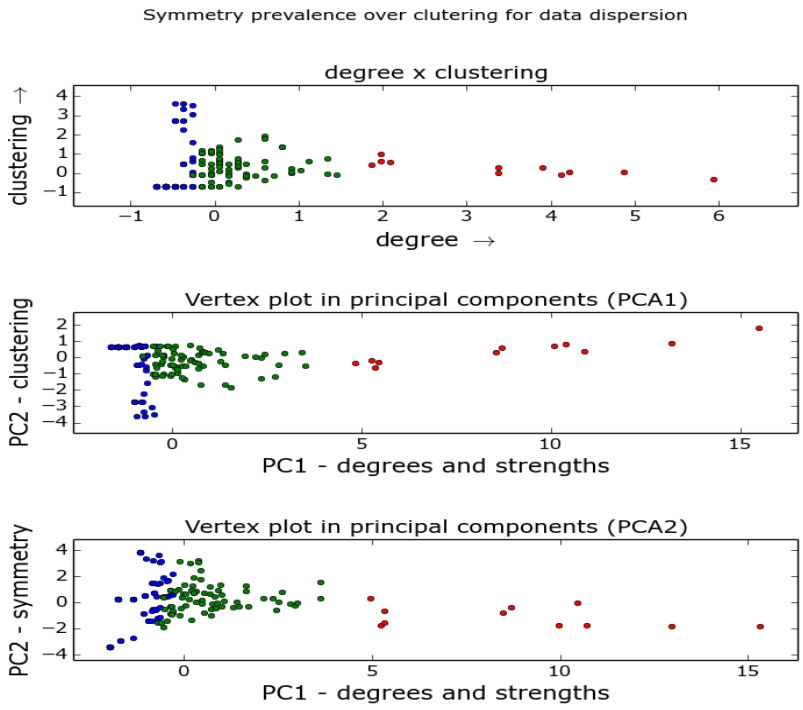

FIG. 4. Prevalence of symmetry measures over clustering for data dispersion. This was considered an important result of PCA analysis, together with stability on component formation. First plot holds the usual degree versus clustering coefficient layout. Second plot is similar, but the first component is an average of centrality measures (see Table V). Third plot resulted from applying PCA with symmetry measures (see Table VI), where greater dispersion can be observed in all sectors, although more pronounced in intermediary and hubs sectors. For this figure, a window size of $ws = 1000$ messages was used from LAU list. The general layout is recurrent and consistent with the literature: most connected vertices have low clusterization while higher clusterization is gradually more incident as the number of connections is lowered. Similar structures were observed in all window sizes $ws \in [500, 10000]$ and for networks of other email lists, which exposes a common relation held by degree, strength and betweenness measures to clustering coefficient. In the third plot, clustering coefficient is omitted as it is only relevant for third component. In this case, second component is representative of symmetry measurements of vertex interactions. Dispersion suggests symmetry related measures are more powerful for characterizing interaction networks than clustering coefficient, specially for hubs and intermediary vertices.

- Core hubs usually have intermittent activity. Very stable activity was found on MET hubs, which motivated its integration to this work. There are reports in the literature of greater stability of participation in smaller communities[4], which is the reason why the smaller number of participants in MET was considered coherent with the stable activity of hubs.

- 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. For example, 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, is found only in peripheral and intermediary sectors.

This "primitive typology", characterized by peripheral, intermediary and hub types, can be further scrutinized using concepts from other typologies, such as Meyer-Briggs, Pavlov or F-Scale. This is not a direct result of numeric analysis, it is a description refinement of the found structure, in typological terms, and coherent with complex networks literature. Although initial, this bridges human and exact sciences in the most pertinent way authors were able to, as is herein considered a result.

Another typology suggested by results is about the networks themselves. This and paired articles[11,35] reports a dipole in human interaction network types: networks with few and stable agents, and (relative) many threads per amount of messages contrasts with networks with many agents of intermittent activity and (relative) few threads per amount of messages. This network typology is endorsed by the seminal Nature Letter by Palla, Barabási and Vicsek[4], which suggests that the smaller size of MET community is responsible for the stronger hubs observed.

V. CONCLUSIONS AND FURTHER WORK

The characterization of interaction networks resulted from stability observations. Along temporal activity statistics, this work reports the stability of the formation of principal components and of the relative sizes of periphery, intermediary and hubs Erdös sectors. These results suggested typologies for both agents and the networks.

Observed systems were coherent with literature in different aspects, such as concentration of activity or such as the clusterization versus connectivity pattern. Even so, analysis of data from other virtual environments, such as Facebook, Twitter and LinkedIn, might help understanding how general are these structures and what are proper uses.

Further work should inspect other topological measures in each sector. The subtraction $\widetilde{P}(k) - P(k)$ (see Section III G) results in two positive clusters for periphery and hubs, and a negative cluster for intermediary vertices. This might support classification of the three sectors by clustering, a more traditional approach to classification. Observance of attributes with greater contribution to principal components of LDA should reveal best chances to present these three sections as clusters in the network measurements space. Another possibility, specially for a brute-force characterization of such sectors, is to remove vertices with degree close to $k_{\Leftarrow}$ or $k_{\Rightarrow}$ depicted in figure 2.

Stability here reported eases tipologization of both outliers and usual participation patterns (see Sections I A and IV D). A paired article with this one is reports significant differences in the textual production of each Edös sector[11]. Resulting knowledge purposes networks and participant typologies. Another paired article describes a network time evolution visualization method[35]. The usage of such results are taking place in software packages, linked data, electronic government technologies, and anthropological physics experiments[36–40].

ACKNOWLEDGMENTS

Renato Fabbri is grateful to CNPq (process: 140860/2013-4, project 870336/1997-5), United Nations Development Program (PNUD/ONU, contract: 2013/000566; project BRA/12/018) 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[41]. Authors thank GMANE creators and maintainers, specifically: GMANE is run by Lars Magne Ingebrigtsen, 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 Sjgren. Authors thank referred email lists communities and welcome feedback as core contribution to this, and similar, research.

Appendix A: More graph-related results

1. Edge existence probability in a directed network without self-loops

Be $\mathcal{N}$ a directed network without self-loops with z edges and N vertices. The probability that an edge exists between two arbitrary vertex is $p_e = \frac{z}{max(\text{number of edges}|\text{ N vertices})}$, where $max(\text{number of edges}|\text{ N vertices}) = 2[(N-1) + (N-2) + ... + 1] = 2[\sum_1^{N-1} i] = 2[\frac{N(N-1)}{2}]$ is the maximum number of edges for a network with N vertices. Therefore:

$$\begin{aligned} p_e &= \frac{z}{max(\text{number of edges}|\text{ N vertices})} \\ &= \frac{z}{2[(N-1)+(N-2)+...+1]} = \frac{z}{2\frac{N(N-1)}{2}} \\ p_e &= \frac{z}{N(N-1)} \end{aligned} \tag{A1}$$

2. Effect of scale in histogram smoothing

There are between 20 and 200 participants in the message window sizes used to derive most of the results ($ws \in [200, 1500]$ messages). As peripheral vertices are abundant and span few degrees, there are more than η vertices with each low degree value. For the case of higher degrees, one should consider that with the ws used, each participant is $p \in [0.1\%, 0.5\%]$ of all participants. Therefore, if incident connectivity is very improbable in an Erdös Rényi network (less than p, the probability that a single participant represents when the histogram is normalized to the density function), than it is not an intermediary connectivity, but a hub. Therefore, using at least η vertices for each bin did not impact the results.

Appendix B: Typologies

There are other ways to split and characterize networks. 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 (the radius is the minimum maximum distance to all vertices , i.e. 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 (diameter being 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 such criteria, the center can often be a factor of 4 times larger than the periphery and the intermediary group often exceeds 90% of the nodes[43].

Models of human dynamics can be used to predict and classify activity. In this case, human 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 bursts of rapidly occurring events[15,44].

Typologies can also be conveniently adapted from psychiatric, psychological and psychoanalytic theories. Concerning empirical research, Theodor Adorno is a core conceiver of an one-of-a-kind typology that resulted from observing authoritarian personality traces to detect Nazism adoption, antisemitism and potential fascists, depicted as an authoritarian syndrome[41].

Other classic typologies of interest 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), each individual manifesting one of these four modes as dominant, and each mode expressed primarily as introverted or extroverted[45]. Myers-Briggs Type Indicator extrapolated Jungian theories into a questionnaire and added perceiving and judging as a fourth dipole[46]. 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[47,48].

It was considered central to benefit from key human typologies, both by describing types and by further characterizing classes in the terms encountered. A primitive physics-based typology is described in Section IV D as a consequence of the periphery, intermediary and hub sectors yielded by comparing the real networks with the Erdös Rényi model. Also, ethic and moral issues are developed by such legacy. For example, Adorno et al. conceptualized that personality is dynamic, not static or immutable, and that recognizing this was important for an ethic empirical study of human authoritarian traces[41]. Indeed, this dynamic typological approach is so vital to secure an ethic study of human systems that our epigraph is devoted to make this point explicit.

Appendix C: Data and scripts

Messages are downloaded from the 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:

- "From" field, as it specifies the sender of the message, in the usual format of "First_name Last_Name $< email >$".
- "Date" field, which is given with the resolution of a second.

- "Message-ID", important to state antecedent/consequent relation between messages and therefore from an author to a replier.
- "References", has the ID of the message it is an answer to, if any, and earlier messages in the thread.

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[11].

1. Python scripts

Basic constructs for obtaining all results are the product of scripts written in the Python programming language. These are kept in a public git repository for backup and sharing with research community[42]. Core scripts, for deriving structures and results exhibited in this article, are in the LEIAME file.

2. Third party libraries and software

The programming framework used is mainly Python-based, with emphasis on usual scientific tools. 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.

Appendix D: Tables

1. PCA tables

TABLE IV. 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 a weighted average of degree and betweenness centrality. The second component is mostly clustering coefficient. The first and second components represent more than 95% of total variance. The λ bottom line holds the percentage of total variance attributed to each component.

	PC1		PC2		PC3	
	μ	σ	μ	σ	μ	σ
d	**48.02**	1.39	2.82	1.74	48.09	0.32
cc	4.12	2.94	**90.45**	3.98	3.98	0.77
bt	**47.87**	1.55	6.74	4.08	47.93	0.46
λ	64.67	0.52	33.26	0.23	2.08	0.40

TABLE V. Principal components composition in percentages. LAU list, $ws = 1000$ messages in 20 disjoint positioning was used for statistics. First component is a weighted average of degree and strength and betweenness centrality. The second component is mostly related to the clustering coefficient. The first and second components represent more than 90% of the variance.

	PC1		PC2		PC3	
	μ	σ	μ	σ	μ	σ
d	**14.58**	0.14	0.43	0.35	1.51	1.08
d^{in}	**14.12**	0.14	1.71	1.22	17.80	6.20
d^{out}	**13.95**	0.12	2.80	1.83	21.15	5.62
s	**14.48**	0.13	0.78	0.65	5.51	4.71
s^{in}	**14.10**	0.14	2.17	1.28	17.32	6.11
s^{out}	**14.05**	0.13	2.08	1.14	19.31	4.86
cc	0.99	0.70	**83.38**	4.83	2.75	1.62
bt	**13.73**	0.19	6.65	1.31	14.66	10.14
λ	81.80	0.83	12.53	0.09	3.24	0.62

2. Tables for activity along time and among participants

TABLE VI. Principal components formation with symmetry-related metrics (see Section III D). LAU list, $ws = 1000$ messages in 20 disjoint positioning was used for statistics. In this case, clusterization is pushed to the third principal component. The second component is primarily derived from symmetry measurements, but also out-degree and out-strength, and disequilibrium standard deviation. Betweenness centrality again has a role similar to degree, but weaker. The clusterization component combines with disequilibrium, while asymmetry is combined to out-degree and out-strength. The three components have in average 80.36% of the variance.

	PC1		PC2		PC3	
	μ	σ	μ	σ	μ	σ
d	**11.51**	0.42	2.00	0.76	2.39	0.49
d^{in}	**11.45**	0.34	2.86	0.91	1.68	0.67
d^{out}	**10.68**	0.60	**7.43**	1.00	3.00	1.02
s	**11.37**	0.42	1.75	0.71	4.31	0.63
s^{in}	**11.33**	0.35	2.39	1.10	3.69	0.86
s^{out}	**10.74**	0.55	**6.14**	1.05	4.75	0.98
cc	0.91	0.64	2.68	1.67	**22.27**	6.43
bt	**10.87**	0.38	1.17	0.93	4.03	1.42
asy	3.99	1.45	**18.13**	1.67	2.55	1.77
μ_{asy}	4.15	1.40	**17.07**	1.78	2.49	1.67
σ_{asy}	1.21	0.67	**17.49**	0.79	3.29	2.33
dis	5.78	0.51	1.94	1.28	**24.75**	3.73
μ_{dis}	0.79	0.49	**14.00**	1.14	3.73	3.13
σ_{dis}	5.18	0.72	4.93	2.48	**17.04**	4.78
λ	51.09	1.07	20.04	1.31	9.23	6.63

TABLE VII. Distribution of activity among agents. First column is dedicated to percentage of messages sent by the most active participant. 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 messages.

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

TABLE VIII. Percentage of activity ($100\frac{\text{counted messages}}{\text{total messages}}$) in each hour, 6 hours and 12 hours. Maximum activity rates are in bold. In 1h 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.

	CPP			MET			LAU			LAD		
	1h	6h	12h	1h	6h	12h	1h	6h	12h	1h	6h	12h
0h	3.66			2.87			3.58			4.00		
1h	2.76			1.77			2.22			2.52		
2h	1.79	10.67		1.04	7.15		1.63	10.14		1.79	10.77	
3h	1.10			0.64			1.06			1.06		
4h	**0.68**			0.47			0.84			0.75		
5h	0.69		33.76	**0.38**		29.33	**0.82**		36.88	**0.66**		33.13
6h	0.83			0.72			1.17			0.85		
7h	1.24			1.33			2.37			1.56		
8h	2.28	23.09		2.67	22.18		3.54	26.74		2.96	22.36	
9h	4.52			4.40			6.04			4.68		
10h	6.62			6.29			**6.83**			5.93		
11h	**7.61**			6.78			6.79			6.40		
12h	6.44			**7.33**			6.11			**6.41**		
13h	6.04			7.08			6.26			6.12		
14h	6.47	**37.63**		7.09	**42.22**		6.38	**35.65**		6.33	**37.25**	
15h	6.10			7.14			5.93			5.98		
16h	6.22			6.68			5.52			6.40		
17h	6.36		**66.24**	6.89		**70.66**	5.46		**63.12**	6.02		**66.87**
18h	6.01			5.99			5.24			5.99		
19h	5.02			5.23			4.52			5.03		
20h	4.85	28.61		4.98	28.44		4.55	27.46		4.63	29.63	
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		

TABLE IX. Percentage 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.

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
CPP	17.06	17.43	17.61	17.13	16.30	6.81	7.67
MET	17.53	17.54	16.43	17.06	17.46	7.92	6.06
LAU	15.71	15.80	15.88	16.43	15.13	10.13	10.91
LAD	14.91	17.73	17.01	15.40	14.25	10.39	10.30

TABLE X. Activity along the days of the month. The pattern is to have no clear prevalent period. One might point a slight tendency for the first two weeks to be more active, although this table does not present statistical foundation for such an assumption. For the scope of this study, differences of activity along the month is assumed to be inexistent.

	CPP			MET			LAU			LAD		
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
1	3.19			3.01			3.34			3.22		
2	3.07			3.38			3.38			3.42		
3	3.20			3.55			3.20			2.87		
4	3.63	23.05		4.34	25.16		3.52	23.06		2.91	21.96	
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		
8	3.12		45.63	3.36		48.08	3.62		47.31	3.72		46.70
9	2.57			3.44			3.82			3.97		
10	2.92			3.17			3.06			3.77		
11	3.54	22.57		3.88	22.92		3.11	24.25		3.27	24.73	
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		
15	3.35			2.72			3.23			3.37		
16	3.77			2.96			2.94			3.37		
17	3.45			3.01			3.02			2.95		
18	3.47	23.02		3.39	21.87		3.63	22.84		3.22	22.82	
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		
22	2.88		46.31	2.92		43.56	3.80		44.01	3.07		46.00
23	4.01			3.27			3.03			3.06		
24	3.13			2.92			2.31			2.72		
25	3.57	23.29		2.83	21.69		2.38	21.17		3.16	23.18	
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		
29	3.68			2.93			3.34			3.15		
30	2.76	8.06	8.06	3.14	8.36	8.36	3.75	8.68	8.68	2.71	7.30	7.30
31	1.63			2.29			1.60			1.45		

TABLE XI. 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 months slots, so every month has the same time of occurrences.

	CPP					MET					LAU					LAD				
	m.	b.	t.	q.	s.	m.	b.	t.	q.	s.	m.	b.	t.	q.	s.	m.	b.	t.	q.	s.
Jan	8.70	17.00	**27.23**	**36.48**	**54.26**	4.88	11.01	16.90	23.32	47.74	10.22	**19.56**	**28.23**	**35.09**	49.17	11.23	18.49	26.43	36.04	**57.95**
Fev	8.29					6.13					9.34					7.26				
Mar	**10.23**	**19.49**				5.89	12.31				8.67	15.52				7.94	17.55			
Apr	9.26		27.03			6.42		30.84			6.85		20.94			9.61		**31.51**		
Mai	9.41	17.78		33.46		10.46	**24.42**		**47.83**		7.27	14.09		30.37		8.94	**21.91**		**37.56**	
Jun	8.37					**13.96**					6.81					**12.97**				
Jul	8.70	15.68	22.94		45.73	13.23	23.41	**31.16**		**52.26**	8.96	16.28	24.47		**50.82**	9.02	15.65	22.29		42.05
Ago	6.98					10.28					7.31					6.63				
Set	7.26	15.36		30.06		7.75	16.80		28.86		8.18	16.24		34.54		6.63	12.38		26.40	
Oct	8.10		22.80			9.05		21.10			8.06		26.36			5.74		19.77		
Nov	7.86	14.69				7.46	12.06				7.63	18.30				7.63	14.02			
Dec	6.81					4.59					**10.66**					6.39				

Appendix E: Figures of vertex classification fractions as the network evolves

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.

[1] M. Newman, *Networks: an introduction* (Oxford University Press, 2010).
[2] B. Travençolo and L. d. F. Costa, "Accessibility in complex networks," Physics Letters A **373**, 89–95 (2008).
[3] M. E. Newman, "Modularity and community structure in networks," Proceedings of the National Academy of Sciences **103**, 8577–8582 (2006).
[4] G. Palla, A.-L. Barabási, and T. Vicsek, "Quantifying social group evolution," Nature **446**, 664–667 (2007).
[5] 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).
[6] P. Doreian and F. Stokman, *Evolution of social networks* (Routledge, 2013).
[7] K. Marek-Spartz, P. Chesley, and H. Sande, "Construction of the gmane corpus for examining the diffusion of lexical innovations," (2012).
[8] M. E. J. Newman, "The structure and function of complex networks," SIAM REVIEW **45**, 167–256 (2003).
[9] M. E. Newman, "Analysis of weighted networks," Physical Review E **70**, 056131 (2004).
[10] K. Gergen and M. Gergen, *Historical social psychology* (Psychology Press, 2014).
[11] R. Fabbri, "A connective differentiation of textual production in interaction networks," (2013), http://arxiv.org/abs/1412.7309.
[12] R. Fabbri, "Participant typologies derived from textual and topological features in interaction networks," (2013).
[13] R. Albert and A.-L. Barabási, "Topology of evolving networks: local events and universality," Physical review letters **85**, 5234 (2000).
[14] 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.
[15] 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).
[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] L. M. Ingebrigtsen, "Gmane," (2008).
[19] Wikipedia, "Gmane — Wikipedia, the free encyclopedi,".
[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] L. d. F. Costa, F. A. Rodrigues, G. Travieso, and P. Villas Boas, "Characterization of complex networks: A survey of measurements," Advances in Physics **56**, 167–242 (2007).
[27] M. O. Jackson,.
[28] I. Jolliffe, *Principal component analysis* (Wiley Online Library, 2005).
[29] U. Brandes, "A faster algorithm for betweenness centrality*," Journal of Mathematical Sociology **25**, 163–177 (2001).
[30] R. Fabbri, L. d. F. Costa, and O. N. d. Oliveira jr, "Video visualizations of email interaction network evolution," (2013), http://www.youtube.com/watch?v=-t5jxQ8cKxM&list=PLf_EtaMqu3jU-1j4jiIUiyMqyVSzIYeh6.
[31] R. Fabbri, L. d. F. Costa, and O. N. d. Oliveira jr, "Image gallery of email interaction networks." (2013), http://hera.ethymos.com.br:1080/redes/python/autoRede/gmane.linux.audio.devel_3000-4200-280/.
[32] R. Fabbri, L. d. F. Costa, and O. N. d. Oliveira jr, "Online gadget for making email interaction network images, gml files and measurements." (2013), http://hera.ethymos.com.br:1080/redes/python/autoRede/escolheRedes.php.
[33] Numpy version 1.6.1, "random.randint" function, was used for simulations, algorithms in https://pypi.python.org/pypi/gmane.
[34] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," Physics reports **424**, 175–308 (2006).
[35] R. Fabbri, "Versinus: a visualization method for graphs in evolution," arXiv preprint arXiv:1412.7311 (2014).
[36] R. Fabbri, "Python package to analyze the gmane database,".
[37] R. Fabbri, R. B. de Luna, R. A. P. Martins, *et al.*, "Social participation ontology: community documentation, enhancements and use examples," arXiv preprint arXiv:1501.02662 (2015).
[38] *Produto 5 da consultoria PNUD/ONU de Renato Fabbri*, https://github.com/ttm/pnud4/blob/master/latex/produto.pdf?raw=true.
[39] R. Fabbri, "Ensaio sobre o auto-aproveitamento: um relato de investidas naturais na participa\ c {c}\~ ao social," arXiv preprint arXiv:1412.6868 (2014).
[40] R. Fabbri, "What are you and i? [anthropological physics fundamentals]," academia.edu (2015), https://www.academia.edu/10356773/What_are_you_and_I_anthropological_physics_fundamentals_.
[41] T. W. Adorno, E. Frenkel-Brunswik, D. J. Levinson, and R. N. Sanford, "The authoritarian personality." (1950).
[42] R. Fabbri, L. d. F. Costa, and O. N. d. Oliveira jr, "Scripts used for obtaining results used in this article ." (2013), sourceforge.net/p/labmacambira/fimDoMundo/ci/master/tree/python/toolkitGMANE/.
[43] N. Developers, "Networkx," (2010).
[44] J. Candia, M. C. González, P. Wang, T. Schoenharl, G. Madey, and A.-L. Barabási, "Uncovering individual and collective human dynamics from mobile phone records," Journal of Physics A: Mathematical and Theoretical **41**, 224015 (2008).
[45] C. G. Jung, H. Baynes, and R. Hull, *Psychological types*, Vol. 4 (Routledge London, UK, 1991).
[46] N. L. Quenk, *Essentials of Myers-Briggs type indicator assessment*, Vol. 66 (Wiley. com, 2009).
[47] S. Freud, "Libidinal types." The Psychoanalytic Quarterly (1932).
[48] H. J. Eysenck, "Types of personality: a factorial study of seven hundred neurotics," The British Journal of Psychiatry **90**, 851–861 (1944).
[49] R. Fabbri, "Complex networks and natural language processing collection and diffusion of information and goods." (2014), wiki.nosdigitais.teia.org.br/ARS.
[50] T. Jia and A.-L. Barabási, "Control capacity and a random sampling method in exploring controllability of complex networks," Scientific reports **3** (2013).
[51] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Control centrality and hierarchical structure in complex networks," Plos one **7**, e44459 (2012).
[52] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Controllability of complex networks," Nature **473**, 167–173 (2011).

[53]J. P. Bagrow, D. Wang, and A.-L. Barabasi, "Collective response of human populations to large-scale emergencies," PloS one **6**, e17680 (2011).
[54]G. Ghoshal, N. Blumm, Z. Forro, M. Schich, G. Bianconi, J.-P. Bouchaud, and A.-L. Barabasi, "Dynamics of ranking processes in complex systems," (2012).
[55]S.-H. Yook, H. Jeong, A.-L. Barabási, and Y. Tu, "Weighted evolving networks," Physical Review Letters **86**, 5835 (2001).
[56]D. Wang, Z. Wen, H. Tong, C.-Y. Lin, C. Song, and A.-L. Barabási, "Information spreading in context," in *Proceedings of the 20th international conference on World wide web* (ACM, 2011) pp. 735–744.
[57]N. Blumm, G. Ghoshal, Z. Forró, M. Schich, G. Bianconi, J.-P. Bouchaud, and A.-L. Barabási, "Dynamics of ranking processes in complex systems," Physical Review Letters **109**, 128701 (2012).
[58]M. E. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," Physical Review E **64**, 026118 (2001).
[59]M. E. Newman, "Random graphs with clustering," Physical review letters **103**, 058701 (2009).
[60]A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," SIAM review **51**, 661–703 (2009).
[61]M. E. Newman, "Assortative mixing in networks," Physical review letters **89**, 208701 (2002).
[62]M. Newman, "Communities, modules and large-scale structure in networks," Nature Physics **8**, 25–31 (2011).
[63]A. Clauset, C. Moore, and M. E. Newman, "Hierarchical structure and the prediction of missing links in networks," Nature **453**, 98–101 (2008).
[64]M. Newman, "Complex systems: A survey," arXiv preprint arXiv:1112.1440 (2011).
[65]B. Ball and M. E. Newman, "Friendship networks and social status," arXiv preprint arXiv:1205.6822 (2012).
[66]G. Deleuze, *Difference and Repetition* (Continuum, 1968).
[67]F. de Saussure, *Course in General Linguistics* (Books LLC, 1916).
[68]A. P. S. U. Pillai, *Probability, Random Variables and Stochastic Processes* (McGraw Hill Higher Education, 2002).
[69]R. A. J. D. W. Wichern, *Applied Multivariate Statistical Analysis* (Prentice Hall, 2007).
[70]C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing* (Prentice Hall, 1992).
[71]R. O. D. P. E. Hart D. G. Stork, *Pattern Classification* (Wiley-Interscience, 2000).
[72]L. da F. Costa R. M. C. Jr., *Shape Analysis and Classification: Theory and Practice (Image Processing Series)* (CRC Press, 2000).
[73]D. Papineau, *Philosophy* (Oxford University Press, 2009).
[74]B. Russel, *A History of Western Philosophy* (Simon and Schuster Touchstone, 1967).
[75]F. G. G. Deleuze, *What Is Philosophy?* (Simon and Schuster Touchstone, 1991).

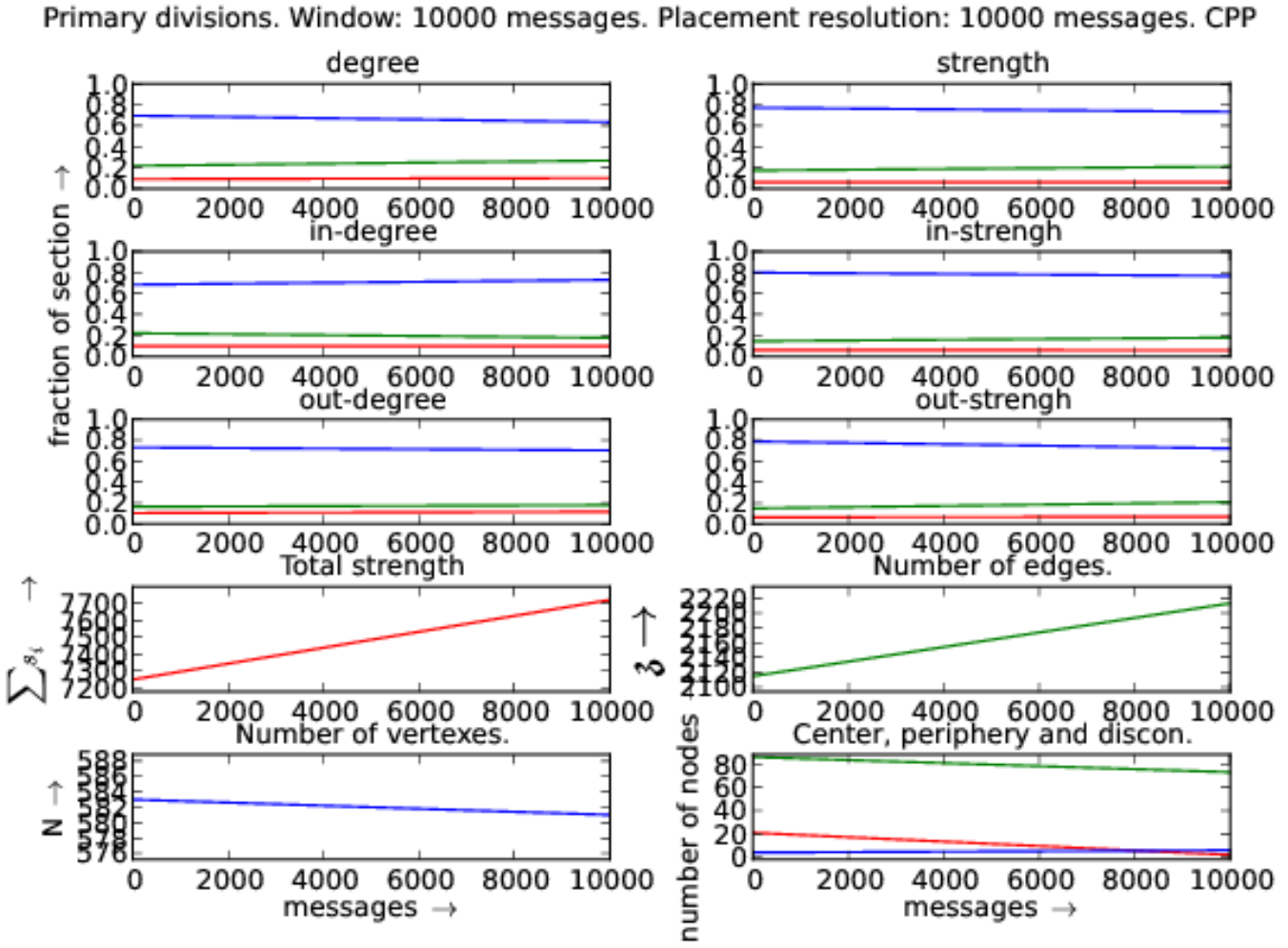

FIG. 5. Distribution of vertices with respect to each centrality measure: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex is equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertices.

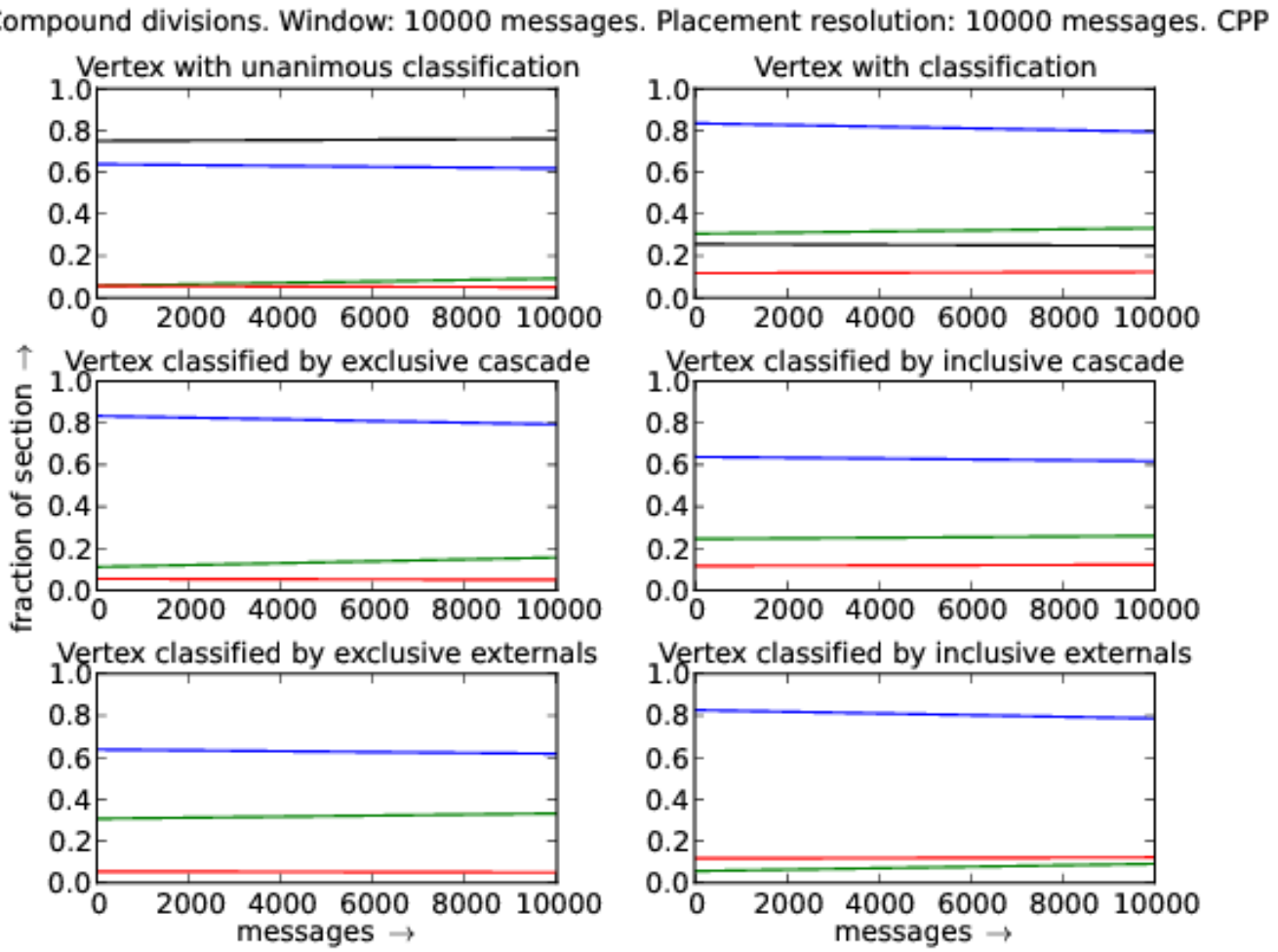

FIG. 6. Distribution of vertex with respect to compound criteria. Red, green and blue designate hubs, intermediary and border (peripheral) vertex fractions. The first two plots exhibit classifications that are not functions. Thus, in the first plot, the fraction of vertices with unique classification is plotted in black. On the second plot, black represents the fraction of vertices that has more than one class: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria is described in Section III C.

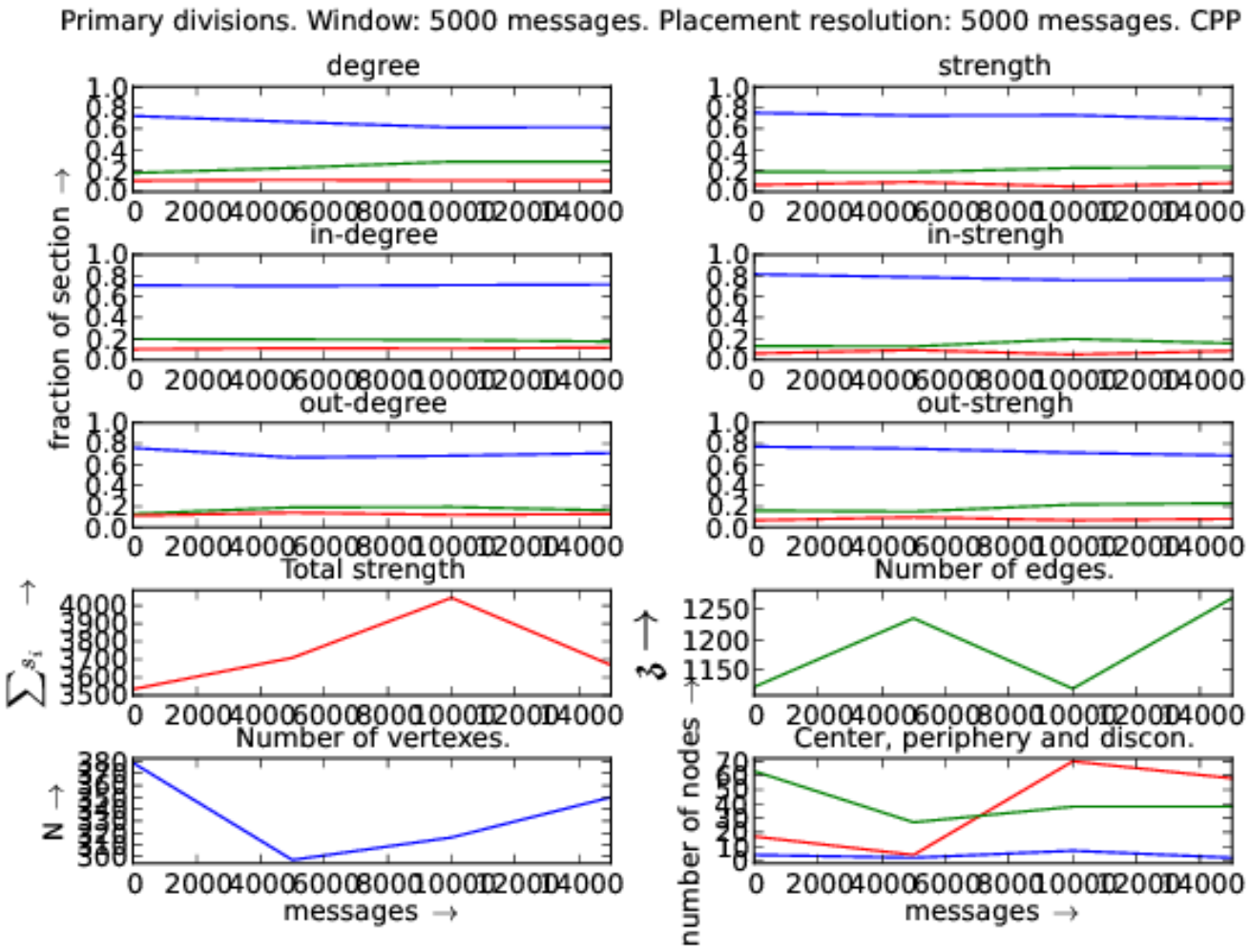

FIG. 7. Distribution of vertices with respect to each centrality measure: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex is equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertices.

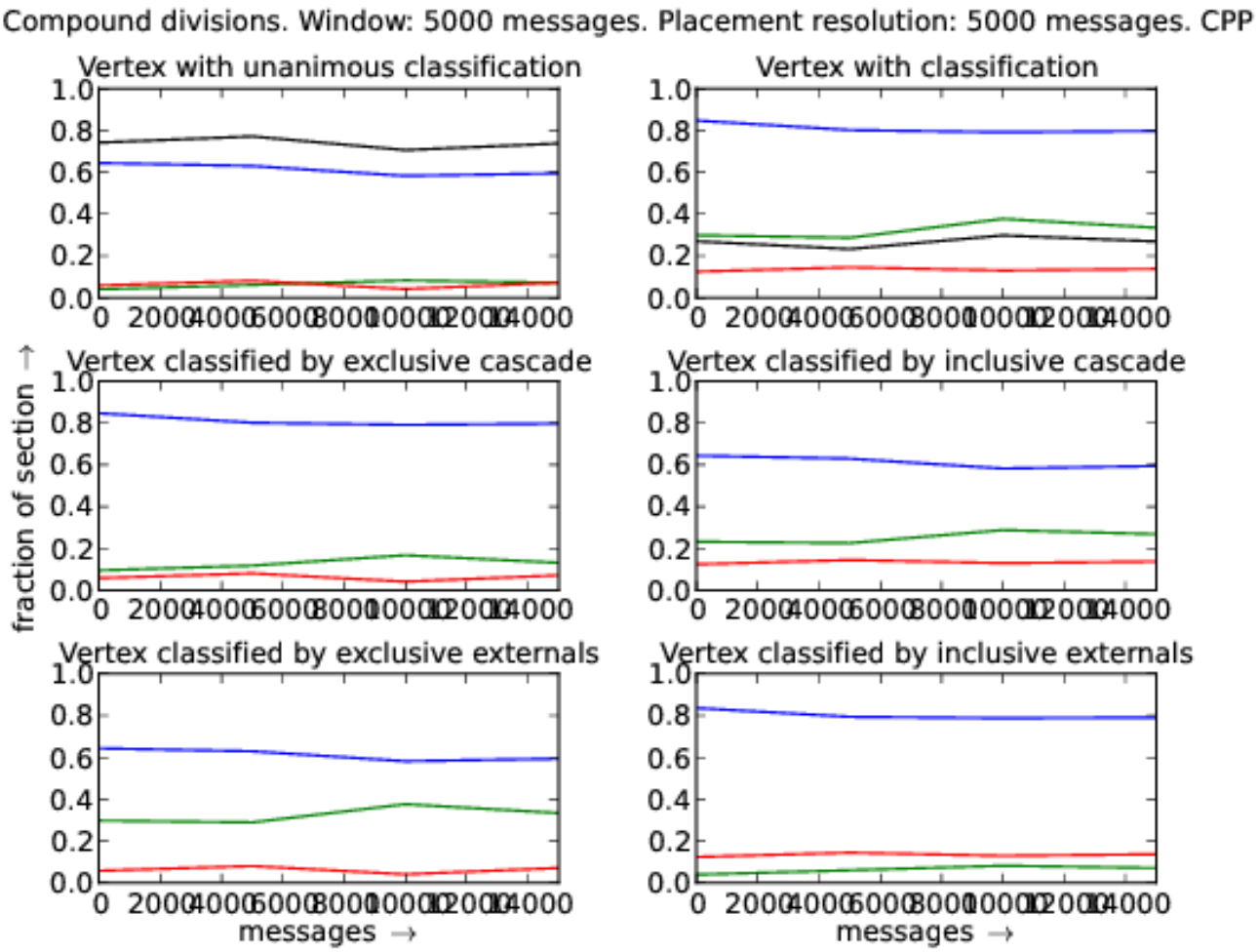

FIG. 8. Distribution of vertex with respect to compound criteria. Red, green and blue designate hubs, intermediary and border (peripheral) vertex fractions. The first two plots exhibit classifications that are not functions. Thus, in the first plot, the fraction of vertices with unique classification in plotted in black. On the second plot, black represents the fraction of vertices that has more than one class: $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$. Compound criteria is described in Section III C.

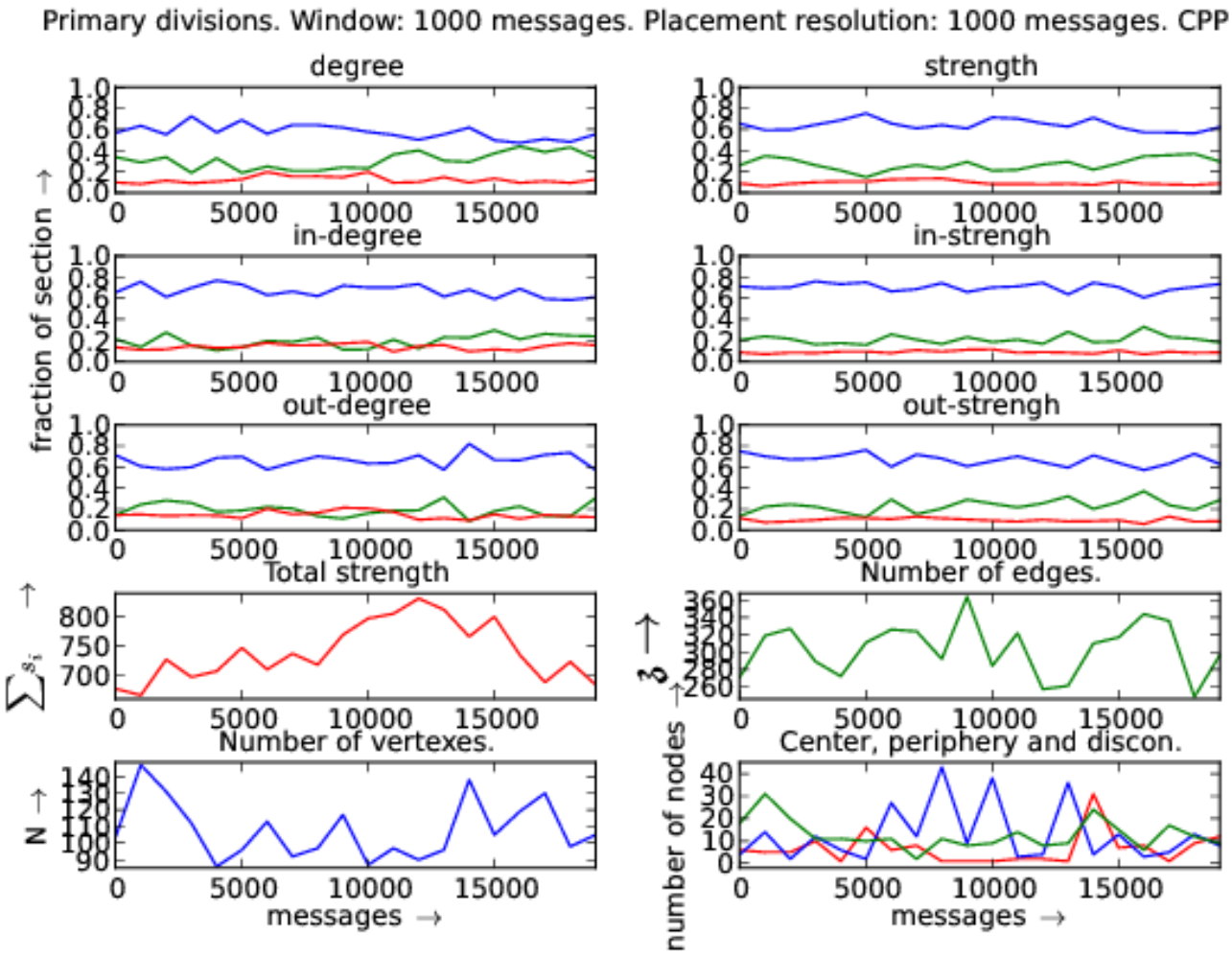

FIG. 9. Distribution of vertices with respect to each centrality measure: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex is equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertices.

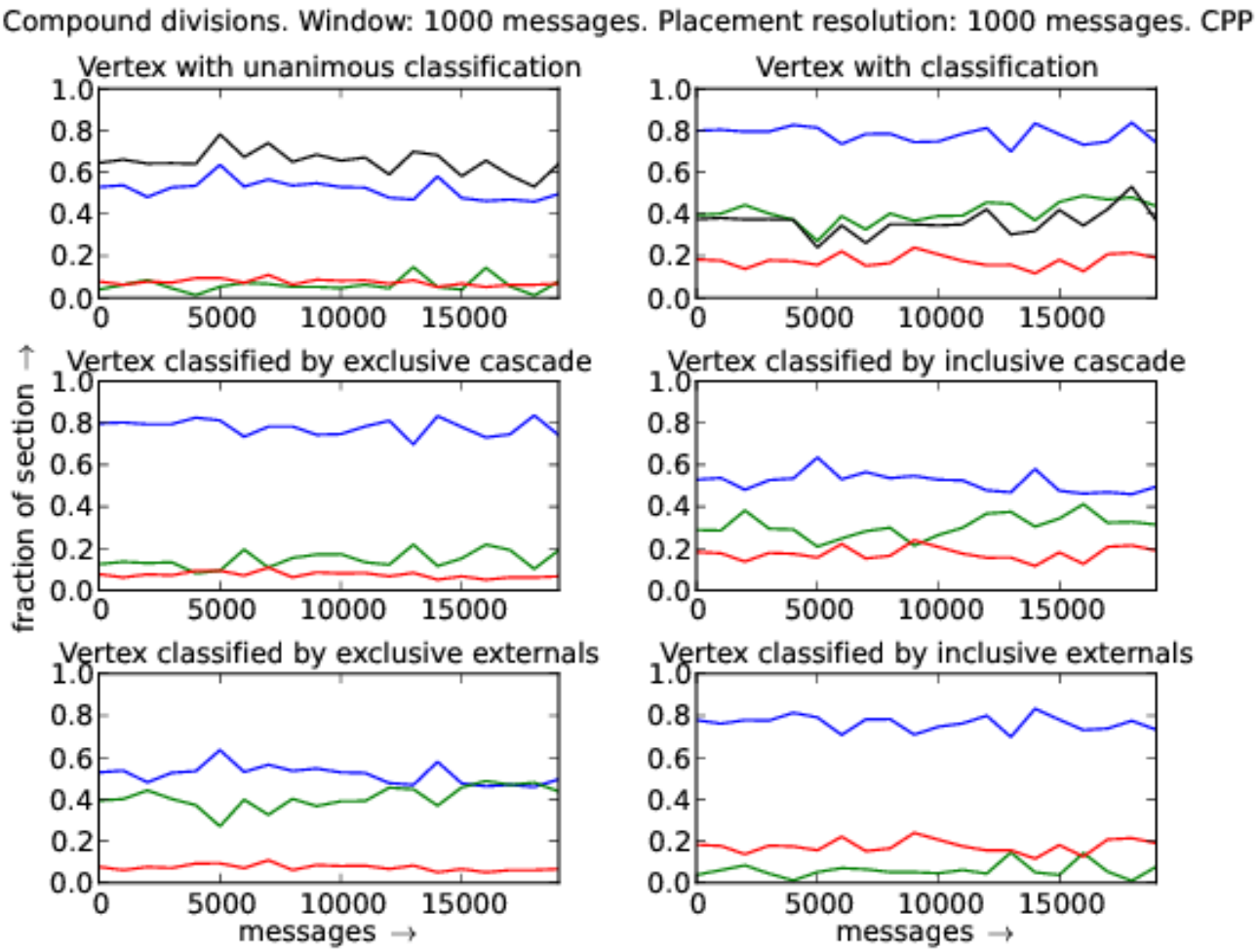

FIG. 10. Distribution of vertex with respect to compound criteria. Red, green and blue designate hubs, intermediary and border (peripheral) vertex fractions. The first two plots exhibit classifications that are not functions. Thus, in the first plot, the fraction of vertices with unique classification in plotted in black. On the second plot, black represents the fraction of vertices that has more than one class: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria is described in Section III C.

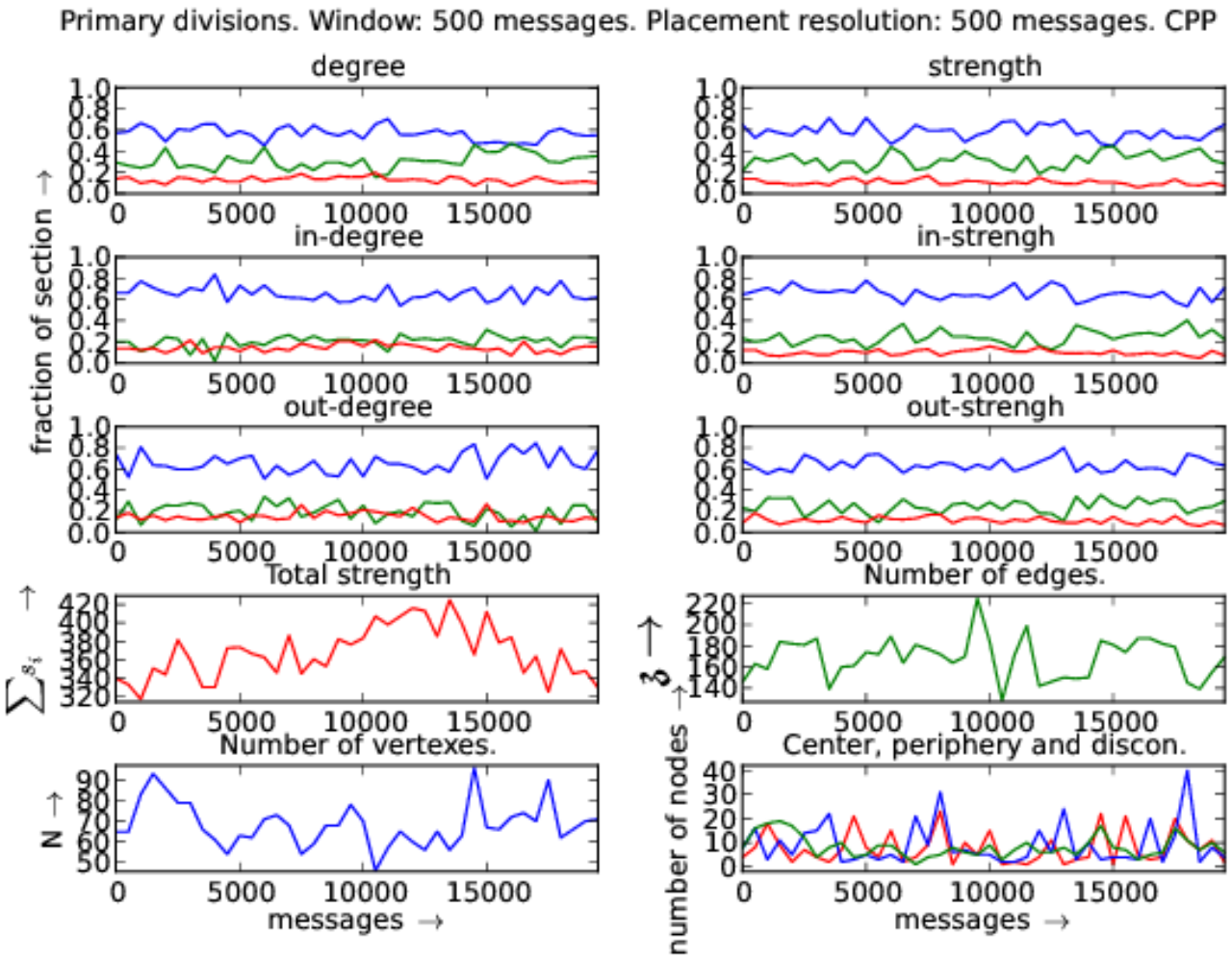

FIG. 11. Distribution of vertices with respect to each centrality measure: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex is equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertices.

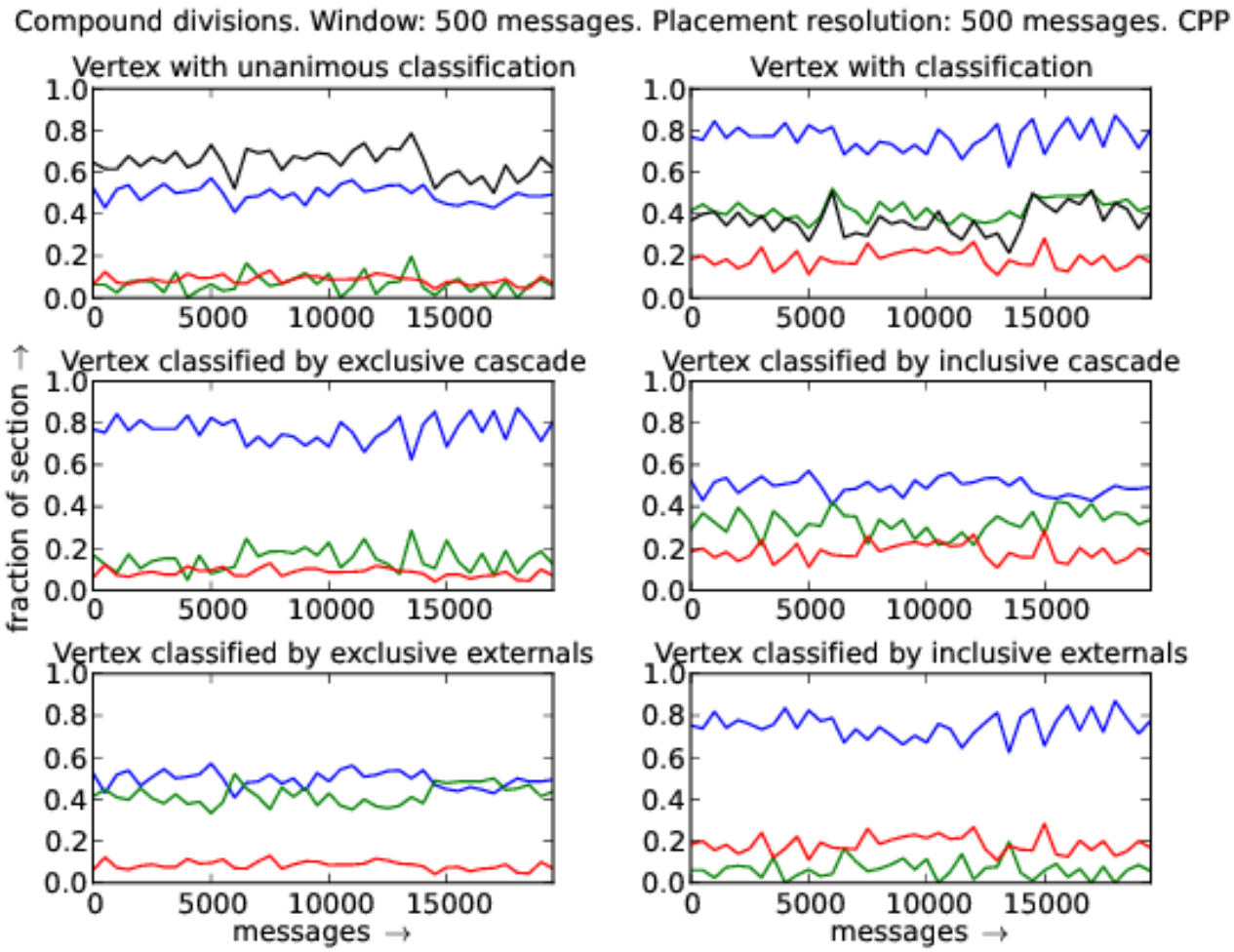

FIG. 12. Distribution of vertex with respect to compound criteria. Red, green and blue designate hubs, intermediary and border (peripheral) vertex fractions. The first two plots exhibit classifications that are not functions. Thus, in the first plot, the fraction of vertices with unique classification in plotted in black. On the second plot, black represents the fraction of vertices that has more than one class: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria is described in Section III C.

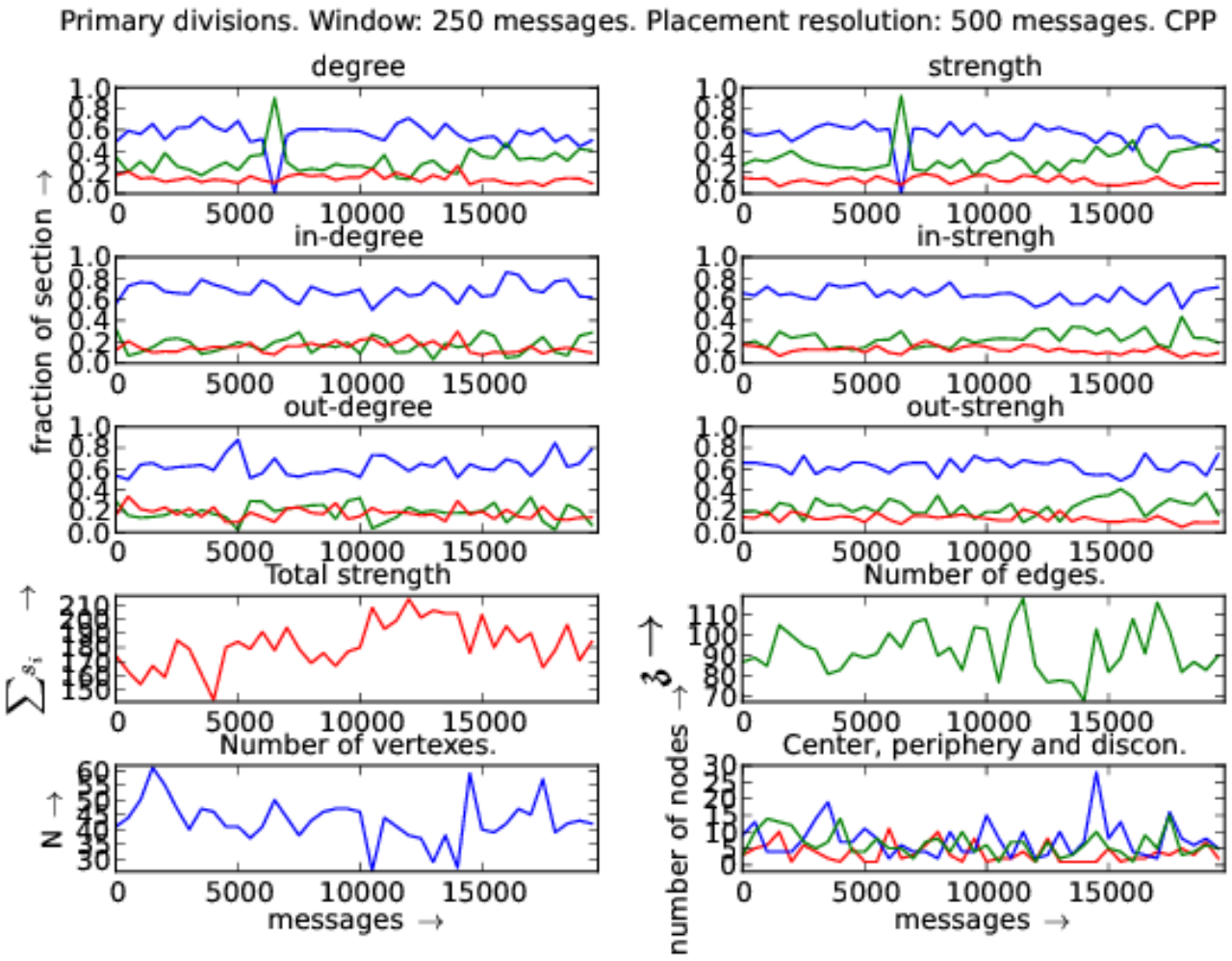

FIG. 13. Distribution of vertices with respect to each centrality measure: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex is equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertices.

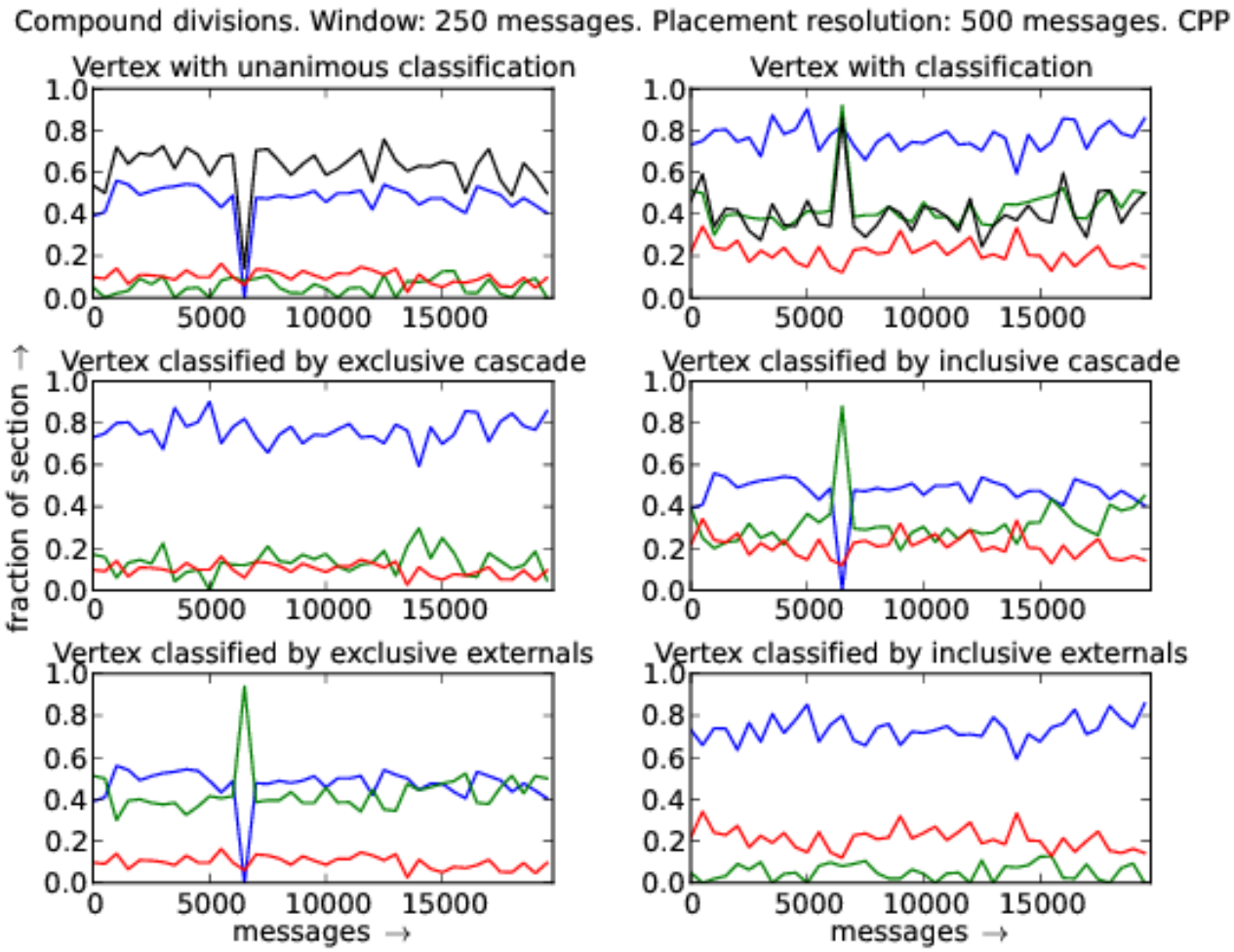

FIG. 14. Distribution of vertex with respect to compound criteria. Red, green and blue designate hubs, intermediary and border (peripheral) vertex fractions. The first two plots exhibit classifications that are not functions. Thus, in the first plot, the fraction of vertices with unique classification in plotted in black. On the second plot, black represents the fraction of vertices that has more than one class: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria is described in Section III C.

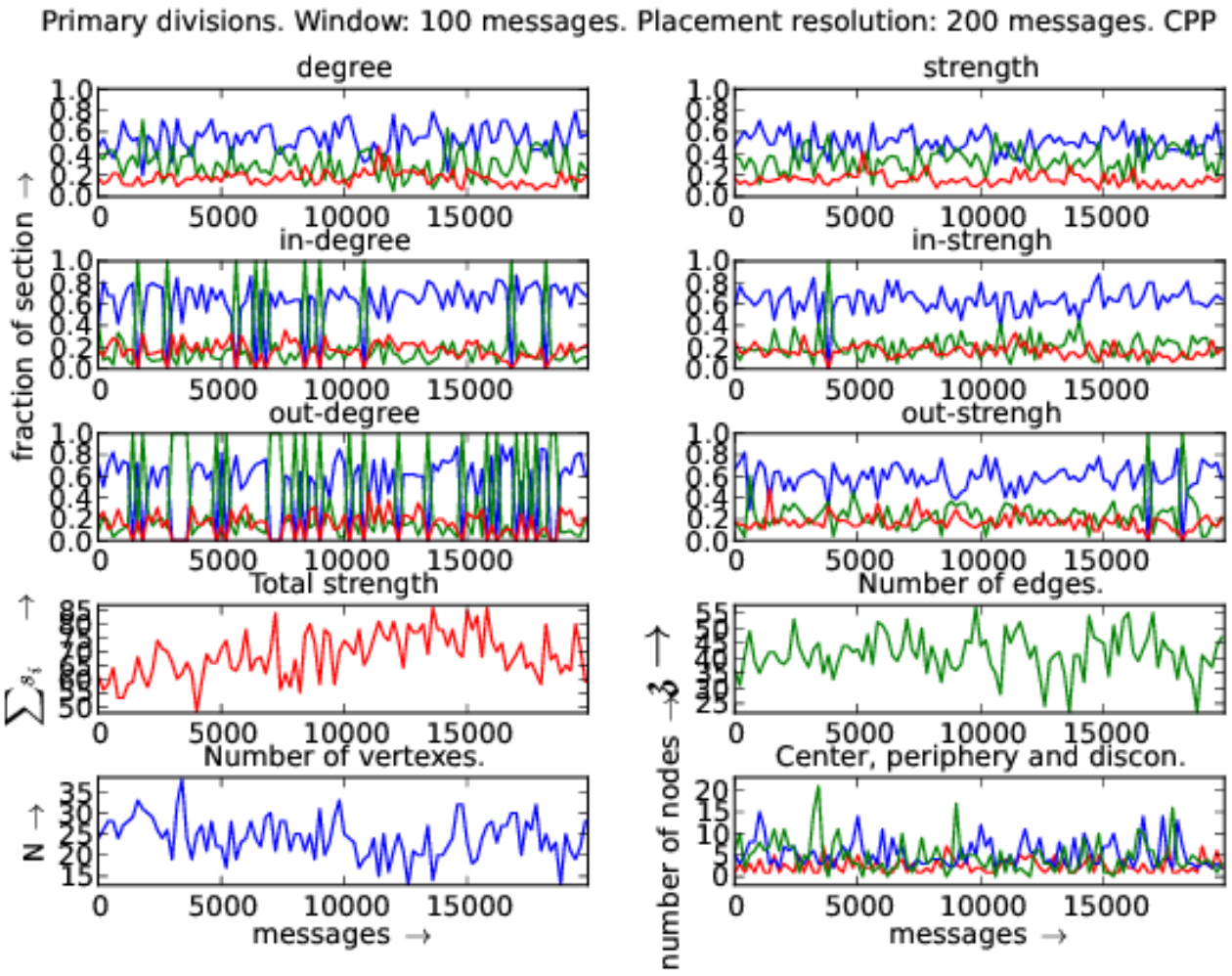

FIG. 15. Distribution of vertices with respect to each centrality measure: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex is equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertices.

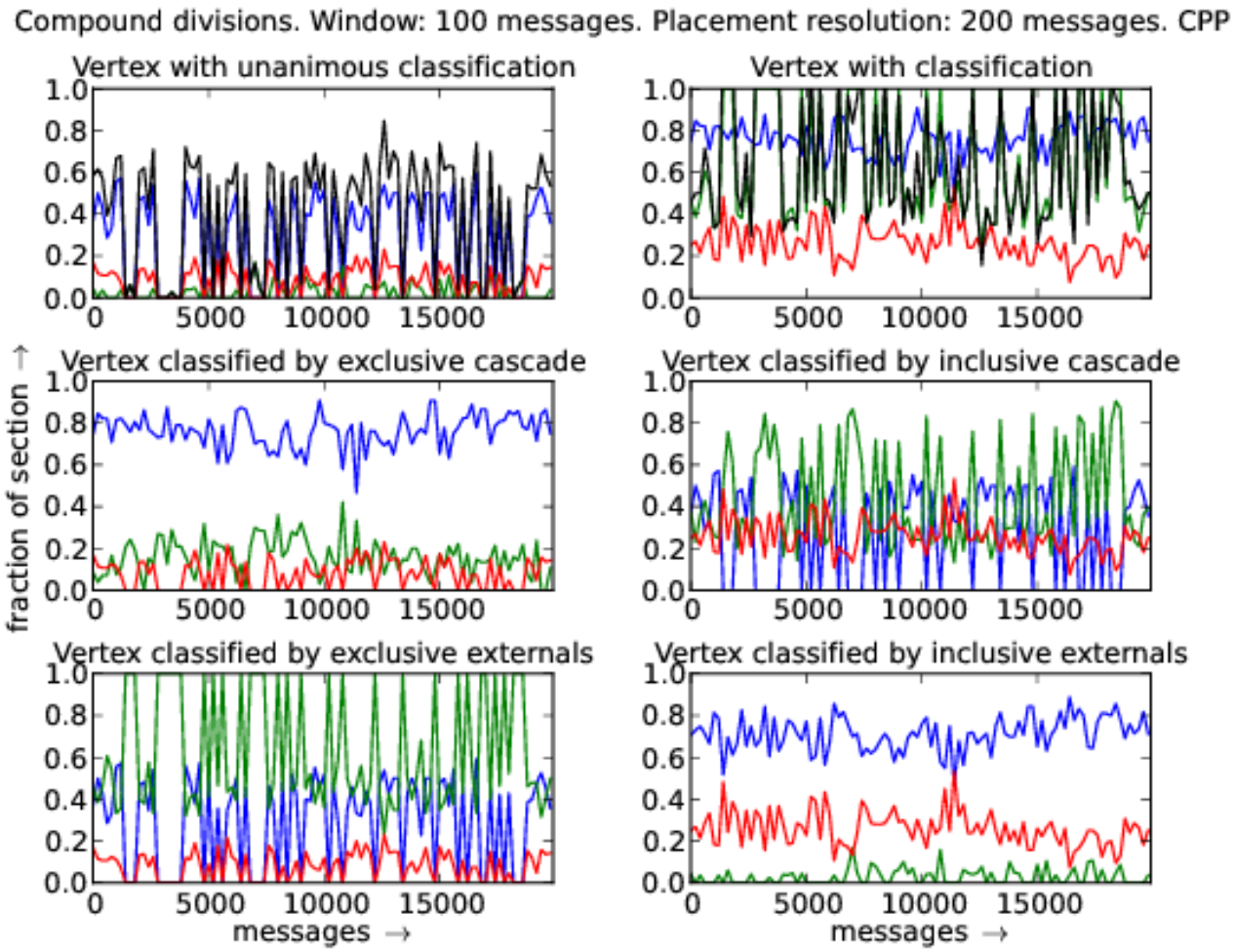

FIG. 16. Distribution of vertex with respect to compound criteria. Red, green and blue designate hubs, intermediary and border (peripheral) vertex fractions. The first two plots exhibit classifications that are not functions. Thus, in the first plot, the fraction of vertices with unique classification in plotted in black. On the second plot, black represents the fraction of vertices that has more than one class: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria is described in Section III C.

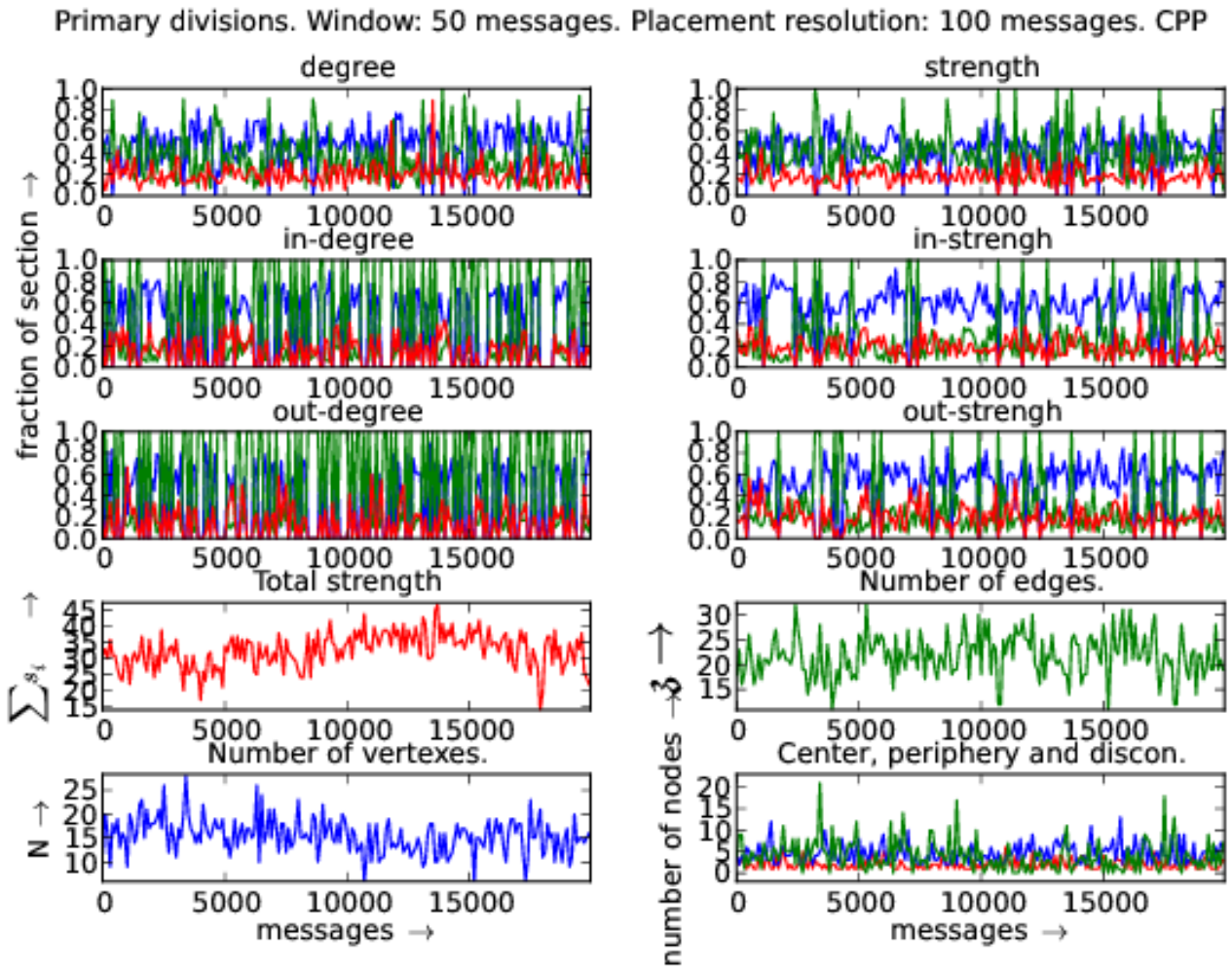

FIG. 17. Distribution of vertices with respect to each centrality measure: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex is equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertices.

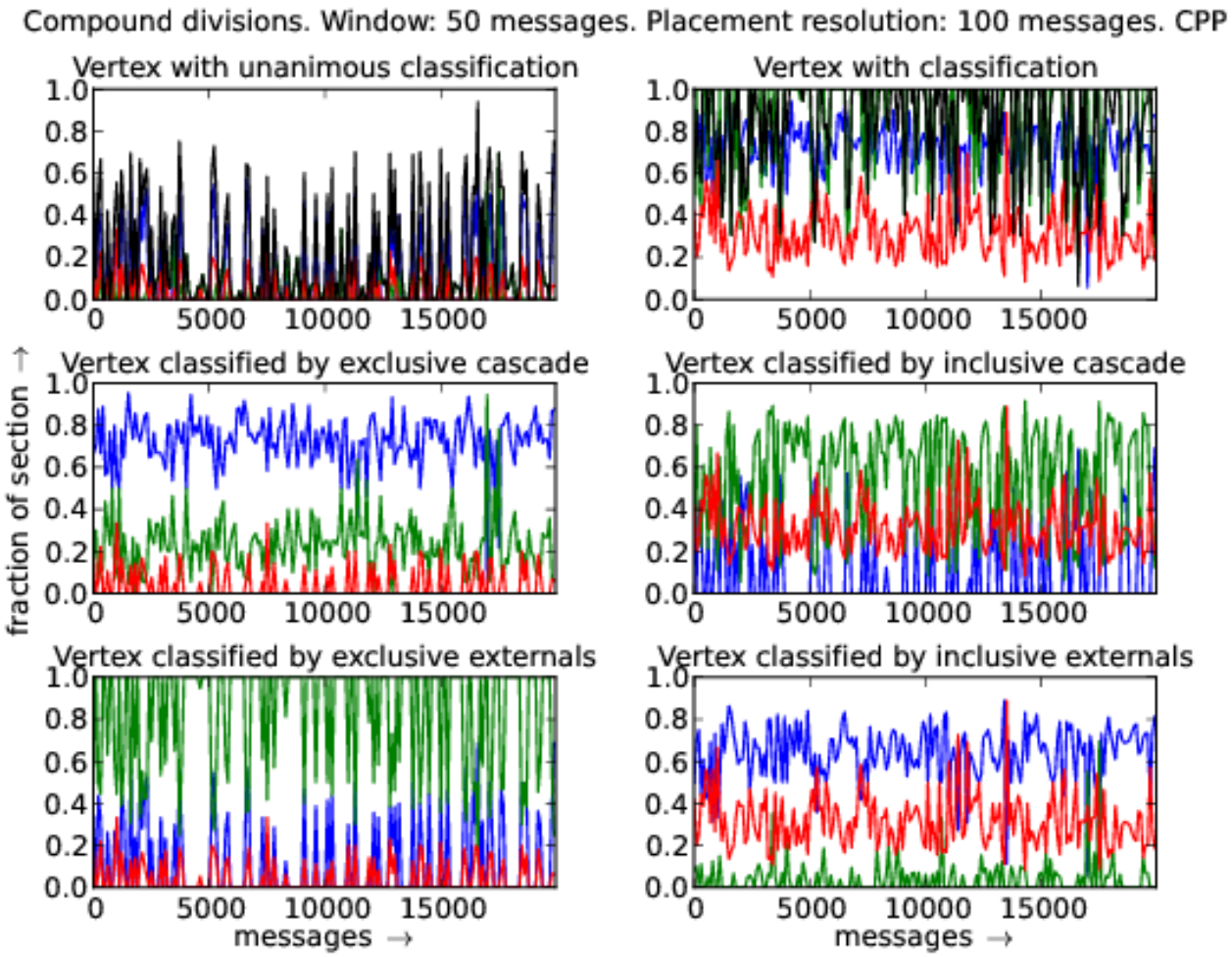

FIG. 18. Distribution of vertex with respect to compound criteria. Red, green and blue designate hubs, intermediary and border (peripheral) vertex fractions. The first two plots exhibit classifications that are not functions. Thus, in the first plot, the fraction of vertices with unique classification in plotted in black. On the second plot, black represents the fraction of vertices that has more than one class: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria is described in Section III C.

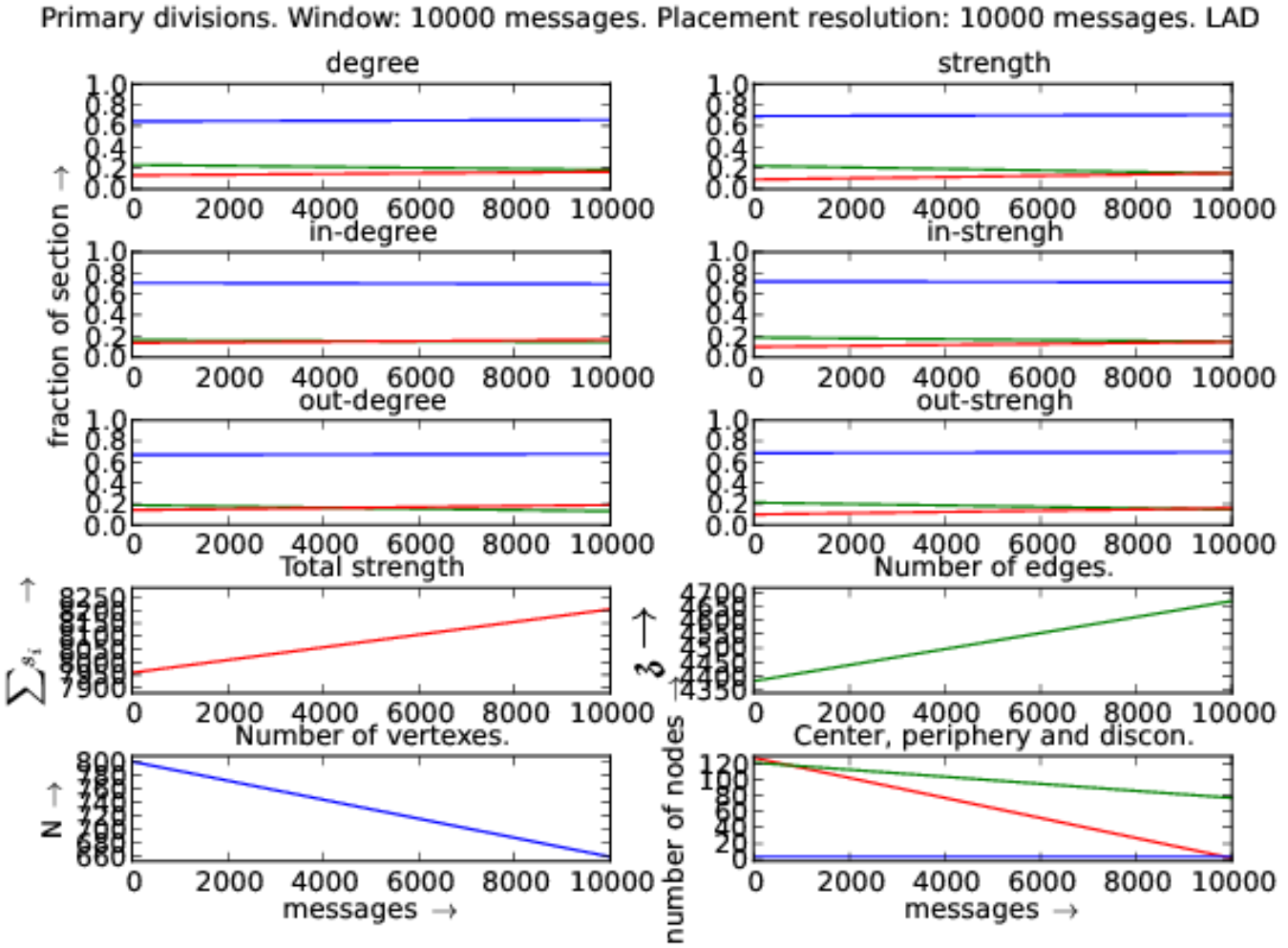

FIG. 19. Distribution of vertices with respect to each centrality measure: in and out degrees and strengths. Linux Audio Users (LAD) official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex is equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertices.

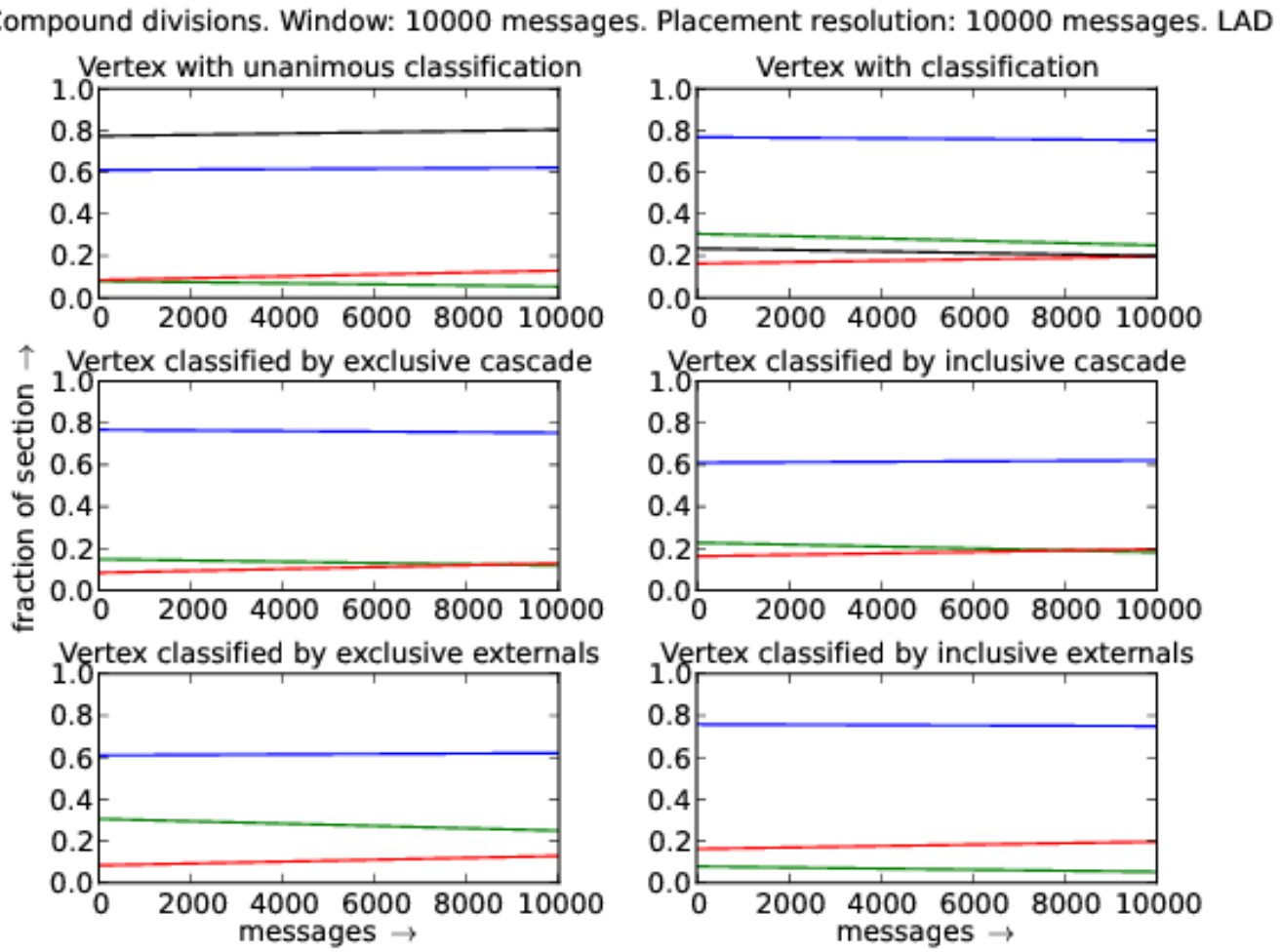

FIG. 20. Distribution of vertex with respect to compound criteria. Red, green and blue designate hubs, intermediary and border (peripheral) vertex fractions. The first two plots exhibit classifications that are not functions. Thus, in the first plot, the fraction of vertices with unique classification in plotted in black. On the second plot, black represents the fraction of vertices that has more than one class: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria is described in Section III C.

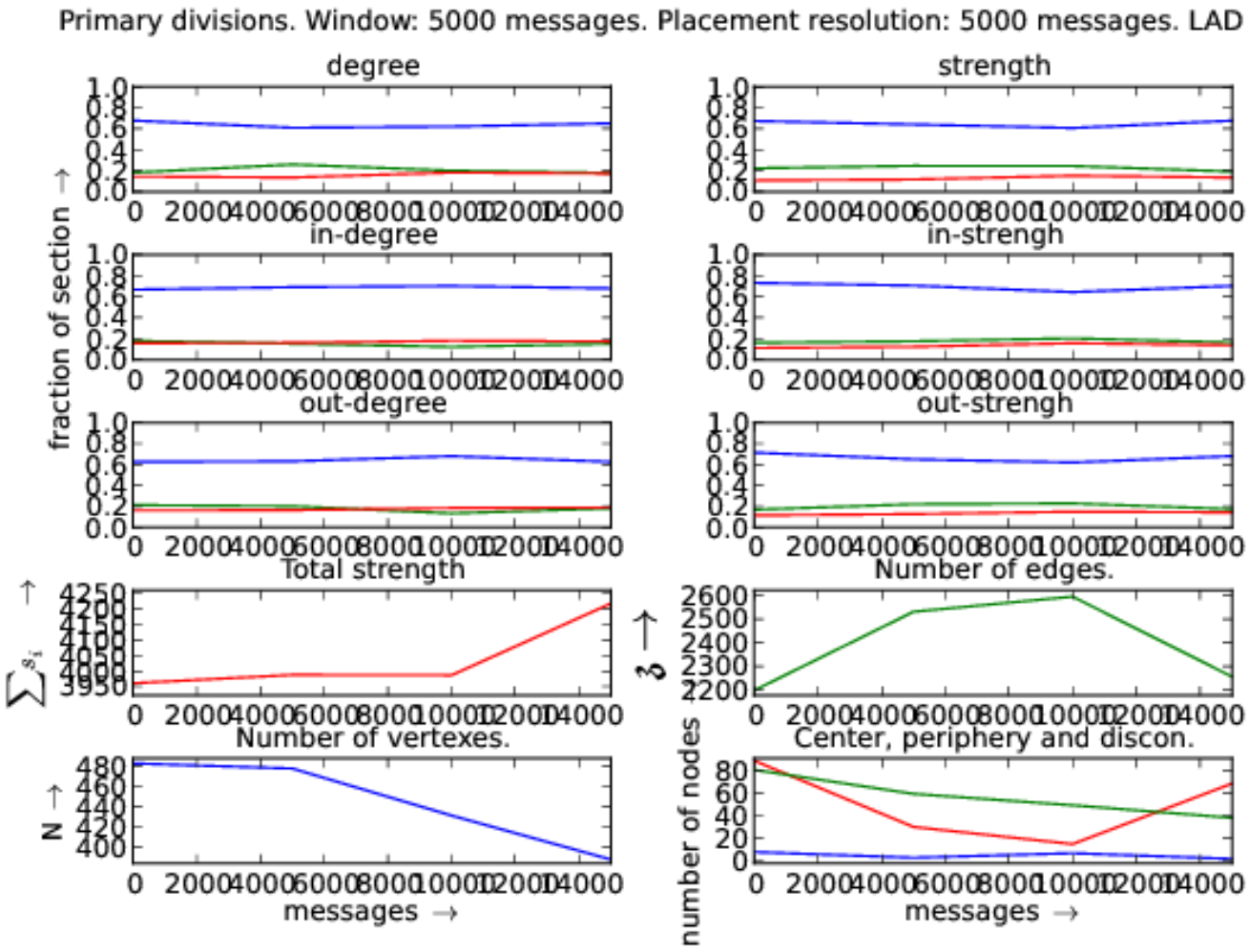

FIG. 21. Distribution of vertices with respect to each centrality measure: in and out degrees and strengths. Linux Audio Users (LAD) official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex is equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertices.

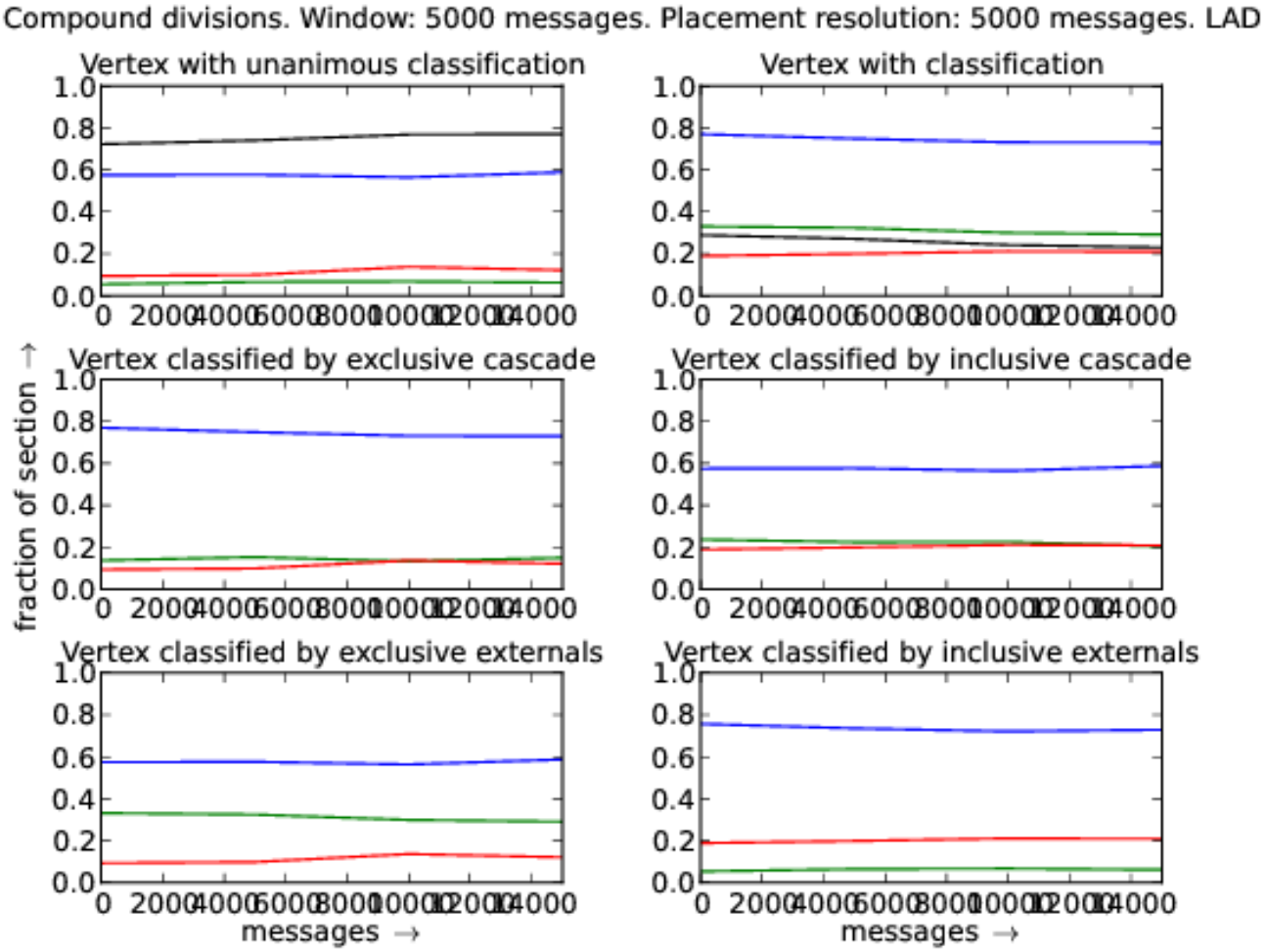

FIG. 22. Distribution of vertex with respect to compound criteria. Red, green and blue designate hubs, intermediary and border (peripheral) vertex fractions. The first two plots exhibit classifications that are not functions. Thus, in the first plot, the fraction of vertices with unique classification in plotted in black. On the second plot, black represents the fraction of vertices that has more than one class: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria is described in Section III C.

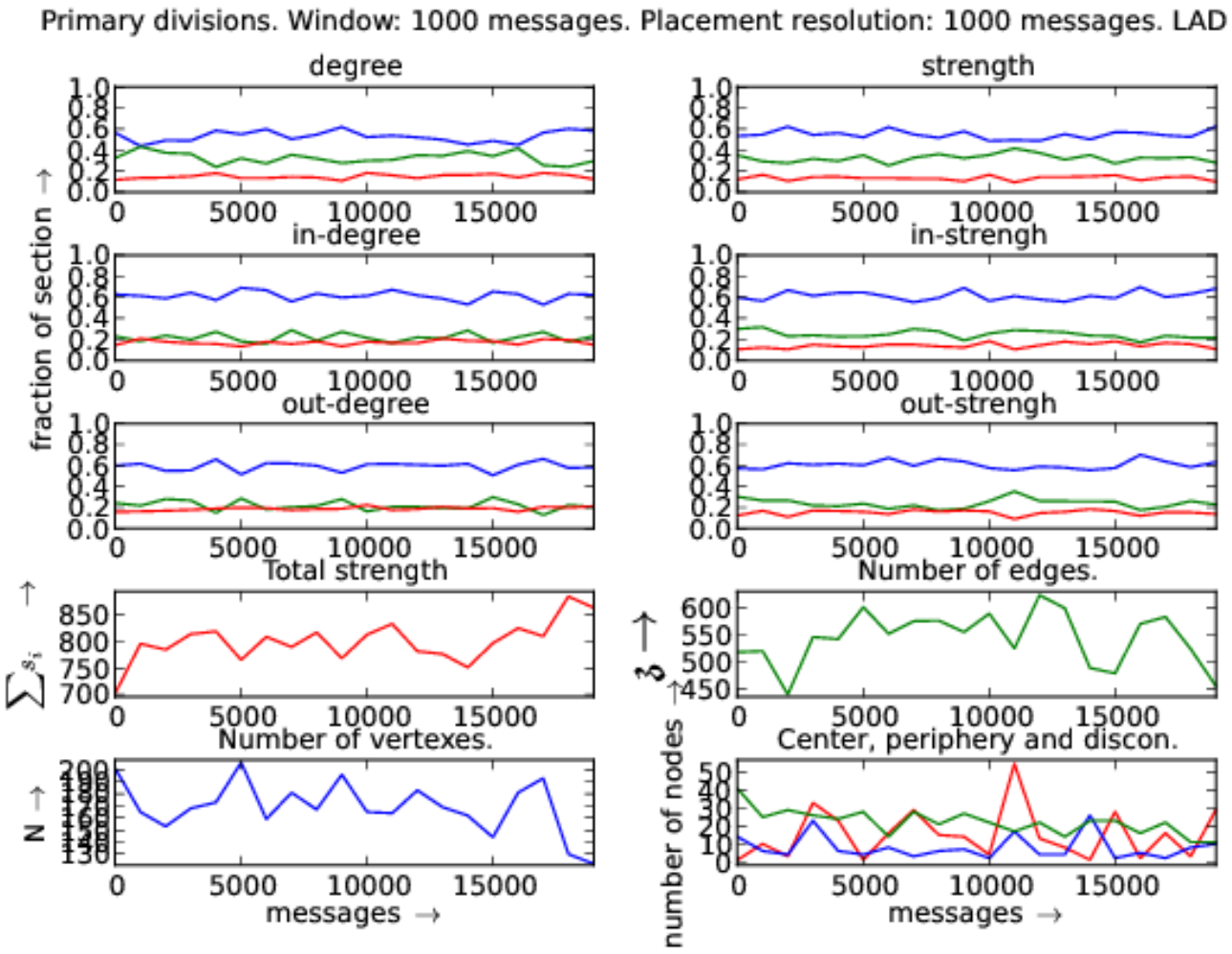

FIG. 23. Distribution of vertices with respect to each centrality measure: in and out degrees and strengths. Linux Audio Users (LAD) official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex is equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertices.

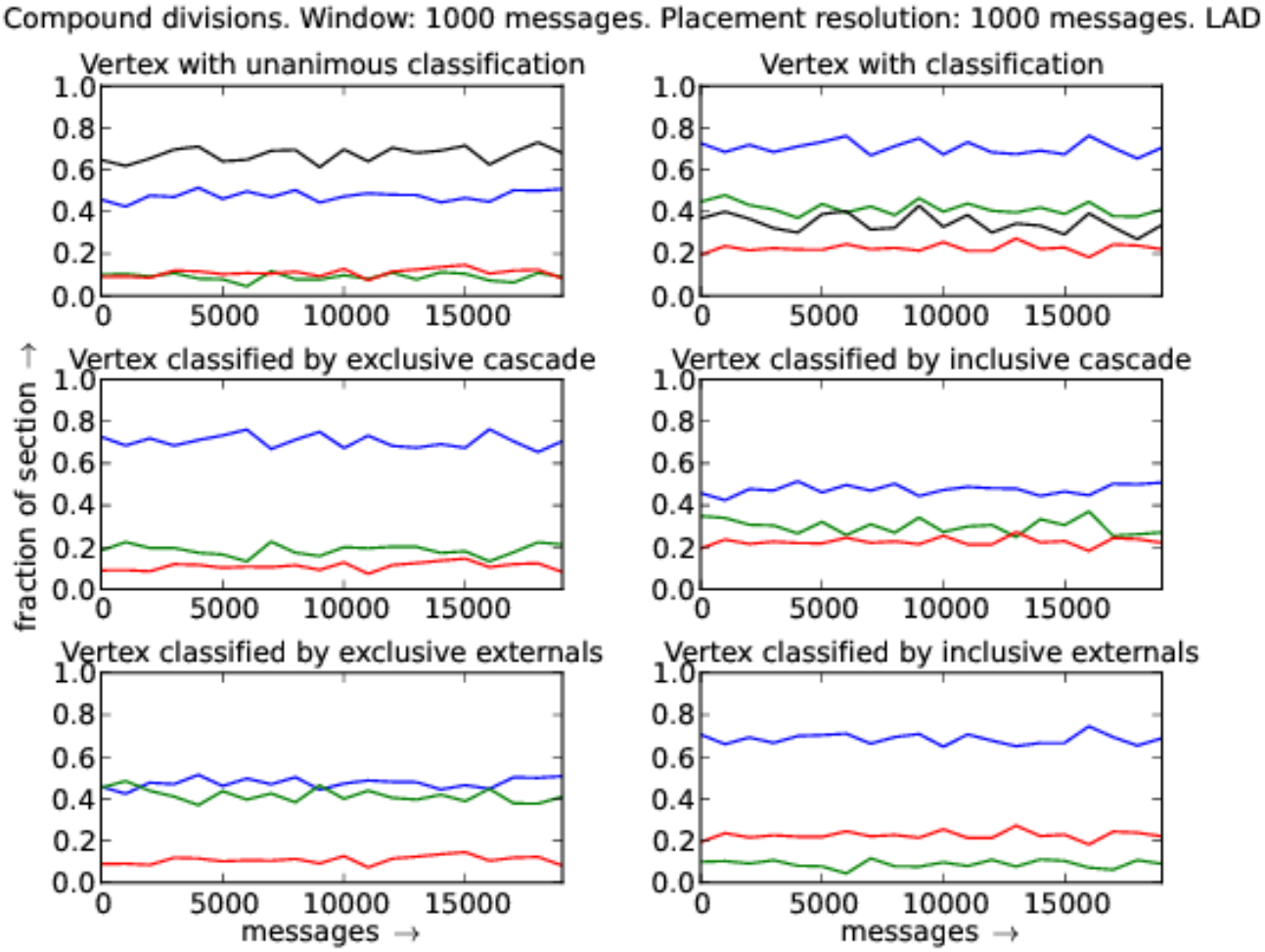

FIG. 24. Distribution of vertex with respect to compound criteria. Red, green and blue designate hubs, intermediary and border (peripheral) vertex fractions. The first two plots exhibit classifications that are not functions. Thus, in the first plot, the fraction of vertices with unique classification in plotted in black. On the second plot, black represents the fraction of vertices that has more than one class: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria is described in Section III C.

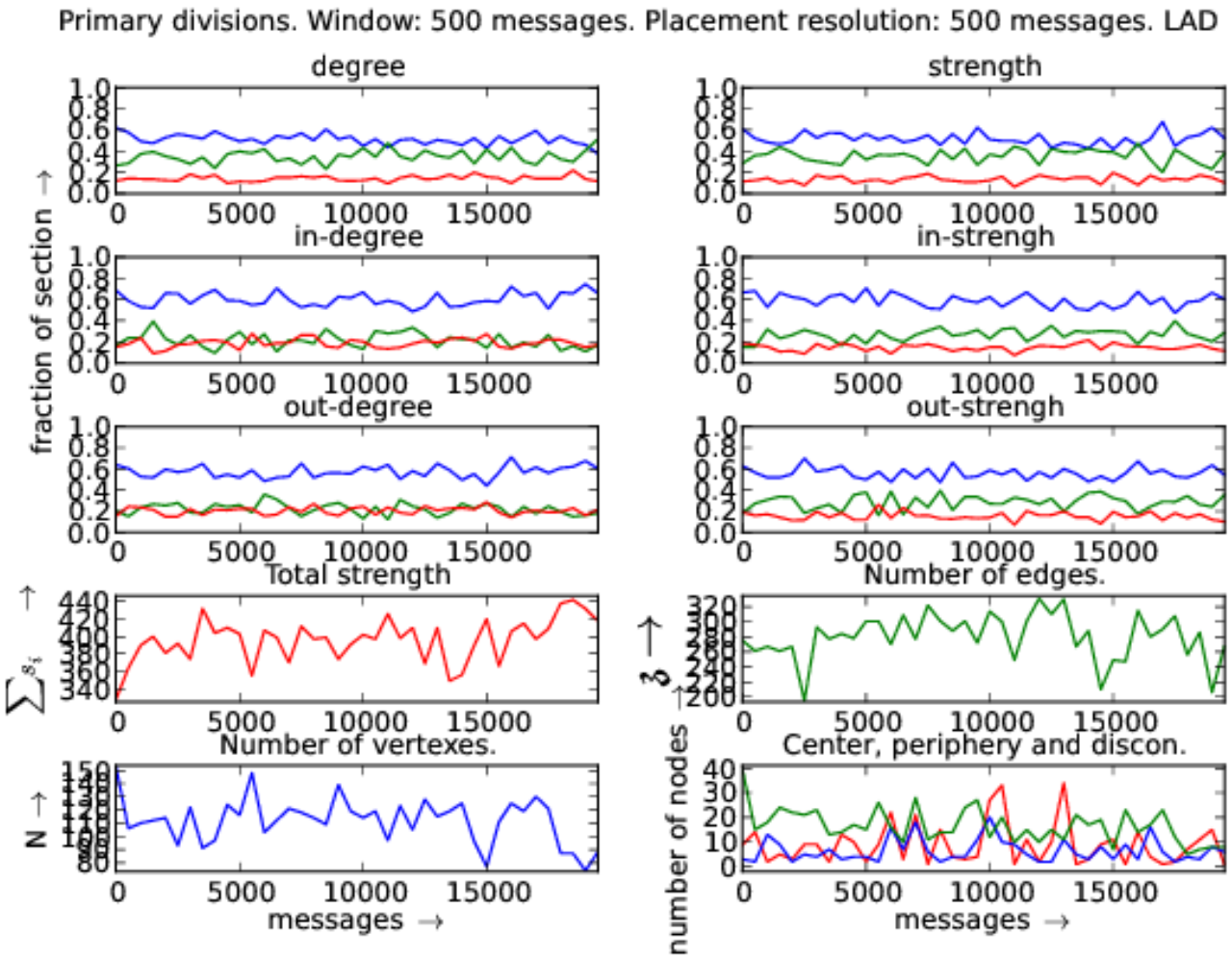

FIG. 25. Distribution of vertices with respect to each centrality measure: in and out degrees and strengths. Linux Audio Users (LAD) official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex is equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertices.

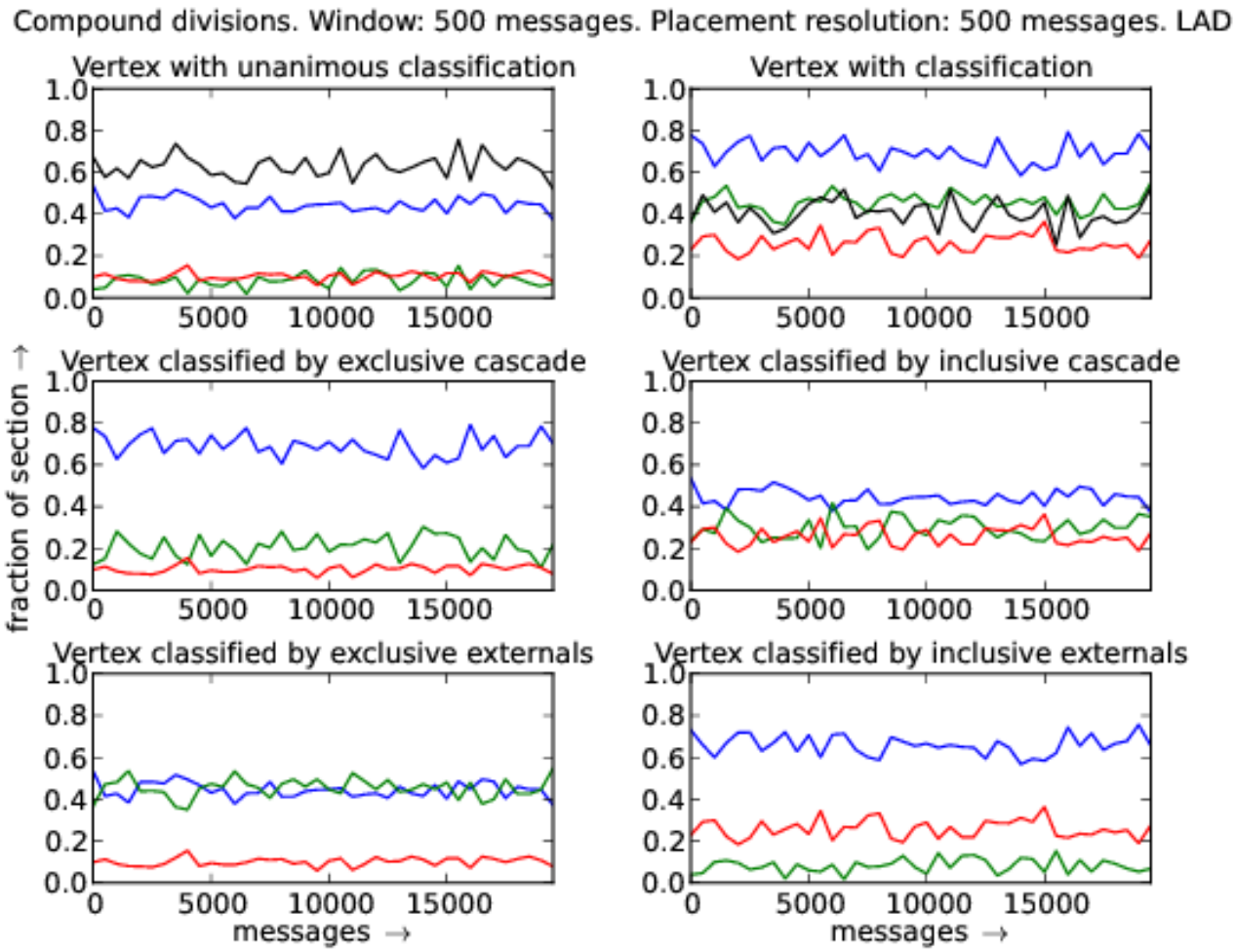

FIG. 26. Distribution of vertex with respect to compound criteria. Red, green and blue designate hubs, intermediary and border (peripheral) vertex fractions. The first two plots exhibit classifications that are not functions. Thus, in the first plot, the fraction of vertices with unique classification in plotted in black. On the second plot, black represents the fraction of vertices that has more than one class: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria is described in Section III C.

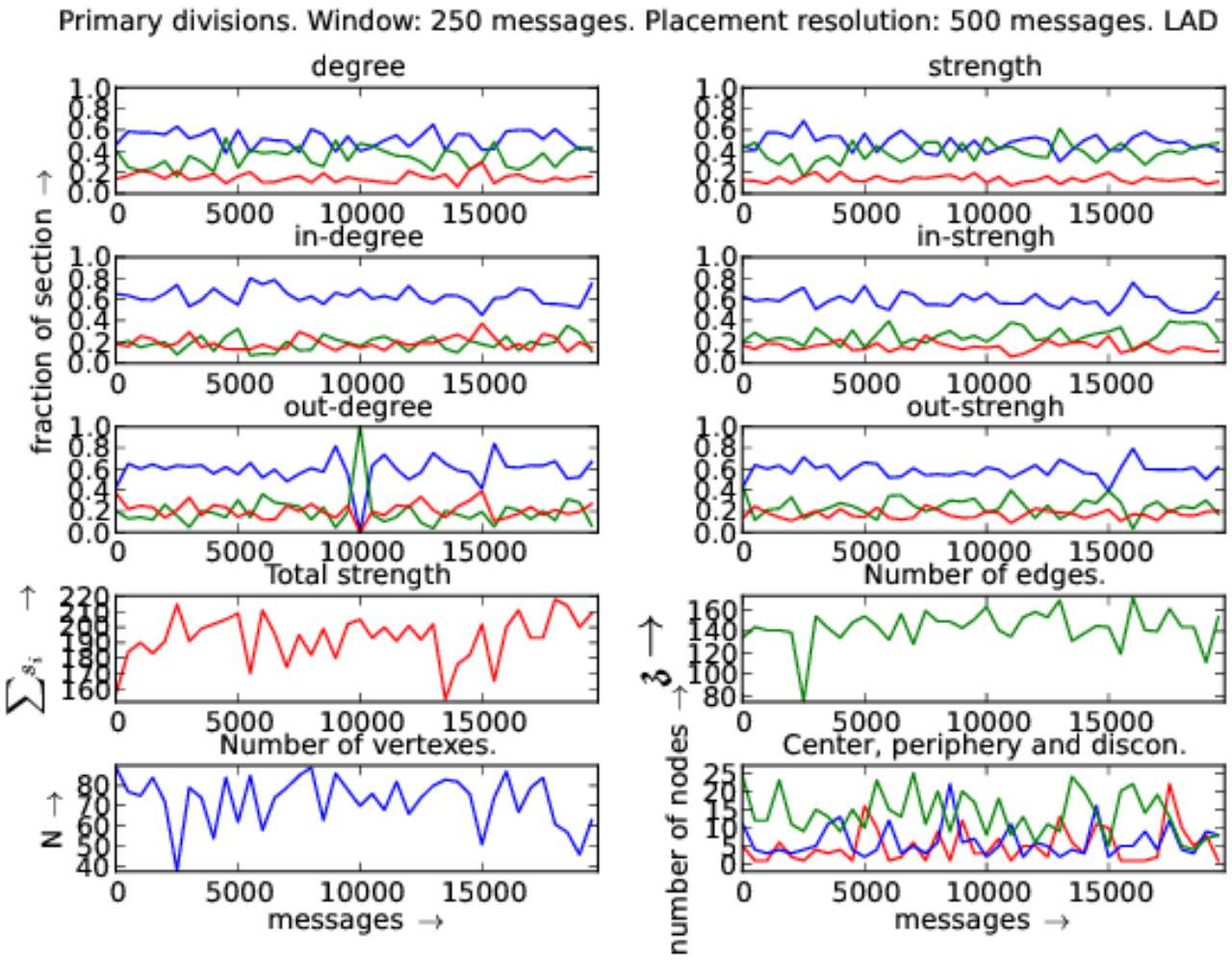

FIG. 27. Distribution of vertices with respect to each centrality measure: in and out degrees and strengths. Linux Audio Users (LAD) official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex is equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertices.

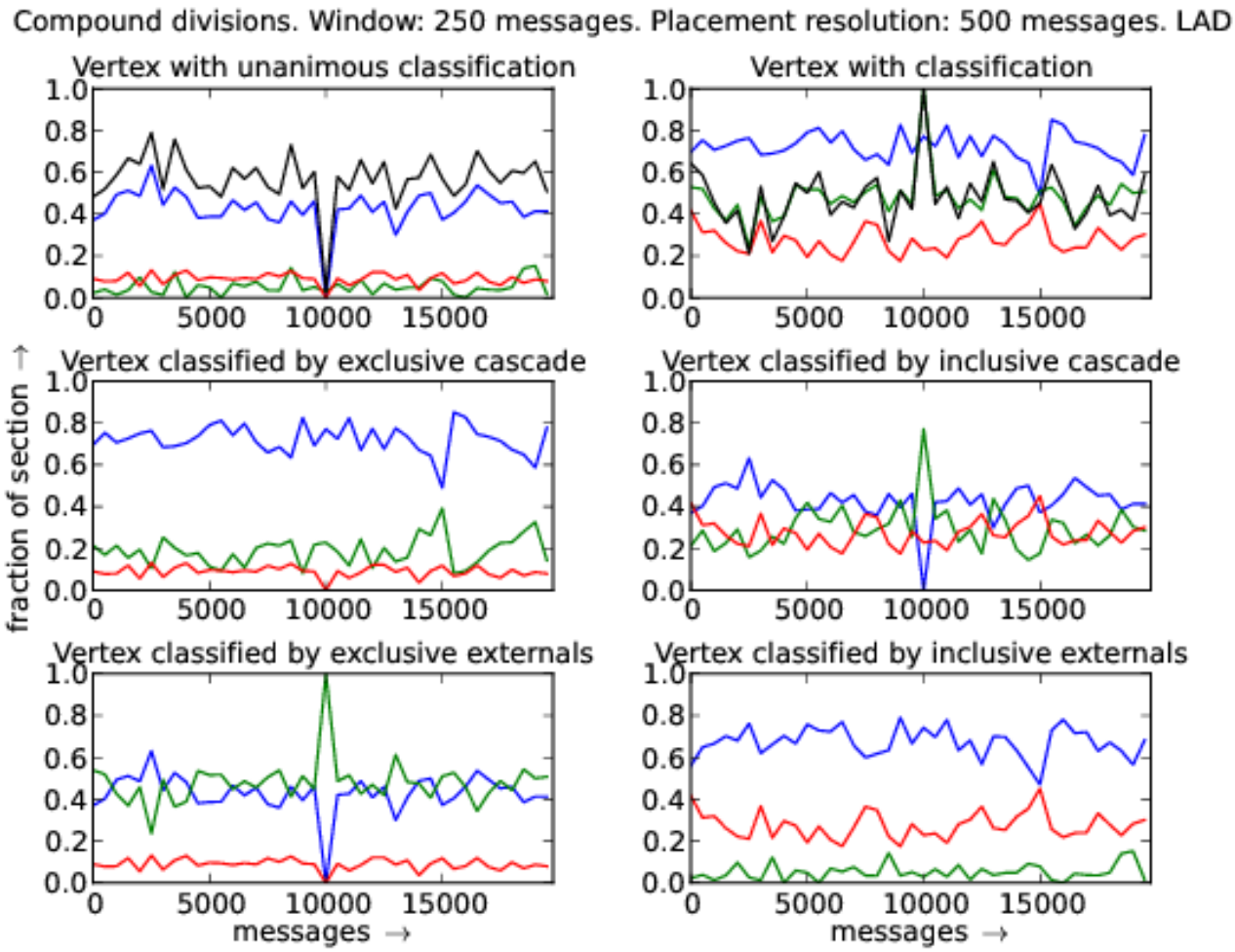

FIG. 28. Distribution of vertex with respect to compound criteria. Red, green and blue designate hubs, intermediary and border (peripheral) vertex fractions. The first two plots exhibit classifications that are not functions. Thus, in the first plot, the fraction of vertices with unique classification in plotted in black. On the second plot, black represents the fraction of vertices that has more than one class: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria is described in Section III C.

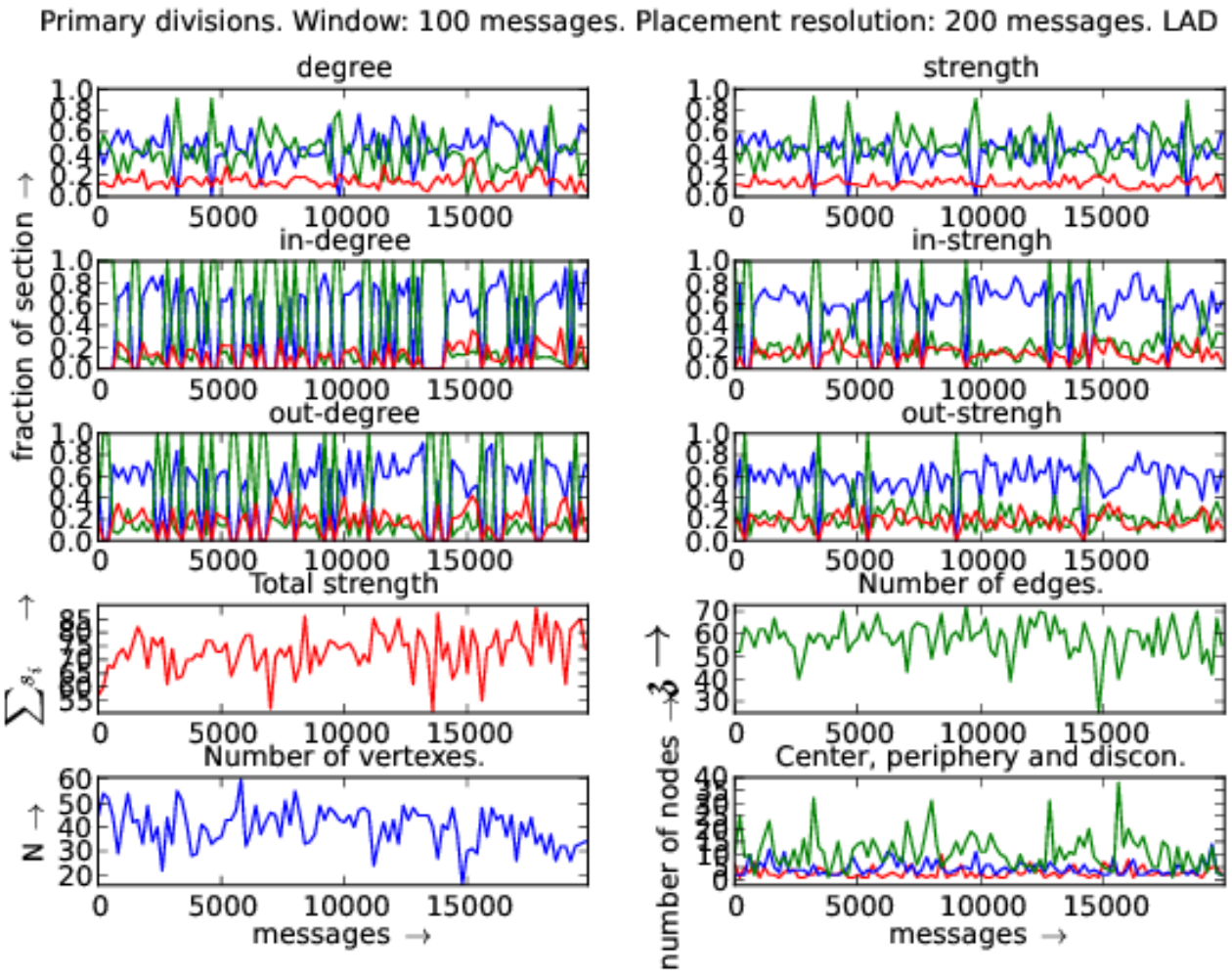

FIG. 29. Distribution of vertices with respect to each centrality measure: in and out degrees and strengths. Linux Audio Users (LAD) official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex is equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertices.

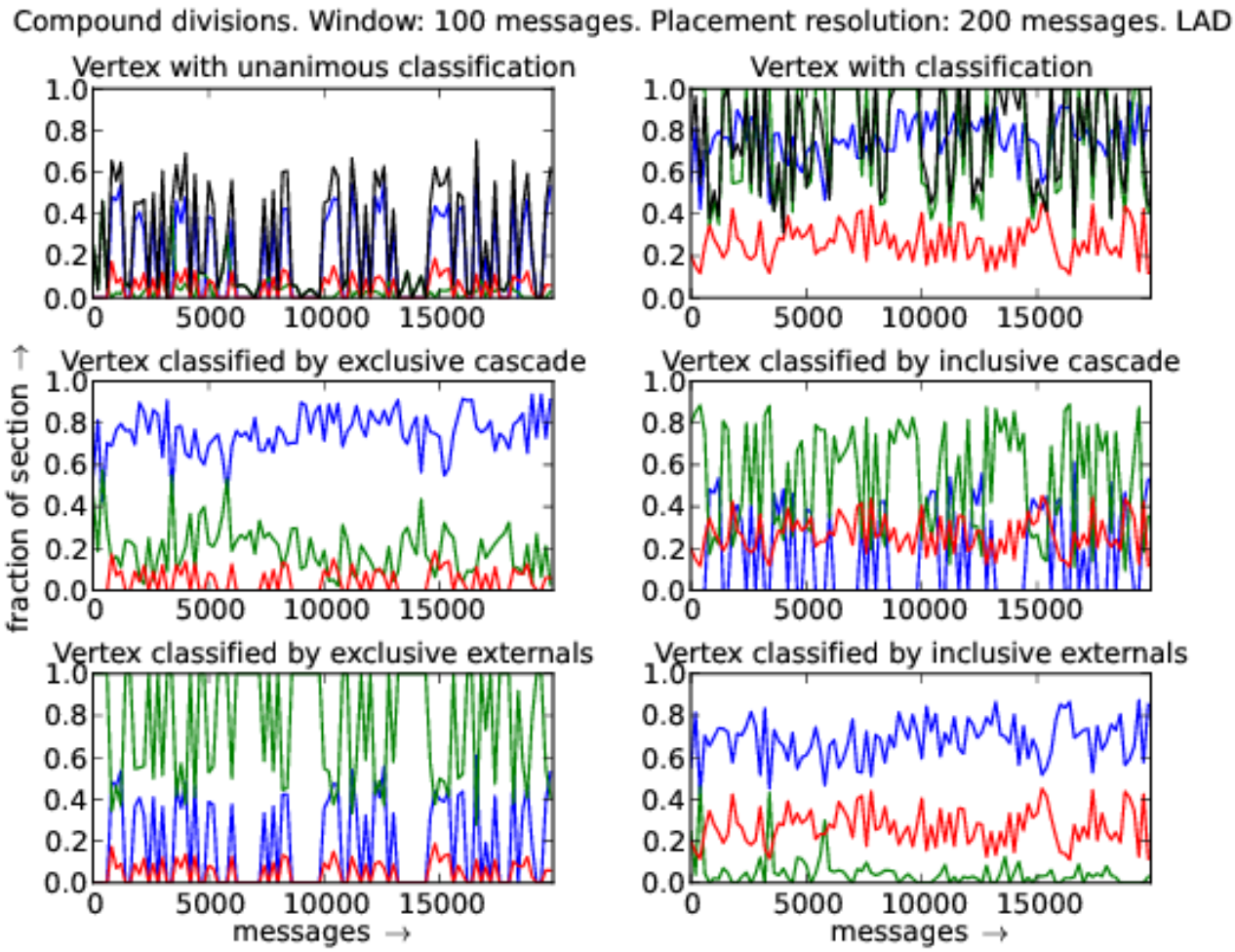

FIG. 30. Distribution of vertex with respect to compound criteria. Red, green and blue designate hubs, intermediary and border (peripheral) vertex fractions. The first two plots exhibit classifications that are not functions. Thus, in the first plot, the fraction of vertices with unique classification in plotted in black. On the second plot, black represents the fraction of vertices that has more than one class: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria is described in Section III C.

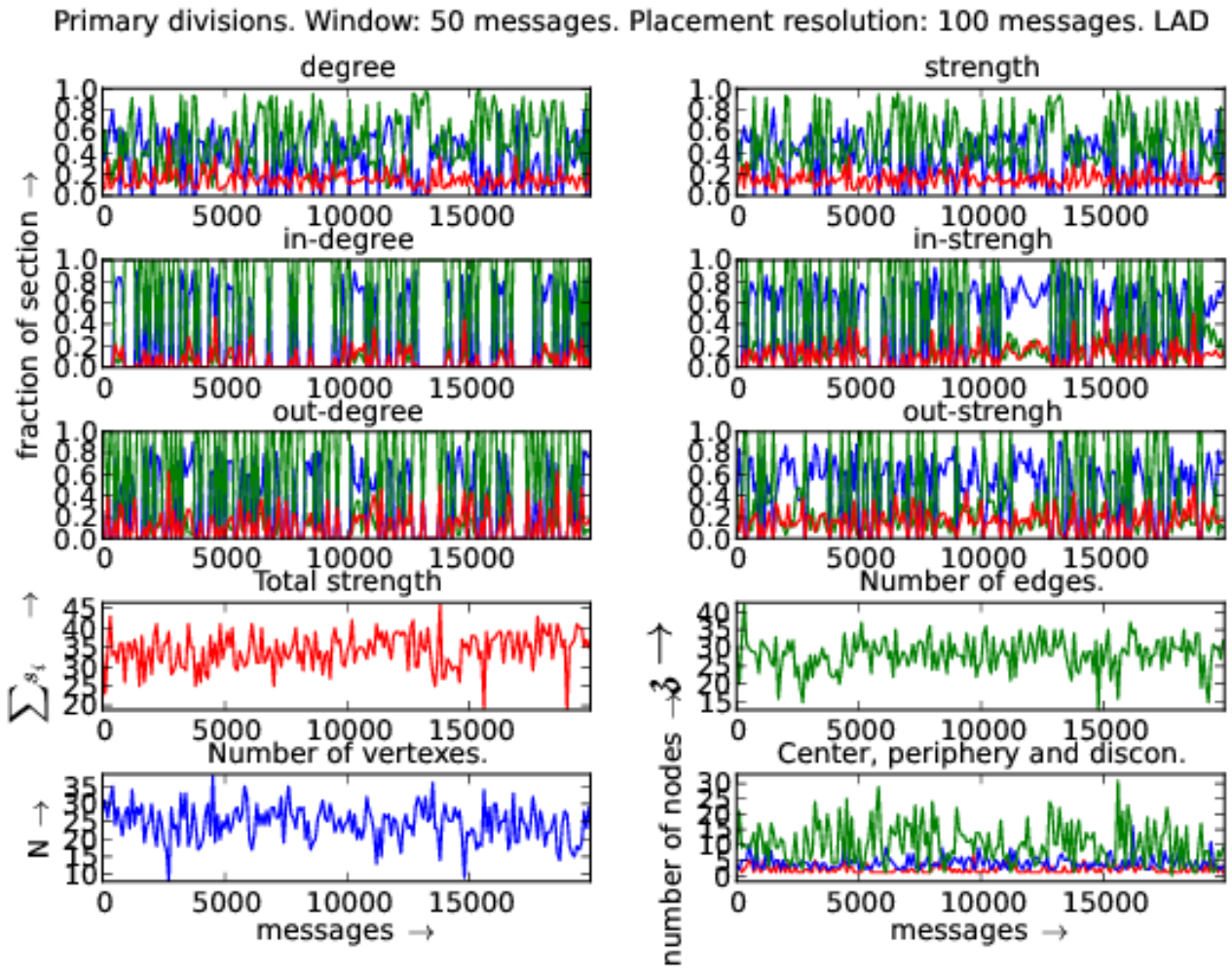

FIG. 31. Distribution of vertices with respect to each centrality measure: in and out degrees and strengths. Linux Audio Users (LAD) official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex is equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertices.

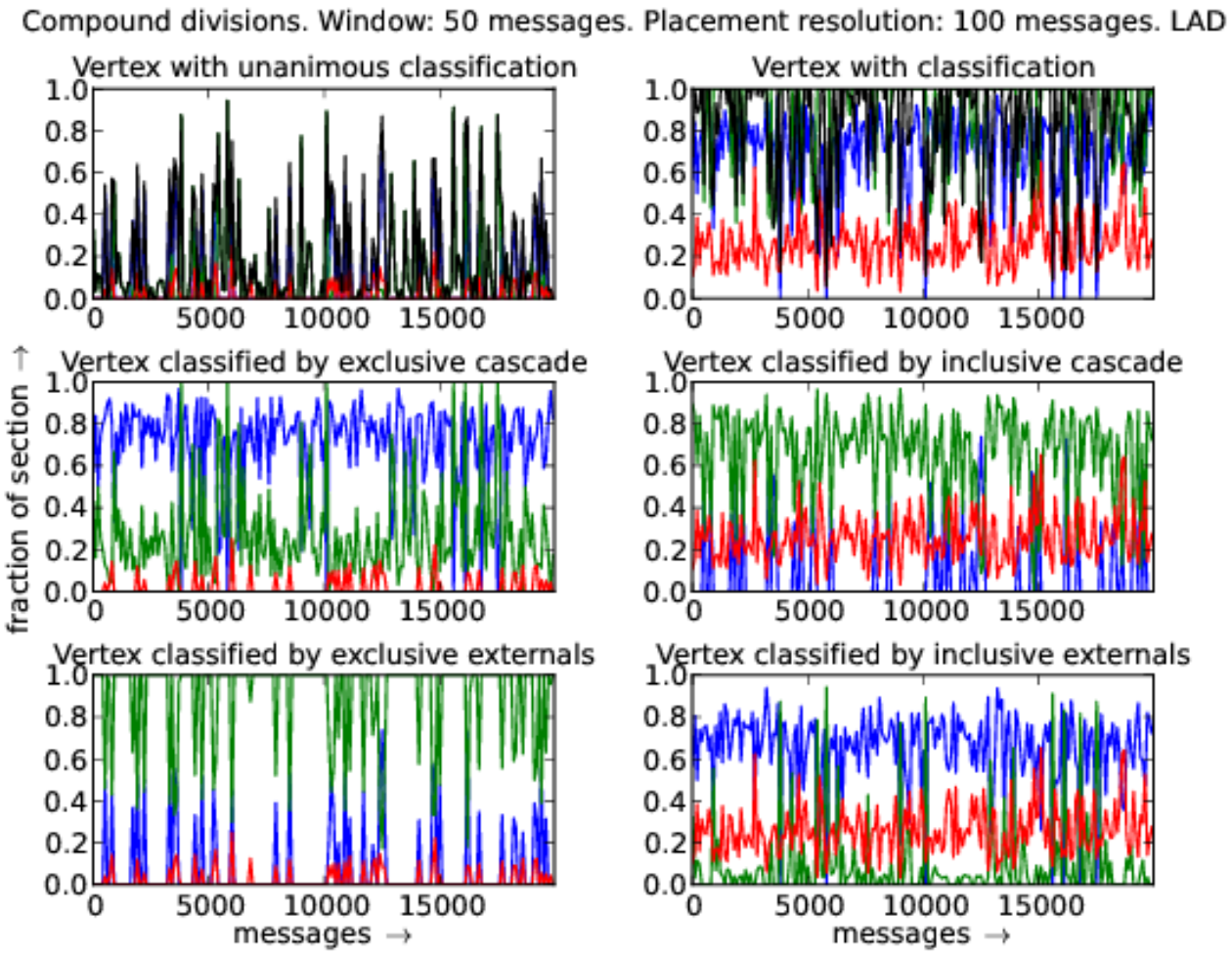

FIG. 32. Distribution of vertex with respect to compound criteria. Red, green and blue designate hubs, intermediary and border (peripheral) vertex fractions. The first two plots exhibit classifications that are not functions. Thus, in the first plot, the fraction of vertices with unique classification in plotted in black. On the second plot, black represents the fraction of vertices that has more than one class: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria is described in Section III C.