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

Renato Fabbri,[1, a] Vilson Vieira da Silva Junior,[b] Ricardo Fabbri,[c] Deborah Christina Antunes,[d] and Marilia Mello Pisani[e]
São Carlos Institute of Physics, University of São Paulo (IFSC/USP)

(Dated: 7 January 2015)

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

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

'The conception of personality structure is the best safeguard against the inclination to attribute persistent trends in the individual to something "innate" or "basic" or "racial" within him. The Nazi allegation that natural, biological traits decide the total being of a person would not have been such a successful political device had it not been possible to point to numerous instances of relative fixity in human behavior and to challenge those who thought to explain them on any basis other than a biological one.'
- *Adorno et al, 1969, p. 747*

I. INTRODUCTION

The present work is aimed at finding common characteristics among (email) interaction networks. This includes observations along time, which imply network evolution, a field that has received dedicated attention from the research community for more than a decade[?] [?] .

While significant measures will depend on the model and system characteristics[?] [?] , this work considers only directed, weighted and human interaction networks. Undirected and unweighted representation of such networks is also seen in the literature and can be obtained by simplification[?] .

Text mining and typologies of online participants benefit from results here presented[?] [?] . 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 (LinkedIn, Facebook, Twitter).

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[?] . Moreover, selected exceptions are reported in this section. The evolution considered in the present article is characterized by a constant number of messages, which is also present in literature, but was less explored to date.

The evolution of interaction networks was addressed with focus on community evolution, a work that ignores the direction of edges[?] . Two topologically different networks are reported to emerge, depending on the frequency of interactions, which present a generalized power law or an exponential connectivity distribution[?] . Recent work on email lists consider an isolated snapshot in order to verify or draw hypothesis. In such, free-scale properties were verified[?] , and different linguistic traces were related to weak and strong ties[?] .

Such results are in accordance with phenomena observed in this work and linguistic characterization is be-

[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)

ing described in a deriving article[?] . See Appendix B for further considerations of related work.

II. DATA DESCRIPTION

A. Email lists and messages

Email list messages were obtained from the GMANE email archive[?] , which consists of more than 20,000 email lists and more than 130,000,000 messages[?] . These lists cover a variety of topics, mostly technology-related. It can be described as a corpus with metadata of its messages, like send time, place, sender name, sender email address etc. GMANE usage in scientific research is reported in studies of isolated lists and of lexical innovations[?, ?] . Appendix A is dedicated scripts for gathering and processing GMANE email messages.

B. Chosen dataset

The four lists below were selected for their diversity, easing initial observance of natural and general properties.

- Linux Audio Developers list[?] . Participants are from different countries, and English is the language used the most. More technical and less active version of LAU. Abbreviated LAD from now on.
- Development list for the standard C++ library[?] . Dominated by specialized computer programmers. Participants are from different countries, and English is the language used the most. Abbreviated as CPP from now on.
- List of the MetaReciclagem project[?] . Dominated by Brazilian activists and digital culture interests. Participants are mostly Brazilians, and Portuguese is the most used language, although Spanish and English usage is commonplace. Abbreviated MET from now on.
- Linux Audio Users list[?] . Dominated by participants with hybrid artistic and technological interests. Participants are from different countries, and English is the language used the most. Abbreviated as LAU from now on.

The first 20,000 messages of each list were considered, with total timespan, authors, threads and missing messages exposed in Table I.

III. CHARACTERIZATION METHODS

After immersion in the data and on appropriated literature, the following methods for network characterization were chosen: 1) statistics of activity along time; 2) division of network in hubs, intermediary and peripheral vertex; 3) prominence of topological measures; 4) evolutive visualizations and quantitative observations; 5) typological elaborations of networks and participants. Each of these methods are described bellow, as other handy alternatives.

list	$date_1$	$date_M$	N	Γ	$\overline{M}$
LAU	Jun/29/2003	Jul/23/2005	1183	3373	5
LAD	Jun/30/2003	Oct/07/2009	1268	3113	4
MET	Ago/01/2005	Mar/07/2008	492	4607	23
CPP	Mar/13/2002	Aug/25/2009	1052	4506	7

TABLE I. Columns $date_1$ and $date_M$ have first and last messages dates from the 20,000 messages considered. 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 messages missing in the 20,000 collection, $100\frac{23}{20000} = 0.115$ percent in the worst case. MET notably has the fewer participants and the larger number of threads. This relation holds for each pair of the lists considered: as the number of participants increases, the number of threads decreases.

A. Temporal activity statistics

Number of messages along time with respect to seconds, minutes, hours, days of the week or the month, and months of the year. These were displayed as tables in Appendix C 2, results are outlined in Section IV A.

B. Interaction network

Regarding literature[?, ?, ?] , interaction networks might be both weighted or unweighted, both directed or undirected. Networks in this article are directed and weighted, considered as more informative among possibilities (directed unweighted, and undirected weighted and undirected unweighted). More precisely, the networks reported are erected as follows: a direct response from participant B to participant A forms an edge from A to B, as information went from A to B. The reasoning is: if B wrote a response to A, he read what A wrote and formulated a response, so B assimilated information from A, thus $A \to B$. Inverting edge direction yields the status network, as B read the message and considered what A wrote worth of responding, giving status to A, thus $B \to A$. This article uses the information network described above and depicted in Figure 1. Edges in both directions are allowed. Each time an interaction occurs, one is added to edge weight. Self-loops were regarded as non-informative and discarded. This networks are described as exhibiting free-scale and small world properties, as expected for a social network[?] .

Previous messages on the thread create directed edges from their author to the observed message's author.

Edges can be created from all antecedent messages on the message-response thread. Is this work, only immediate predecessors are linked to new message's 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 connection between $x \to z$. Potential interpretations for this weaker connection are usually common sense, such as: double length, half weight or with one more "obstacle". This suggests the adoption of other centrality measures that account for the connectivity with all nodes, such as betweenness centrality and accessibility[?] [?] .

1. Sectioning network in periphery, intermediary and hubs classes

Because of social networks' tendency to have scale-free distribution of properties, one can compare it to an Erdös-Rényi random graph and consider peripheral, intermediary and hub sectors[?] , as depicted in Figure 2. The degree distribution $\widetilde{P}(k)$ of an ideal scale-free network $\mathcal{N}_f$ with N vertexes and z edges, has less average degree nodes when compared with the distribution $P(k)$ of an Erdös-Rényi random graph with the same number of vertexes and edges:

$$\widetilde{P}(k) < P(k) \Rightarrow \text{k is intermediary degree} \qquad (1)$$

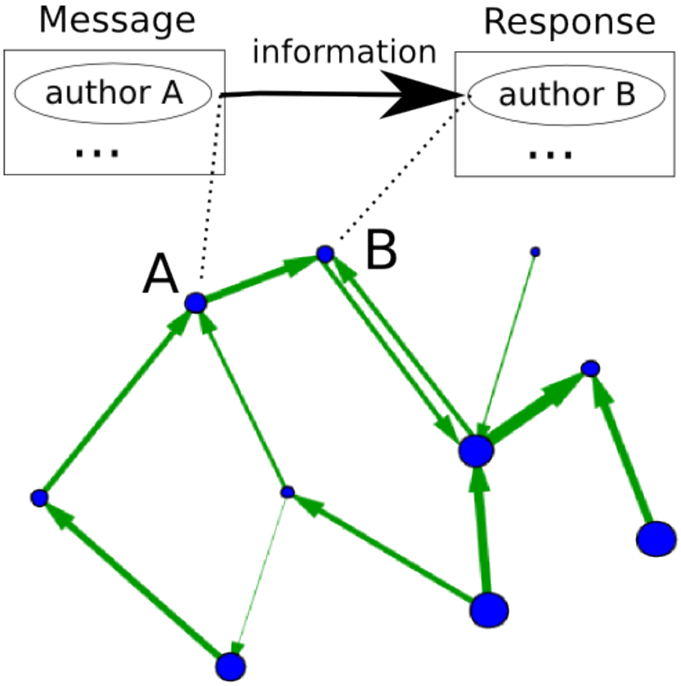

FIG. 1. Formation of interaction network from email messages. Each vertex represents a participant. If participant B replies participant A, that is regarded as evidence that B received information from A. Multiple messages add "weight" to directed edge. Further details are in Section III B.

If $\mathcal{N}_f$ is directed and has self-loops, the probability of the presence of an unknown edge is $p = \frac{z}{N(N-1)}$, where $N(N-1)$ is the maximum number of edges for a network with N vertexes, directed edges and without selfloops. A vertex in the ideal Erdös-Rényi digraph with the same number of vertexes and edges, and thus the same probability p for the presence of an edge, will have degree k with probability:

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

The lower degree fat tail constitute the border vertexes or peripheral sector. The higher degree fat tail is the hub sector.

The arguments behind this classification are: 1) vertexes so connected that they are virtually inexistent in networks connected at pure chance, specially without preferential attachment, are correctly associated to hubs sector. Vertexes with very few connections, which are way more abundant than expected by pure chance, are correctly associated to periphery. Degree values predicted as the most abundant if connections are created by pure chance, near the average, and less frequent in free-scale phenomena, are correctly associated to intermediary vertexes.

To assure statistical validity, bins can be chosen to span the average of η gaps between measure values. Thus, each bin, starting at degree k_i, spans $\Delta_i = \frac{\sum_{i=1}^{\eta}(k_{i+1}-k_i)}{\eta}$ values, with borders k_i and $k_i + \Delta_i$. This changes equation 1 to:

$$\sum_{x=k_i}^{k_i+\Delta_i} \widetilde{P}(x) < \sum_{x=k_i}^{k_i+\Delta_i} P(x) \Rightarrow \text{i is intermediary} \qquad (3)$$

If instead strength s is used for comparison, P remains the same, but $P(\kappa_i)$ with $\kappa_i = \frac{s_i}{\overline{w}}$ should be used for comparison, with $\overline{w}$ the average weight of an edge and s_i the vertex strength. For in and out degrees and strengths, comparisons should be made with $\kappa_i = 2k_i^{in}$, $\kappa_i = 2k_i^{out}$,

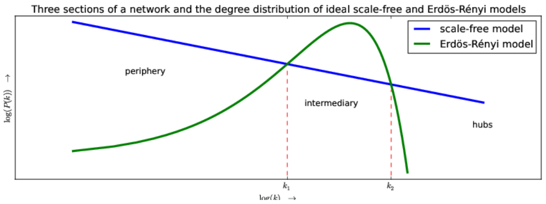

FIG. 2. Degree distribution on scale-free and Erdös-Rényi ideal networks. The later has more intermediary, as the former has more peripheral and hubs. Sections are given by the two intersections of the distributions k_1 and k_2. Characteristic degrees are in compact intervals of degree: $[0, k_1]$, $(k_1, k_2]$, $(k_2, k_{max}]$ for the three sections considered (periphery, intermediary and hubs).

$\kappa_i = 2\frac{s_i^{in}}{\overline{w}}$ and $\kappa_i = 2\frac{s_i^{out}}{\overline{w}}$. Results of these segmentations are discussed in subsection IV B.

As a further refinement of the network segmentation, compound criteria is used for classification of vertex, considering all measures: total, in and out degree and strength. After a careful inspection of possible combinations, these were abbreviated to six:

- Exclusivist criteria: vertex are only classified if the class is the same with respect to all measures. In this case, total vertex classified (usually) does not reach 100%, which is specified by a black line in Appendix D.
- Inclusivist criteria: vertex has class given by any of the measures. In this case, a vertex can have more than one class and the fraction of class attribution beyond total number of vertexes is also represented by a black line in Appendix D.
- Exclusivist cascade: hubs are only hubs if classified as hub with respect to all measures. Intermediary are the vertexes classified as intermediary or hub with respect to all measures. Vertexes left are regarded as peripheral vertex.
- Inclusivist cascade: hubs are vertexes classified as hubs by any of the measures. From the vertexes left, if any is classified as intermediary by any measure, than it is intermediary. The rest of the vertexes are peripheral.
- Exclusivist externals: hubs have unanimous classification with respect to all measures. Of vertexes left, peripheral vertexes are the ones classified as hub or peripheral by simple criterion. The rest represent intermediary sector.
- Inclusivist externals: hubs are vertexes classified as hubs with respect to any measure. From vertexes left, if a vertex is classified as peripheral with respect to any measure, than it is peripheral. The rest is regarded intermediary sector.

These compound criteria, and reduction of possibilities to them, can be formalized in strict mathematical terms. This was considered out of the scope of the present article.

Results from applying this classification method is further reported in Section IV B.

2. Topological measures

This article restricts this analysis to a small selection of the most basic measures of each vertex:

- Degree d_i: number edges linked to node i.
- In-degree d_i^{in}: number of edges ending at node i.
- Out-degree d_i^{out}: number of edges departing from node i.
- Strength s: sum of weights of all edges linked to node i.
- In-strength s_i^{in}: sum of weights of all edges ending at node i.
- Out-strength s_i^{out}: sum of weights of all edges departing from node i.
- Clustering coefficient cc_i: fraction of pairs of neighbors of i that are linked. Standard clustering coefficient for undirected graphs was used.
- Betweenness centrality bt_i: fraction of geodesics that contain the node i. Betweenness centrality index considered directions and weight, as specified in[?] .

In order to capture asymmetries in the activity of participants, the following metrics were introduced (see subsection IV C):

- asymmetry of note i: $asy_i = \frac{d_i^{in} - d_i^{out}}{d_i}$.
- mean of asymmetry of edges: $\mu_i^{asy} = \frac{\sum_{j \in J_i} e_{ji} - e_{ij}}{|J_i|}$. Where e_{xy} is 1 if there is and edge from x to y, 0 otherwise. $|J_i|$ is the number of neighbors of vertex i.
- standard deviation of asymmetry of edges: $\sigma_i^{asy} = \sqrt{\frac{\sum_{j \in J_i} [\mu_{asy} - (e_{ji} - e_{ij})]^2}{|J_i|}}$
- disequilibrium: $dis_i = \frac{s_i^{in} - s_i^{out}}{s_i}$.
- mean of disequilibrium of edges: $\mu_i^{dis} = \frac{\sum_{j \in J_i} \frac{w_{ji} - w_{ij}}{s_i}}{|J_i|}$, where w_{xy} is weight of edge $x \rightarrow y$ and zero if there is no such edge.
- standard deviation of disequilibrium of edges: $\sigma_i^{dis} = \sqrt{\frac{\sum_{j \in J_i} [\mu_{dis} - \frac{(w_{ji} - w_{ij})}{s_i}]^2}{|J_i|}}$

C. Evolutive observations

Evolution of network is observed within a fixed number of messages (window size: ws) that shifts in the message timeline. All ws =50, 100, 200, 400, 500, 800, 1000, 2000, 2500, 5000 and 10000 were used. Within a same ws, the number vertex and edges vary in time, as do other network characteristics. Further work should deepen inspection of measure interdependence, this article holds to measures in Section III B 2.

Visualization of network evolution

In refining hypotheses, visualization of the network was crucial. Animations, image galleries and an online gadgets were made[? ? ?] . Mapping of various topological measures to glyphs and layouts are being further explored as a parallel research. Furthermore, stable aspects of measures prominence along time are captured through mean and standard deviation (see Section III B 2). Constant sector sizes along time are observed in a timeline fashion in Appendix D.

D. Typological deepening

There are other ways to split a network. To point a common example, the center of the network is defined as all the nodes whose maximum distance to any other node is the radius[?] . In the same framework, the periphery (as opposed to the center) consists of the nodes whose maximum distance to any node is the diameter[?] . Accordingly, the intermediary sector can be defined as the nodes that are not in the center or in the periphery. Interestingly, in the email networks analyzed, with these criteria, the center can often be a factor of 4 times larger than the periphery and the intermediary group often exceed 93% of the nodes[?] .

Models of human dynamics can be used to predict and classify activity. In this case, agent activity is commonly considered a Poisson process, as a consequence of the randomly distributed events in time. Even so, evidence-based models suggests that human activity patterns follow non-Poisson statistics, characterized by a long tail of inactivity with of bursts of rapidly occurring events[? ?] . Emails are reported as having a heavy tailed distribution with $\alpha = 1$, together with web browsing and library loans[?] .

Typologies can also be conveniently adapted from psychiatric, psychological and psychoanalytic theories. Concerning empirical research, Theodor Adorno was a core conceiver of an one-of-a-kind typology that resulted from observing authoritarian personality traces[?] , sometimes depicted as an authoritarian syndrome. Other typologies include Jung's extroversion-introversion trait with four modes of orientation. This four modes are divided in two perceiving functions (sensation and intuition) and two judging functions (thinking and feeling)[?] . Myers-Briggs Type Indicator extrapolated Jungian theories into a questionnaire and added perceiving and judging as a fourth dipole[?] . 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[? ?] .

It was considered central to benefit from key human typologies, both by adding descriptions to a type and by further characterizing classes in the terms encountered.

IV. RESULTS AND DISCUSSION

A. Constancy and discrepancy of activity along time

1. Seconds and minutes

The incidence of messages at each second in a minute and at each minute in an hour is compatible with uniform distribution tests. If compared to simulations using an uniform distribution[?] , messages were slightly more evenly distributed in all lists: for both seconds and minutes $\frac{max(incidence)}{min(incidence)} \in (1.26, 1.275]$. 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 $\sigma_\xi = 0.04619$.

2. Hours of the day

Table V shows how the four lists distribute activity along the day. Afternoon was the most active period 6h of the day. Second 12h more active than first 12h. Even so, activity peak occurs around midday, with a slight skew towards earlier hours.

3. Days of the week

Weekdays also exhibit an interesting pattern, to which is dedicated Table VI. Days exhibit pattern on weekdays, with a decrease of at least one third and reaching two thirds on weekends.

4. Days along the month

Table VII shows activity along the month.

5. Months and larger divisions of the year

Table VIII is dedicated to activity in months and larger divisions of the year.

As for months along the year, there seems to be two periods of more prominent activities: Jun-Aug (MET and LAD), and Dec-Mar (CPP, LAU and LAD). These observations fit academic calendars, vacations and end-of-year holidays.

Variation of activity in the days along the month are less prominent, one cannot point much more than a (probably not statistically relevant) tendency to first and second weeks to be more active. The most important result in the days along the month is their homogeneity with respect to activity. Last days of the month (29, 30 and 31) are not present in every month, and observed activity is proportional to incidence rates.

B. Scalable fat-tail structure

There is a concentration of hub activity and of vertex with few connections. Table IX is dedicated to exposing this distribution of activity among participants.

As specified in Section III B 1, in order to classify nodes as hubs, intermediary or peripheral, $\widetilde{P}(k)$ is compared to $P(k)$. For the networks studied here, i.e. derived from public email lists, if degree distribution is used classification, hubs sector size reaches peaks with 10% of all vertexes. If strength is used for comparison ($\widetilde{P}(k_i = \frac{s_i}{\overline{w}})$) is compared to $P(k)$), hubs account for approximately 5% of all vertex, i.e. strength classification yields half the number of hubs as plain degree. This results hold for in and out degrees and strengths,

Classification criteria exposed in subsection III B 1 was used efficiently with windows with at least 200 messages. Specially with 1000 or more messages, criteria yields stable fractions of $\approx 5\%$ of hubs, $\approx [15-20]\%$ of intermediary and $\approx [75-80]\%$ peripheral vertex. The compound criteria, also described in Section III B 1, can be used as a classification refinement. This is specially useful in dealing with fewer messages, in which case the structure degenerates with respect to some of the degree and strength measures, but not all.

For the networks analysed, differences of using this smoothing process were not significant. There were between 20 and 1200 participants, so each participant were between 5% and 0.08% of all participants. The bottom line is: for network sizes considered, if connectivity intensity would most probably not exist in an Edös Renyi network, than it is not an intermediary intensity. As peripheral vertex are abundant, this statistic discussion has no relevance.

A reasonable window size for observation can be inferred by monitoring the giant component size and the degeneration of the hub, intermediary and peripheral sections. This degeneration is critical in the span of 50-100 messages. With compound criteria, such as exclusive cascade of Figure **??**, the network seems to hold basic structure even with as few as 20-50 messages. This indicates that concentration of activity and low-activity participants occurs even with very few messages.

Appendix D is dedicated to figures on these networks and their evolution.

C. Prevalence of centrality over asymmetries and asymmetries over clusterization

The principal component exhibit ponderation of centrality measures: degrees, strengths and betweenness centrality. Clustering coefficient is presented in almost perfect orthogonality. Dispersion is more prevalent in symmetry related measures than clustering coefficient. This holds for all network snapshots observed, even with as few messages as to degenerate structure. Symmetric and asymmetric edges have been reported as bounded to different roles played by participants and relations[?] . Principal components formation from original measures can be observed in Tables II, III and IV. Individual vertexes relation with top two principal components can be appreciated in Figures 3 and 4. This peculiar first component that consists of the averaged sum of degree, strength and betweenness measures was verified to be incident in virtually all networks with 500 or more messages and most smaller networks (degeneration of basic structure is critical with $ws \approx 50-100$ messages). This composition of principal component suggests that all six degree and strength measures are equally important for system characterization, although it is known that they do not relate to the same participation characteristics.

As expected, degree and strength are highly correlated, with Spearman correlation coefficient $\in [0.95, 1]$ and Pearson coefficient $\in [0.85, 1)$ for window larger sizes ($ws > 1000$); and high degree is associated with low clustering coefficient, as can be observed in Figure 5.

When symmetry of node connectivity is considered, the first component remains mostly the same, but clustering coefficient is only relevant to third and fourth components. A snapshot of vertexes with respect to these first two principal components are in Figure 4. This asymmetry and disequilibrium measures revealed as more proper measures to characterization of hubs and intermediaries, as seen in greater spreading of second PCA plot. Both symmetry of node overall activity and of individual relations play important role in second component, as can be observed in Table IV.

D. Primitive typology

This work aimed at finding common characteristics among (email) interaction networks. The analysis involved primary measures observance and a formal criteria for coherent ratios of hub, intermediary and hub sectors. Nevertheless, inspection done by visualizations and raw data manipulations suggest peculiarities of interest, specially:

- Core hubs usually have intermittent activity. Very stable activity was found on MET hubs, which motivated its integration to this work. Literature reports greater stability of participation in smaller communities[?] , reason why smaller number of participants in MET was considered a direct cause of stable hubs activity.

- Typically, hub activity is trivial: they interact as much as possible, in every occasion with everyone. Peripheral vertex activity also follows a simple patters: they will interact very rarely, in very few occasions. Intermediary vertexes seem responsible for network structure.

- Network operation modes dictated by intermediary vertexes behavior. These can exhibit preferential communication to peripheral or hub vertex.

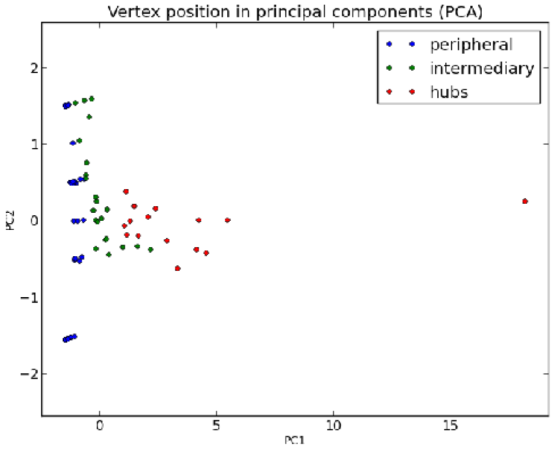

FIG. 3. PCA of in and out degree and strength, betweenness centrality and clustering coefficient, as specified in Section III B 2. Table III has the composition of principle components from the original measures. First principle component is a pondered sum of centrality measures: degrees, strengths and betweenness centrality. Second component is mostly clustering coefficient in this figure, but asymmetries holds second component if also considered. Similarity to plot in Figure 5 was verified with all window sizes considered ($ws \in [100, 10000]$), which exposes a common relation is held by degree, strength and betweenness measures to clustering coefficient.

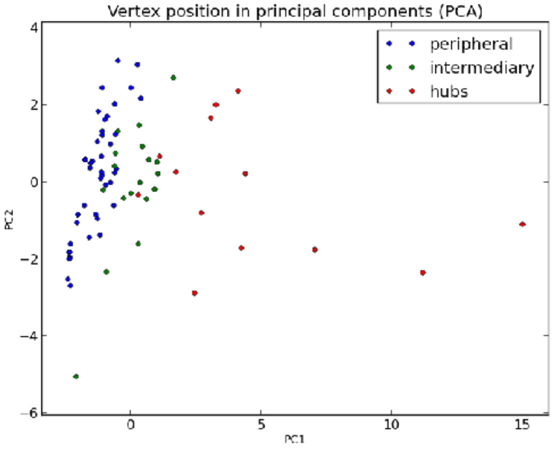

FIG. 4. Degree and strength, clustering coefficient, betweenness centrality and symmetry related measures are used for this scatter plot of principal components. Compositions of first three components are in table IV and measure details in subsection III B 2. Most importantly, clustering coefficient is only relevant for third component, being second component representative of symmetry measurements of vertex interactions.

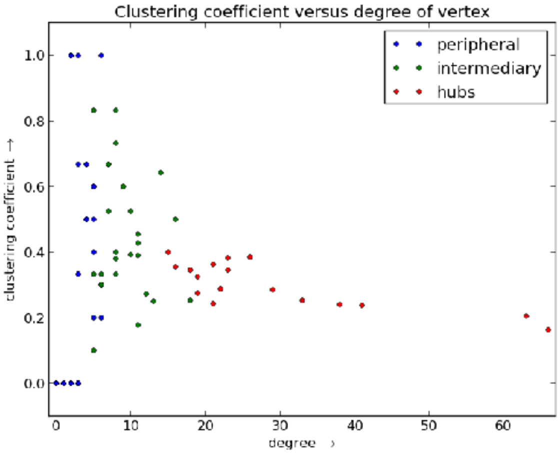

FIG. 5. Clustering versus degree of vertex in interaction network observed, $ws = 1000$ email messages, LAU list. General layout is in accordance with literature: connected vertexes have low clusterization while higher clusterization is gradually more incident as number of connections is lowered.

- Some of the most active participants receive many responses with relative few messages sent, and are never top hubs. These seem as authorities and contrast with participants that respond way more than receive responses.
- Most obvious community structure, as observed by hung 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 involved in other typologies, such Meyer-Briggs, Pavlov or F-Scale. This has no pretension of being a direct result from numeric analysis, it is a refinement the description of found structure and classes considered.

V. CONCLUSIONS AND FUTURE WORK

In fact, some characterization of chosen networks resulted from stable observations. Along temporal activity statistics, this work reports the stability of the principal components (in the concentration of dispersion and composition) and of the ternary partitioning relative sizes (evident with the comparison with the Erdös-Renyi model).

The task of delivering a first and general characterization of chosen interaction networks involved starting a larger effort. The different aspects covered requires not only different analytical background, but also considerations about textual production and social psychology. These are receiving attention within dedicated works and are summarized in this section.

A. Further work

1. Constancy of general characteristics, easing tipologization of outliers and mundane occurrences

Regarding topological aspects of interaction networks, further work should inspect other measures (e.g. closeness centrality, accessibility), and statistics in each of the three connective sectors: hubs, intermediary and peripheral.

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 sections, is to remove vertexes with degree close to k_1 or k_2 depicted in figure 2.

2. Interaction network characterization

Observed networks were coherent with literature in different aspects, such as concentration of activity, and clusterization versus connectivity patterns. Even so, verification of results in other virtual environment, such as Facebook, Twitter and LinkedIn, should verify the generality of this report.

3. Textual productions

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

4. Typology enhancements

From the stable characteristics, outliers can be pointed and further developed in terms of networks and participants typology.

5. Results exploitation

Usage of such characteristics are taking place in linked data and electronic government technologies[?][?]. Further steps involve elaboration and tests of social dynamics that takes advantages of these results

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[?] . Authors thanks 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: Data and scripts further description

Messages are downloaded from GMANE database by RSS in the mbox email text format. They are requested one by one to avoid reaching maximum size of the requests accepted by GMANE API.

Every message has about 30 fields, from which the following are crucial for the present work:

- “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 for, 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.

Basic constructs for obtaining all results in this article are described in A 2. Scripts, written in Python programming language, are publicly available at[?] and very briefly specified below.

1. Third party libraries and software

Programming resources used were mainly Python and part of the common scientific bundle for the language. More specifically, scripts where written for 2.7.3 version of Python, with the following third party libraries: Numpy, Pylab/Matplotlib, NetworkX, IGraph. Behind the scenes, Graphviz is accessed via PyGraphviz to make network drawings.

2. Python scripts

All results were obtained with scripts written in the Python programming language. These are kept in a public repository for backup and sharing with research community[?] . Core scripts, for deriving structures and results exhibited in this article, are in the LEIAME file.

Appendix B: Further consideration of related work

Unreciprocated edges often exceed 50%, which matches empirical evidence reported in[?] . Although no correlation of topological characteristics and geographical position was found in a pertinent study[?] , geographical incidences should be present in further refinement of the analysis.

The seminal Nature Letter by Palla, Baraási and Vicsek[?] has strong confluence with this work, suggesting that smaller size of MET community is responsible for the stronger hubs observed.

Controllability of these networks is also an uncovered issue. These has unintuitive properties and might bring into forefront crucial differences between email interaction networks and interaction networks in Facebook or Twitter[?] [?] [?] .

Gender related behavior has been notified in mobile phone datasets[?] , which can be further investigated to hold in email lists and in evolving terms, as a community oriented, non-private interactions are drawn from public emails groups with hundred of participants.

Considered years altogether, hundreds to thousands of participants post on a list, more rarely dozens or tenths of thousands. The most active lists usually reaches a few thousands of participants. Authors have not checked each list (more than 20 thousand public email groups[?]), and this might lead to a deeper insights in community-related network evolution.

Appendix C: Tables

1. PCA tables

	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 II. 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. First component is a pondered sum of degree and betweenness centrality measures. Second component is mostly clustering coefficient. First and second components sum more than 95% of total dispersion.

	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

TABLE III. Principal components' composition in percentages. LAU list, $ws = 1000$ messages in 20 disjoint positioning was used for statistics. First component is a pondered sum of degree and strength measures and betweenness centrality. Second component is mostly clustering coefficient. First and second components sum more than 90% of dispersion.

	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 IV. Distribution of components, added measures of symmetry described in subsection III B 2. LAU list, $ws = 1000$ messages in 20 disjoint positioning was used for statistics. In this case, clusterization is pushed to third component, with disequilibrium measures. Second component is primarily symmetry measures, but also some out degree and strength contribution. Betweenness again has a role similar to degree, but weaker. Clusterization component combines with disequilibrium, while asymmetry is related to out degree and strength. Three components has in average 80.36% of dispersion.

2. Tables for activity along time

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

	CPP			MET			LAU			LAD		
	1h	6h	12h	1h	6h	12h	1h	6h	12h	1h	6h	12h
0h	3.66	10.67	33.76	2.87	7.15	29.33	3.58	10.14	36.88	4.00	10.77	33.13
1h	2.76			1.77			2.22			2.52		
2h	1.79			1.04			1.63			1.79		
3h	1.10			0.64			1.06			1.06		
4h	**0.68**			0.47			0.84			0.75		
5h	0.69			**0.38**			**0.82**			**0.66**		
6h	0.83	23.09		0.72	22.18		1.17	26.74		0.85	22.36	
7h	1.24			1.33			2.37			1.56		
8h	2.28			2.67			3.54			2.96		
9h	4.52			4.40			6.04			4.68		
10h	6.62			6.29			**6.83**			5.93		
11h	**7.61**			6.78			6.79			6.40		
12h	6.44	**37.63**	**66.24**	**7.33**	**42.22**	**70.66**	6.11	**35.65**	**63.12**	**6.41**	**37.25**	**66.87**
13h	6.04			7.08			6.26			6.12		
14h	6.47			7.09			6.38			6.33		
15h	6.10			7.14			5.93			5.98		
16h	6.22			6.68			5.52			6.40		
17h	6.36			6.89			5.46			6.02		
18h	6.01	28.61		5.99	28.44		5.24	27.46		5.99	29.63	
19h	5.02			5.23			4.52			5.03		
20h	4.85			4.98			4.55			4.63		
21h	4.38			4.37			4.42			4.59		
22h	4.06			4.24			4.51			4.88		
23h	4.30			3.64			4.23			4.53		

TABLE VI. Concentration of activity on days along the week. Weekend days are at least $\frac{1}{3}$ less active and can reach $\frac{1}{3}$ of activity. MET concentrates activity in weekdays the most, leaving only 13.98% of total activity to Saturday and Sunday. LAU is the one that less concentrates activity in weekdays, reaching 20.94% of total activity in weekends. These might suggest professional relation of CPP and MET participants to the topics of interest, or a hobby relation of LAU and LAD participants.

	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 VII. Activity along the days of the month. As can be noted, 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 non existent.

	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	23.05	45.63	3.01	25.16	48.08	3.34	23.06	47.31	3.22	21.96	46.70
2	3.07			3.38			3.38			3.42		
3	3.20			3.55			3.20			2.87		
4	3.63			4.34			3.52			2.91		
5	2.85			3.93			2.68			3.30		
6	3.67			3.76			3.18			3.52		
7	3.45			3.18			3.77			2.27		
8	3.12	22.57		3.36	22.92		3.62	24.25		3.72	24.73	
9	2.57			3.44			3.82			3.97		
10	2.92			3.17			3.06			3.77		
11	3.54			3.88			3.11			3.27		
12	3.23			2.94			3.40			2.75		
13	3.39			3.29			3.55			3.34		
14	3.81			2.83			3.69			3.93		
15	3.35	23.02	46.31	2.72	21.87	43.56	3.23	22.84	44.01	3.37	22.82	46.00
16	3.77			2.96			2.94			3.37		
17	3.45			3.01			3.02			2.95		
18	3.47			3.39			3.63			3.22		
19	2.90			3.42			3.16			3.59		
20	2.80			3.09			3.25			3.21		
21	3.29			3.27			3.61			3.13		
22	2.88	23.29		2.92	21.69		3.80	21.17		3.07	23.18	
23	4.01			3.27			3.03			3.06		
24	3.13			2.92			2.31			2.72		
25	3.57			2.83			2.38			3.16		
26	3.27			2.97			3.49			3.57		
27	3.27			3.41			2.92			3.92		
28	3.17			3.36			3.26			3.69		
29	3.68	8.06	8.06	2.93	8.36	8.36	3.34	8.68	8.68	3.15	7.30	7.30
30	2.76			3.14			3.75			2.71		
31	1.63			2.29			1.60			1.45		

TABLE VIII. 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				

TABLE IX. 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 activity (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%)

Appendix D: 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.

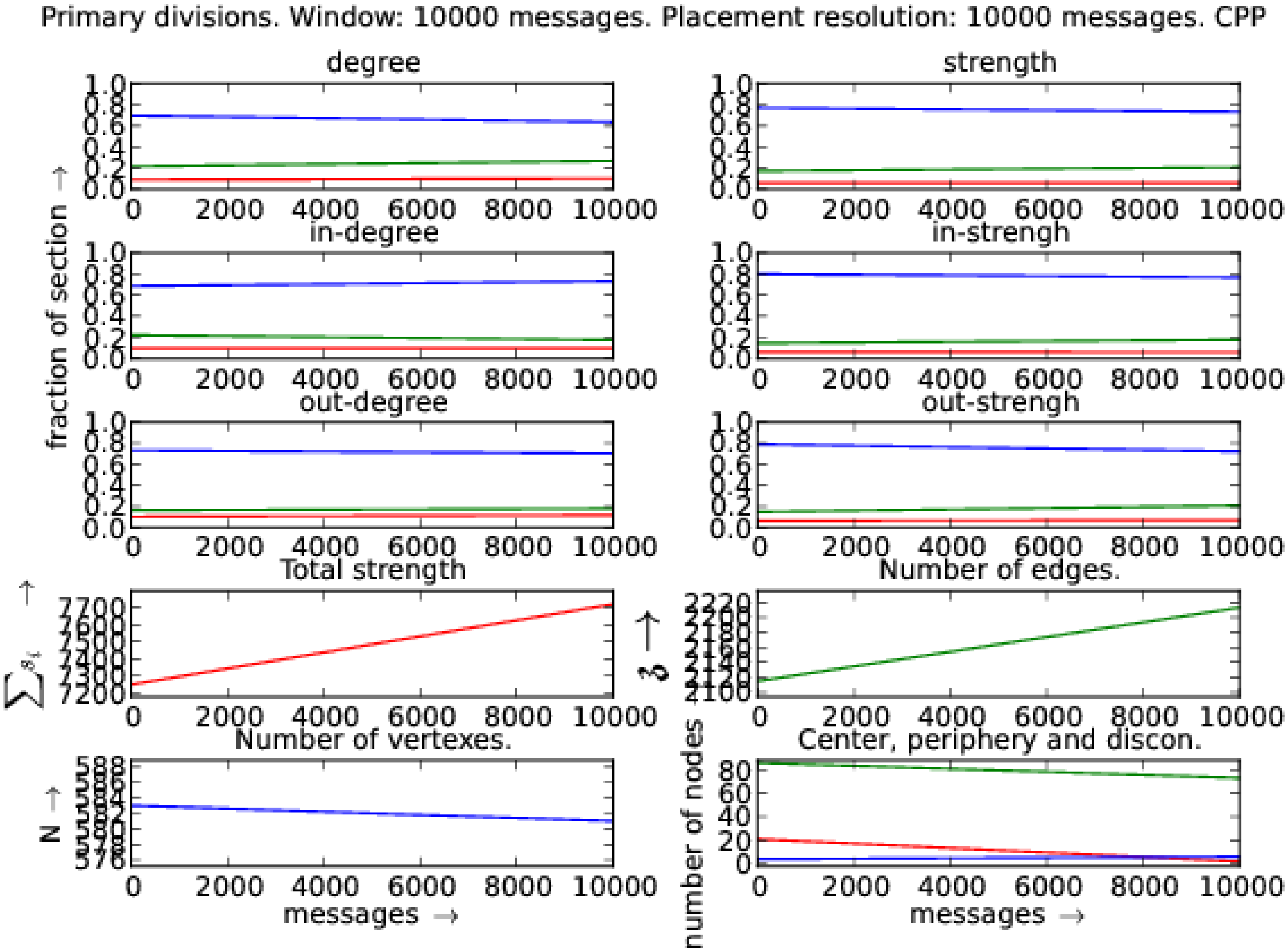

FIG. 6. Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.

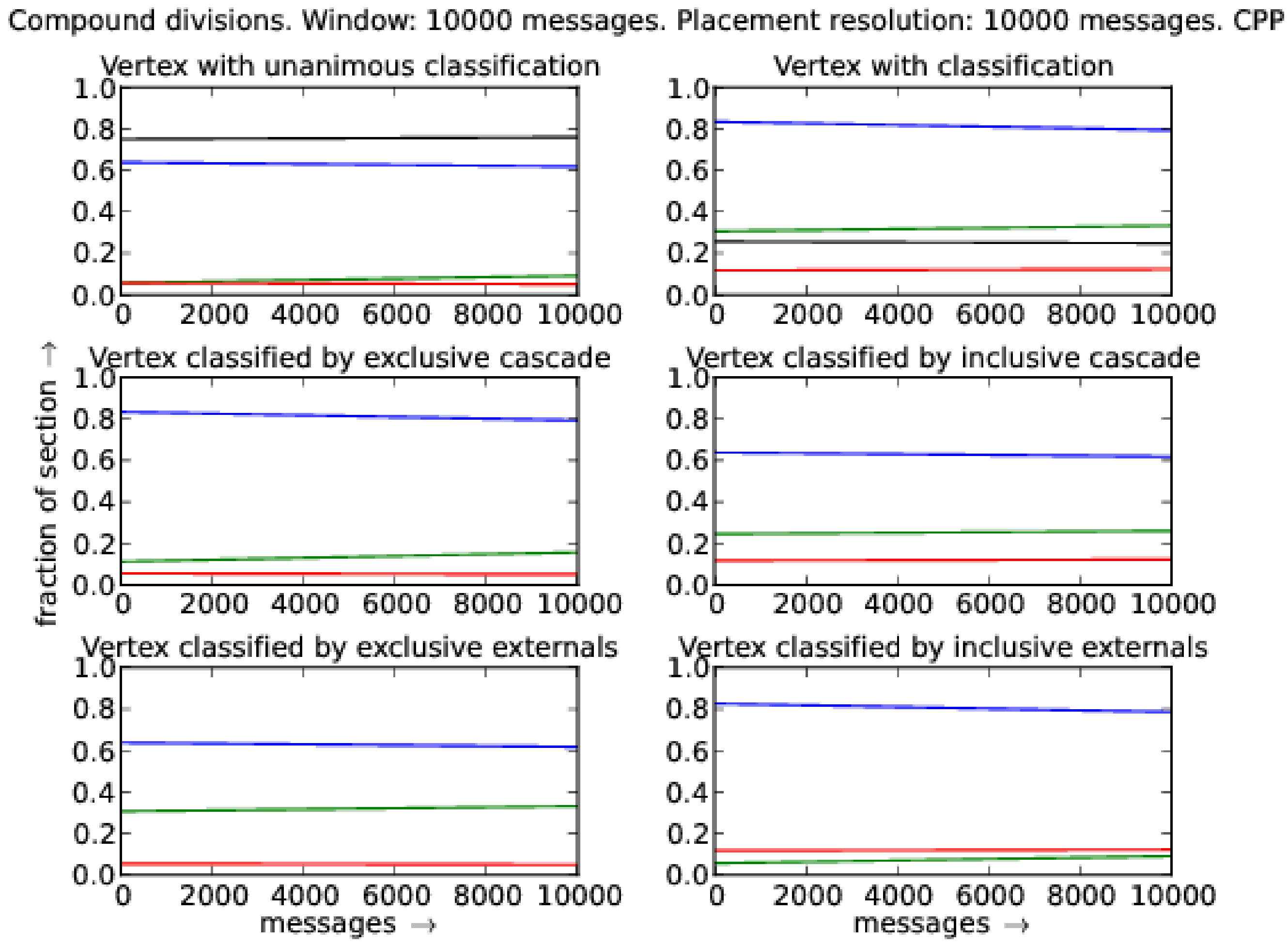

FIG. 7. Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection III B 1.

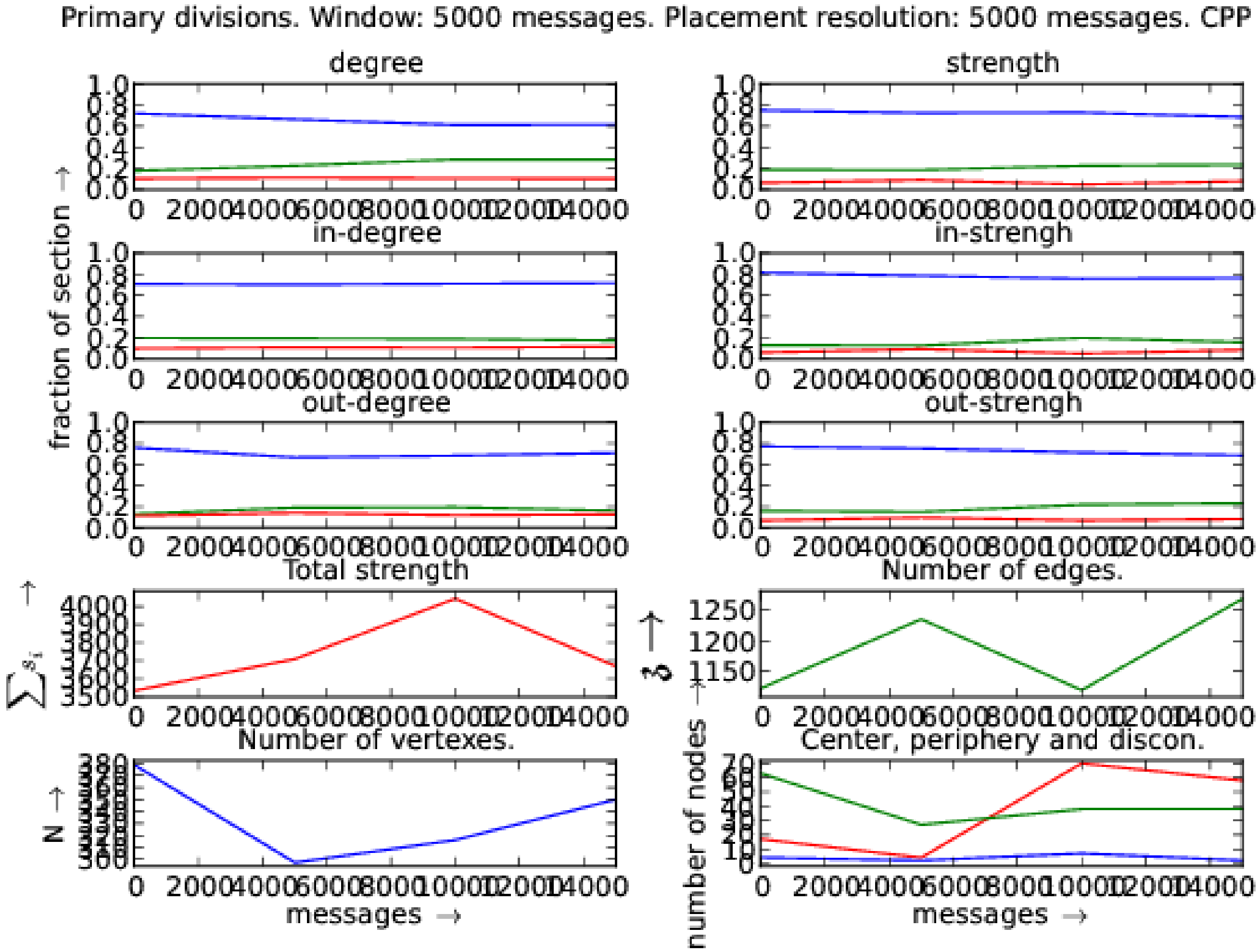

FIG. 8. Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.

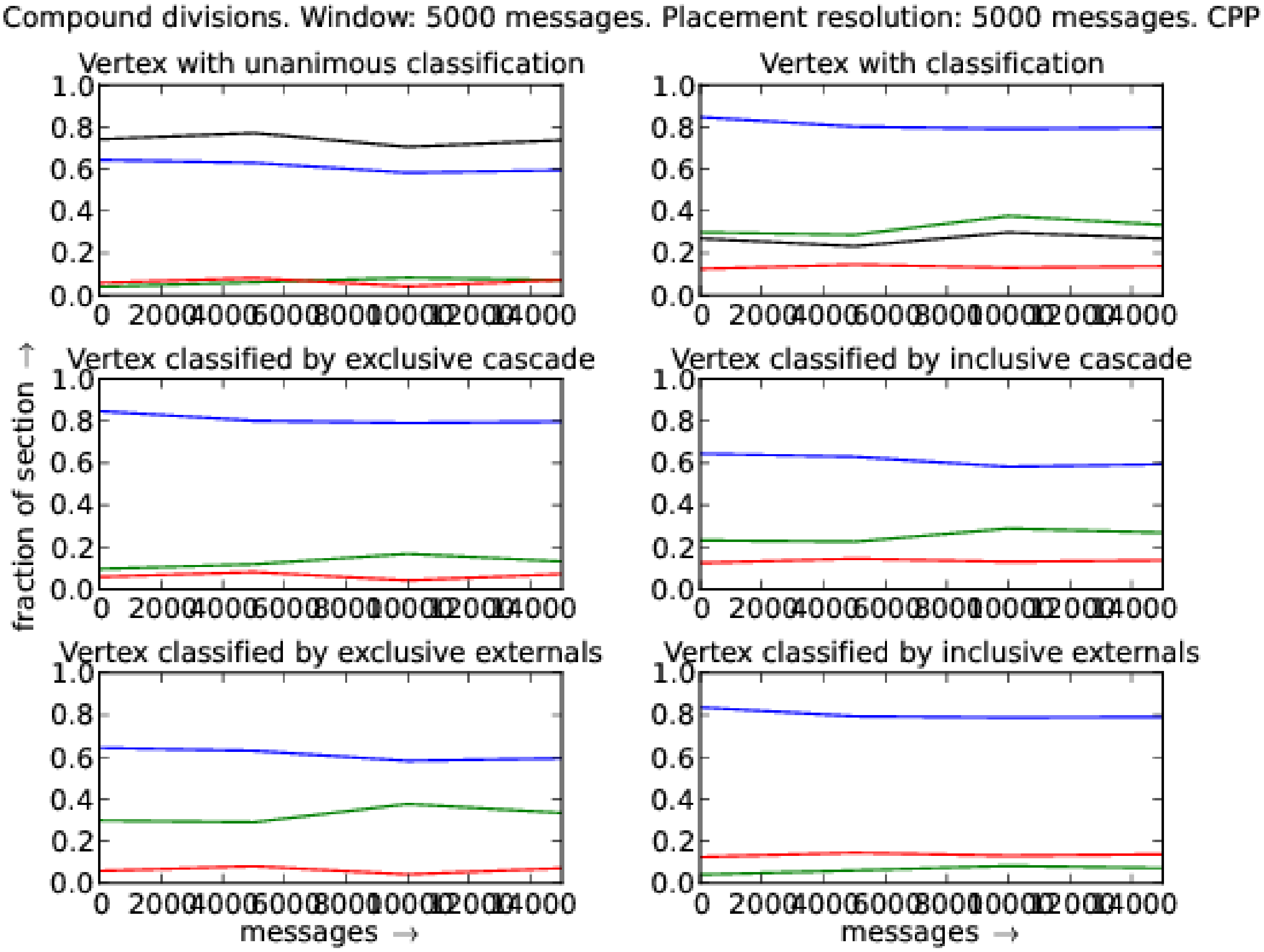

FIG. 9. Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection III B 1.

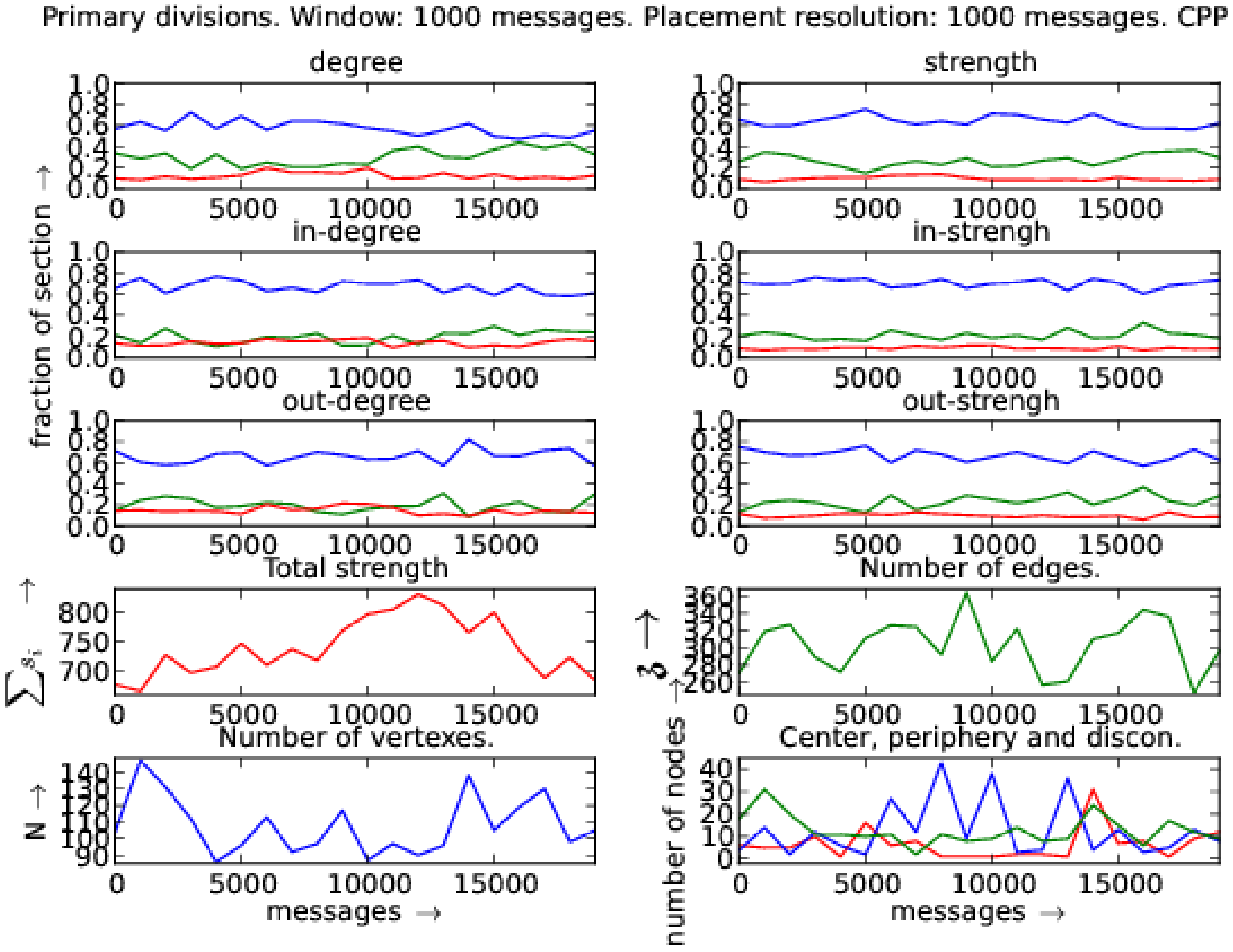

FIG. 10. Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.

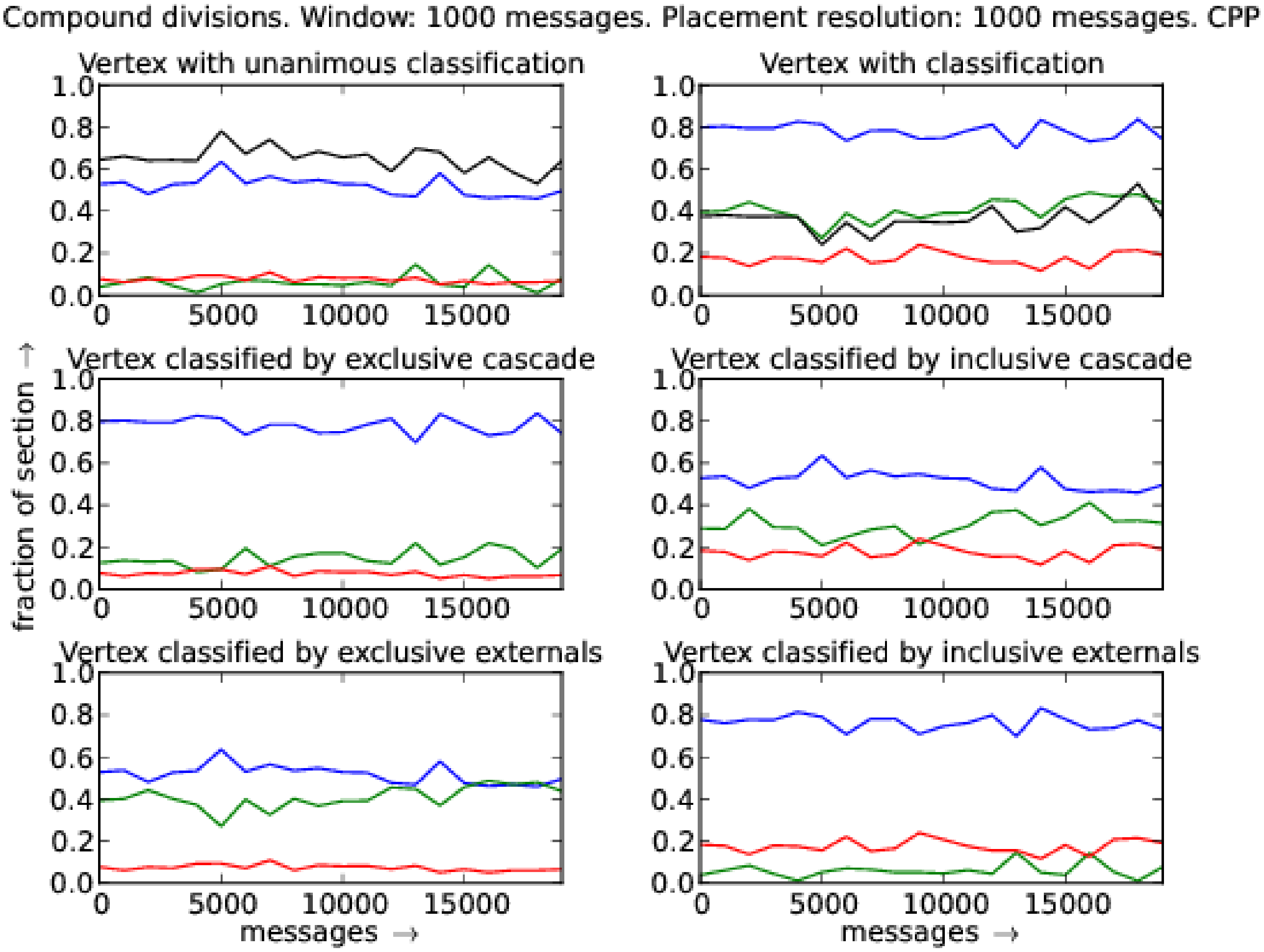

FIG. 11. Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection III B 1.

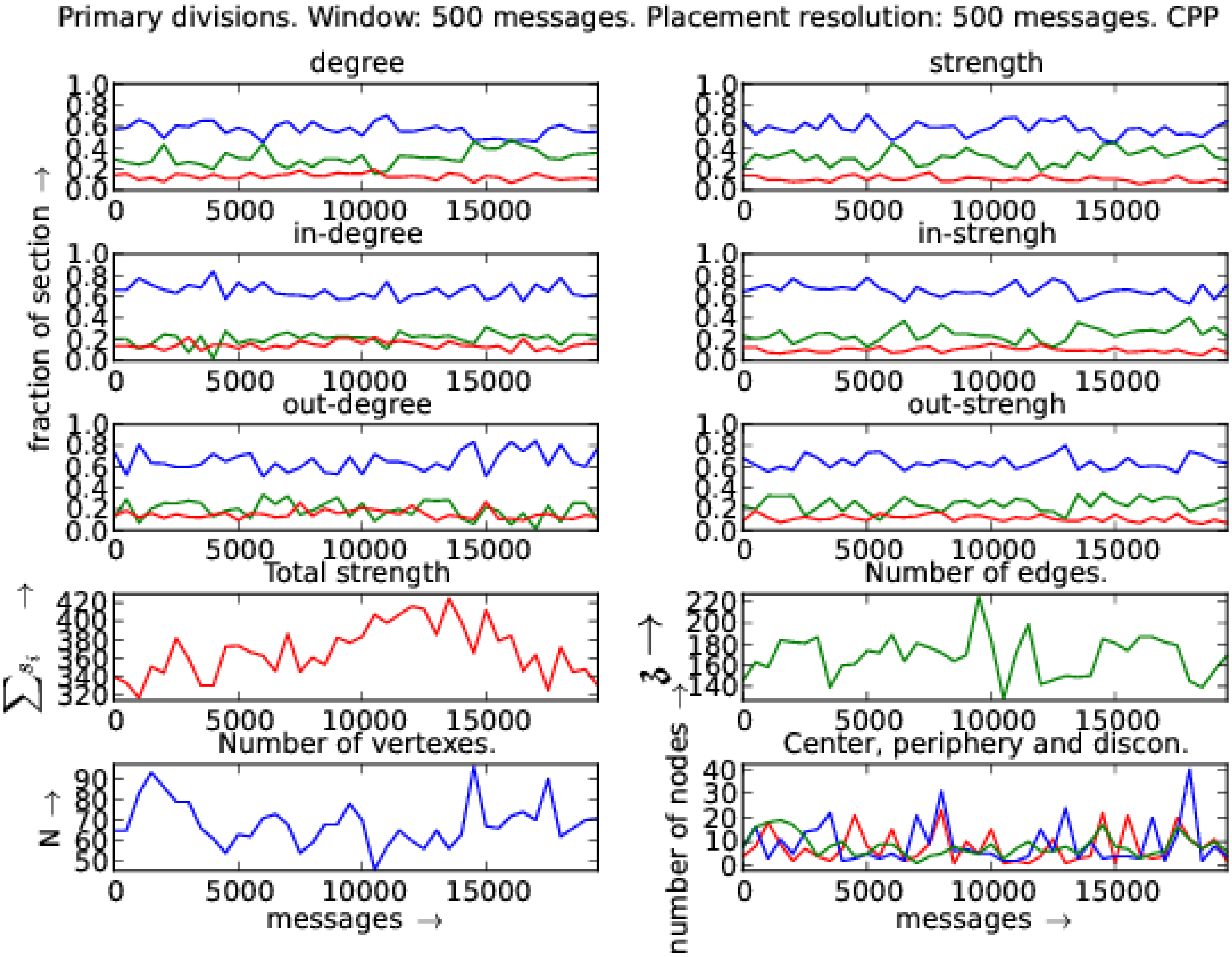

FIG. 12. Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.

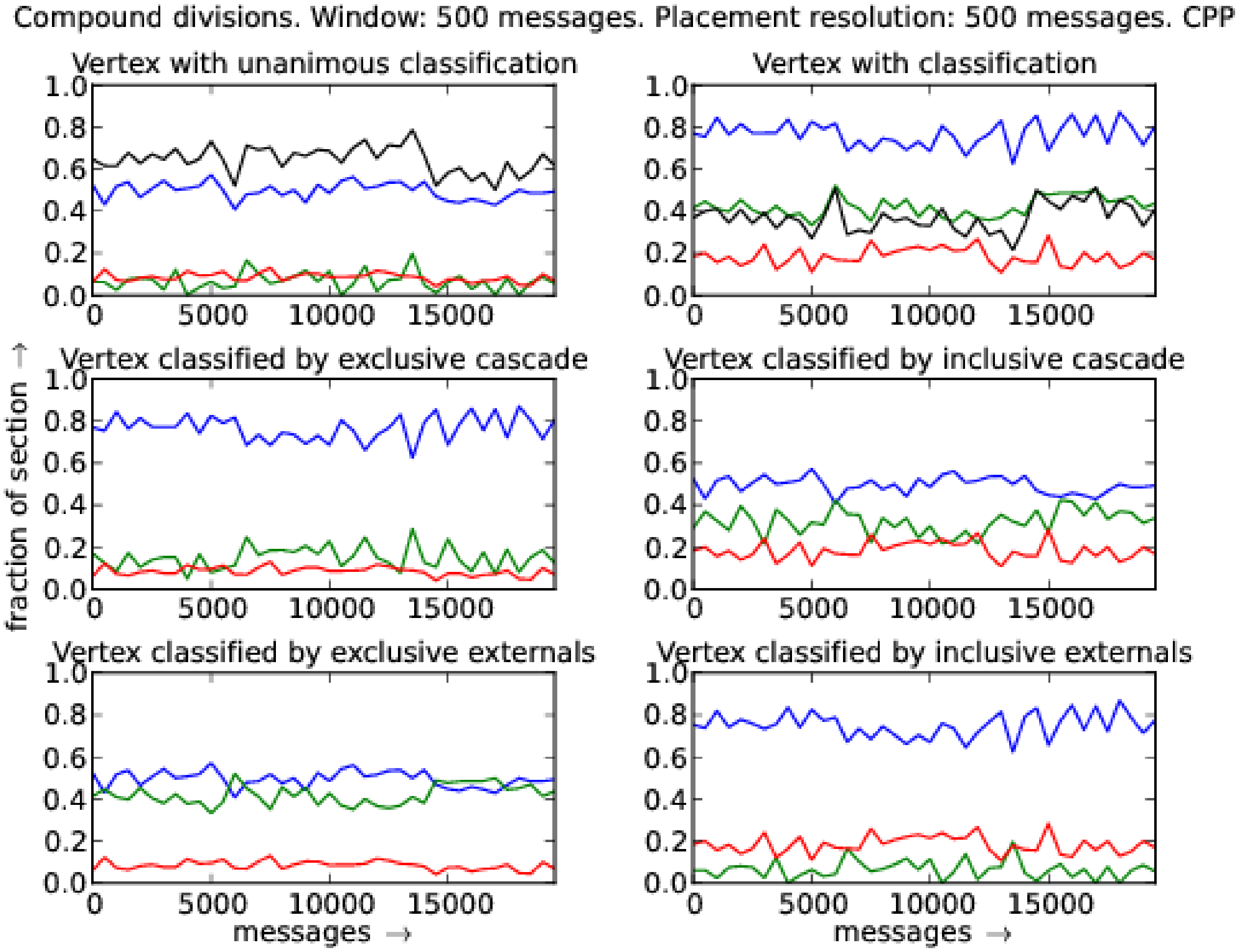

FIG. 13. Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection III B 1.

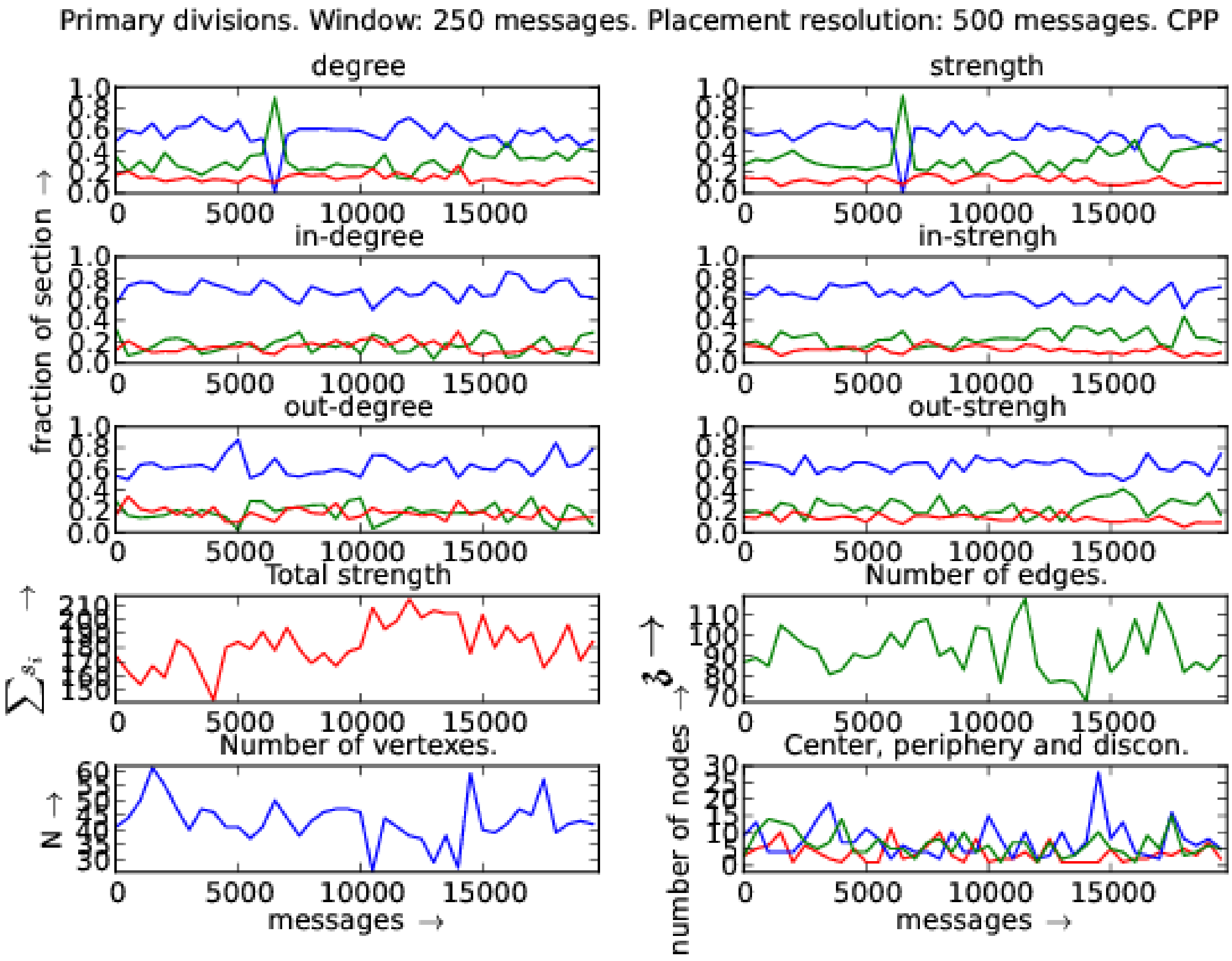

FIG. 14. Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.

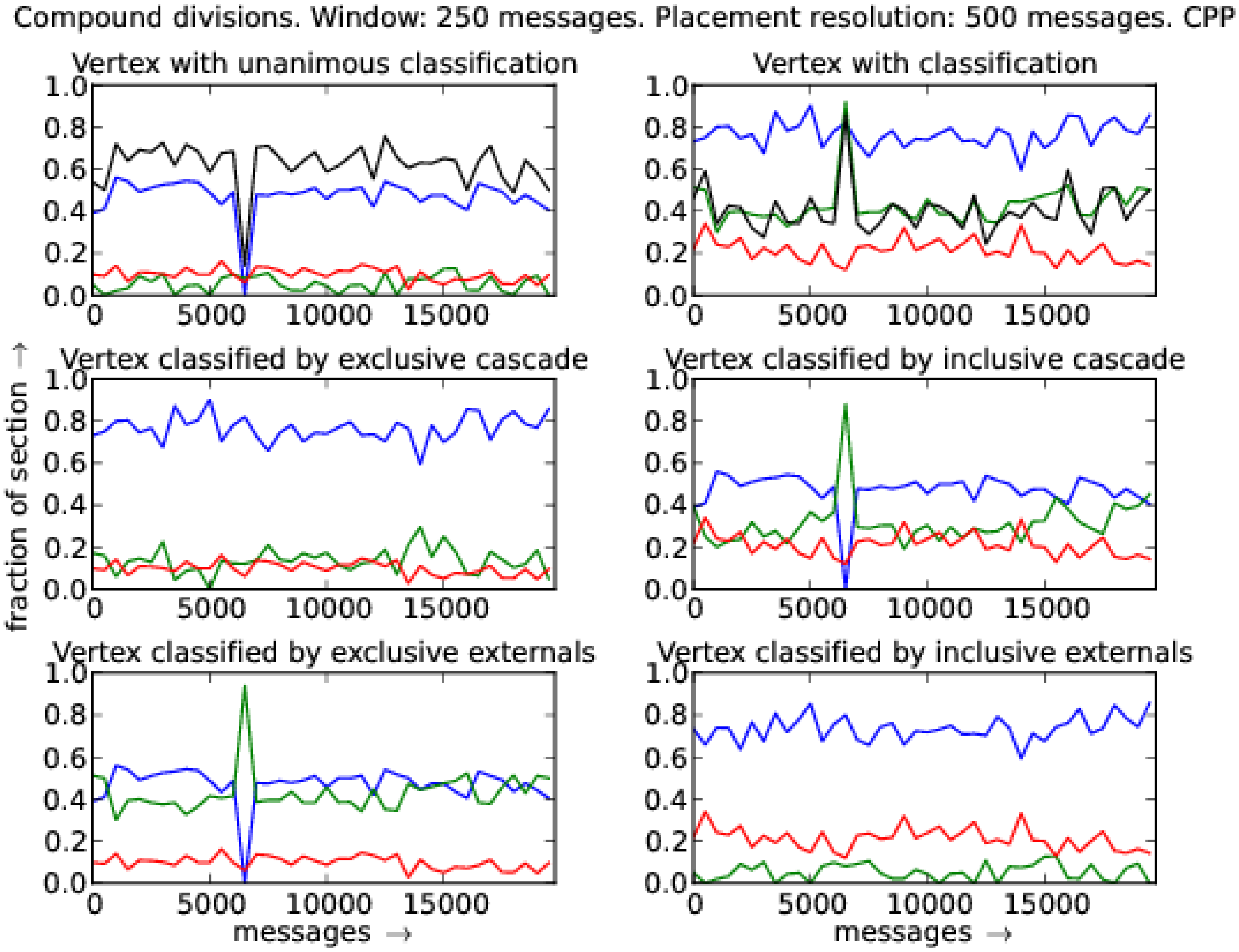

FIG. 15. Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection III B 1.

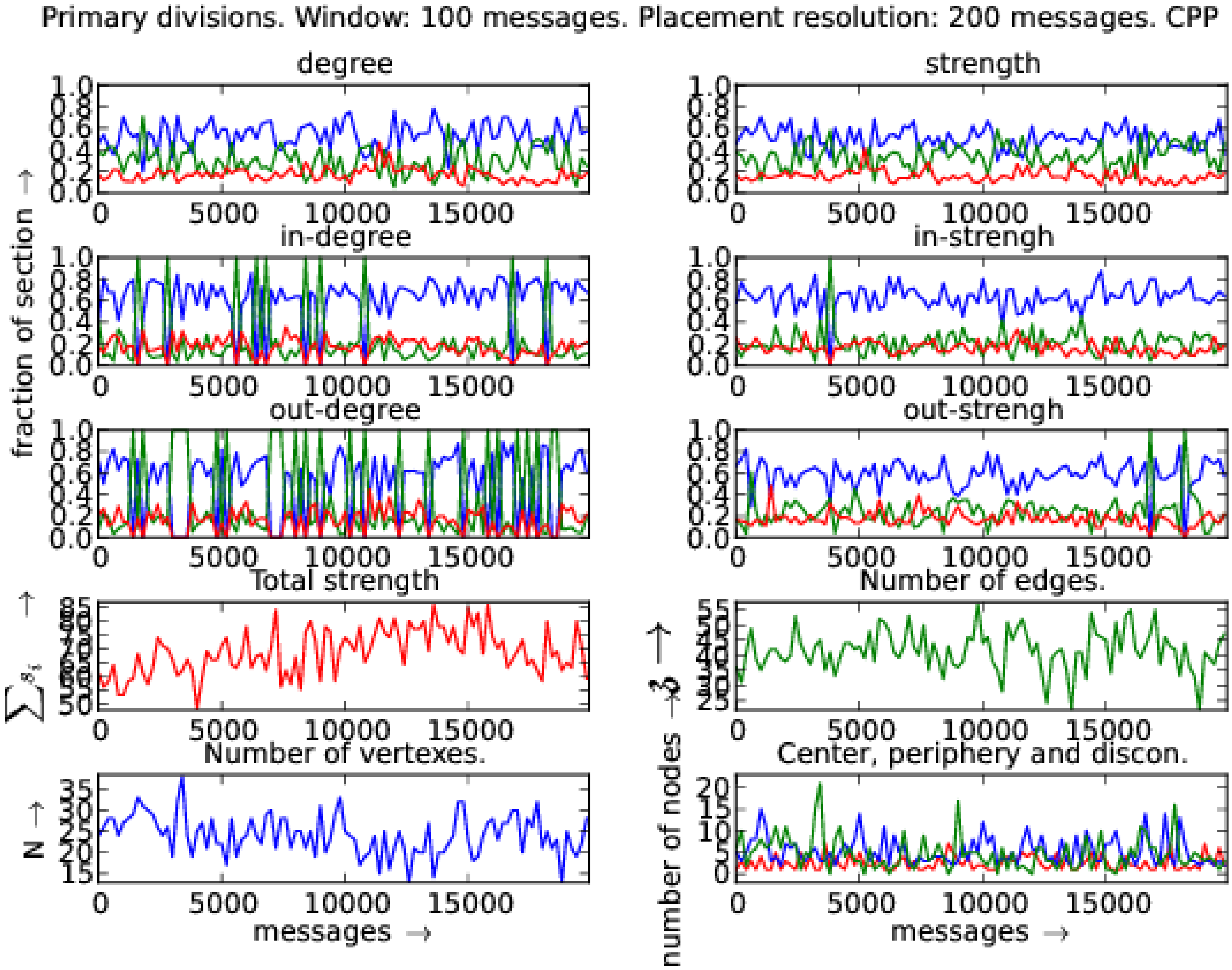

FIG. 16. Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.

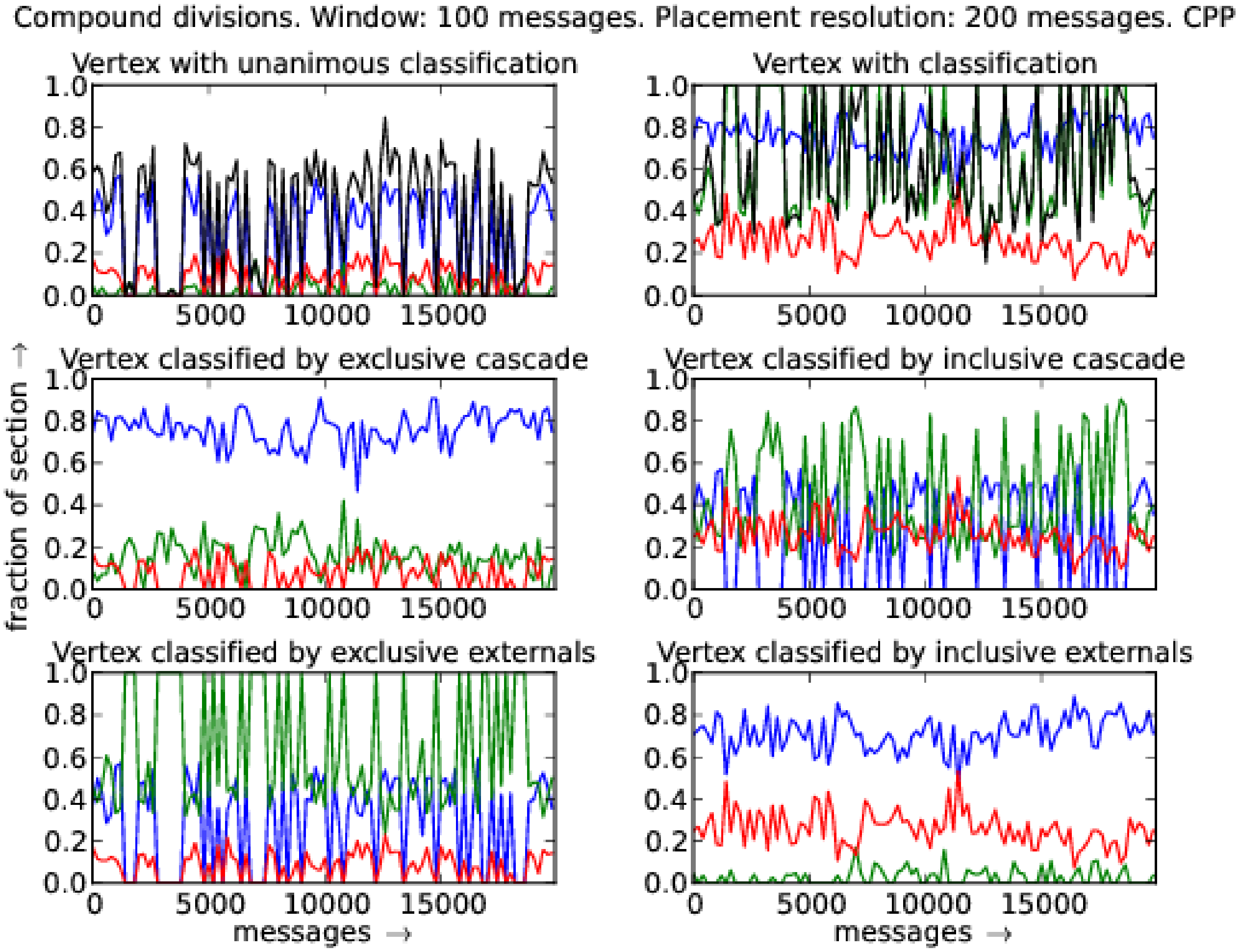

FIG. 17. Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection III B 1.

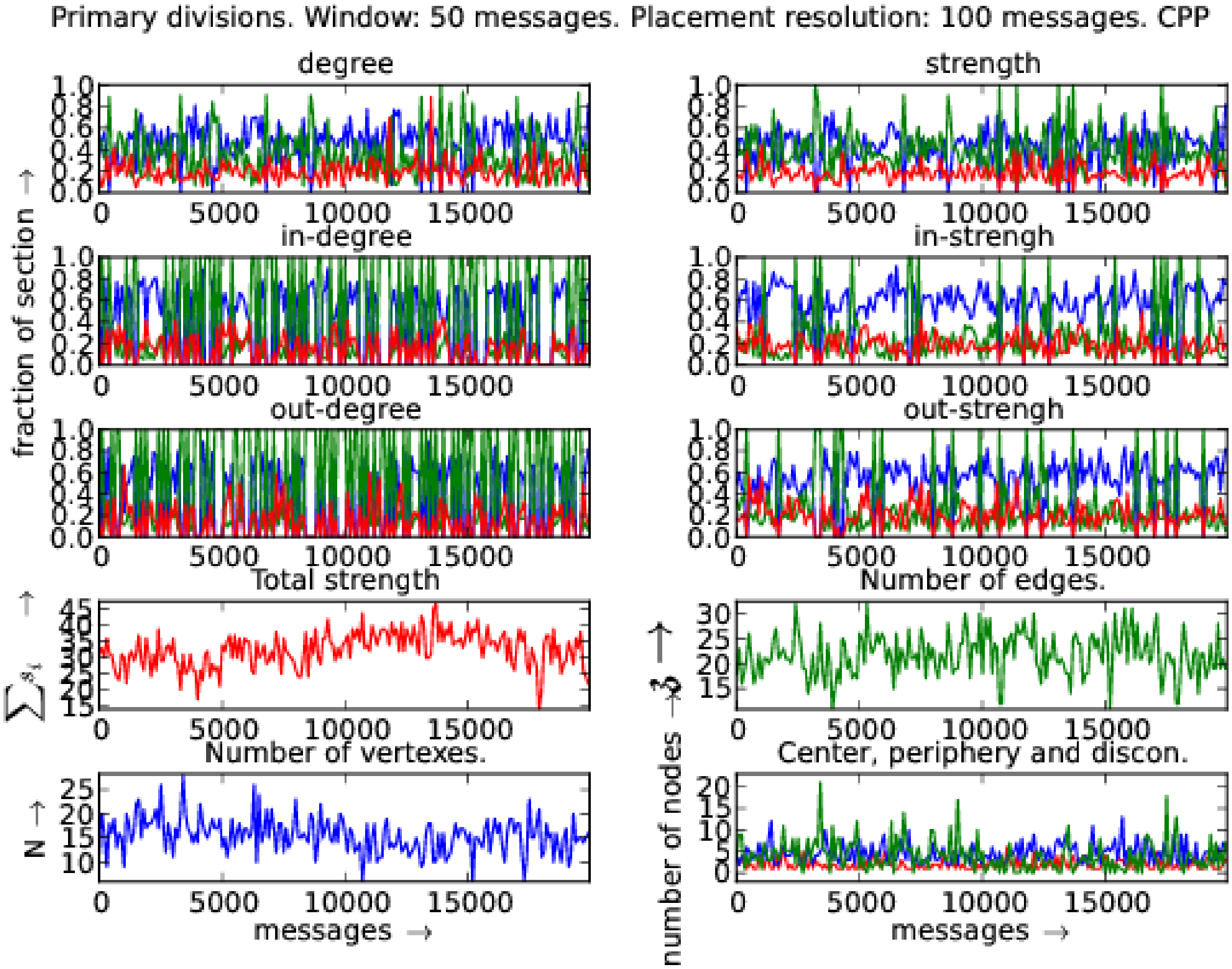

FIG. 18. Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.

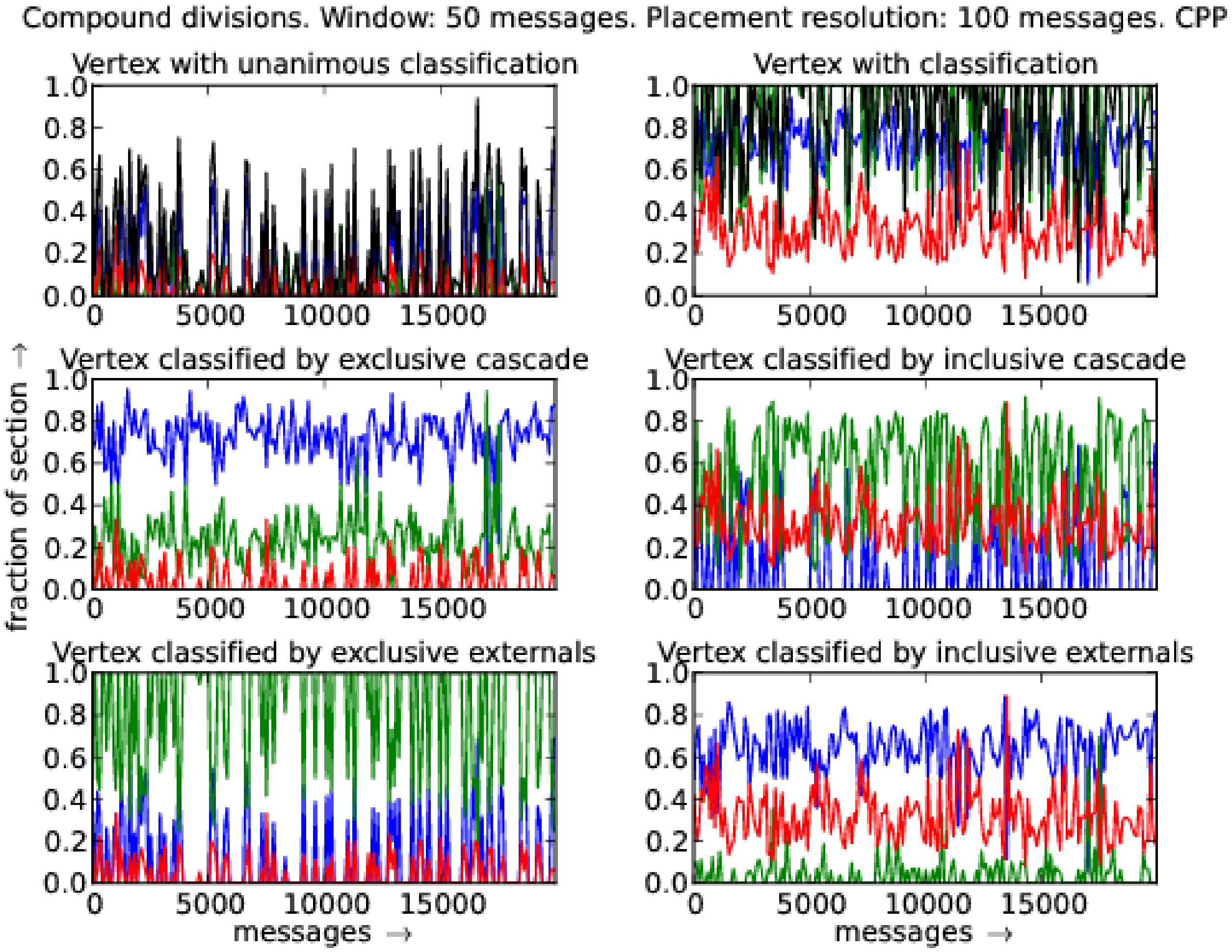

FIG. 19. Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection III B 1.

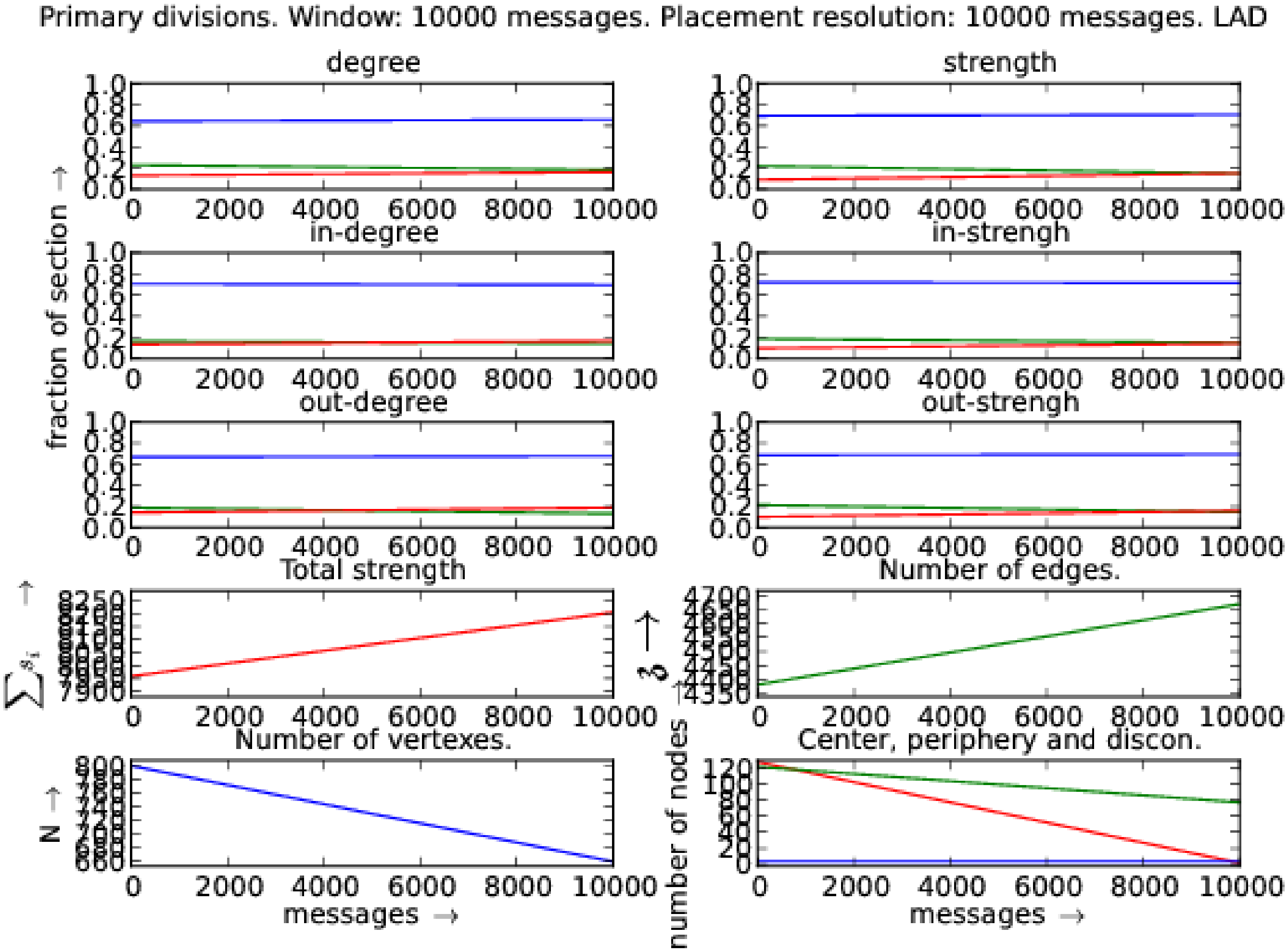

FIG. 20. Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.

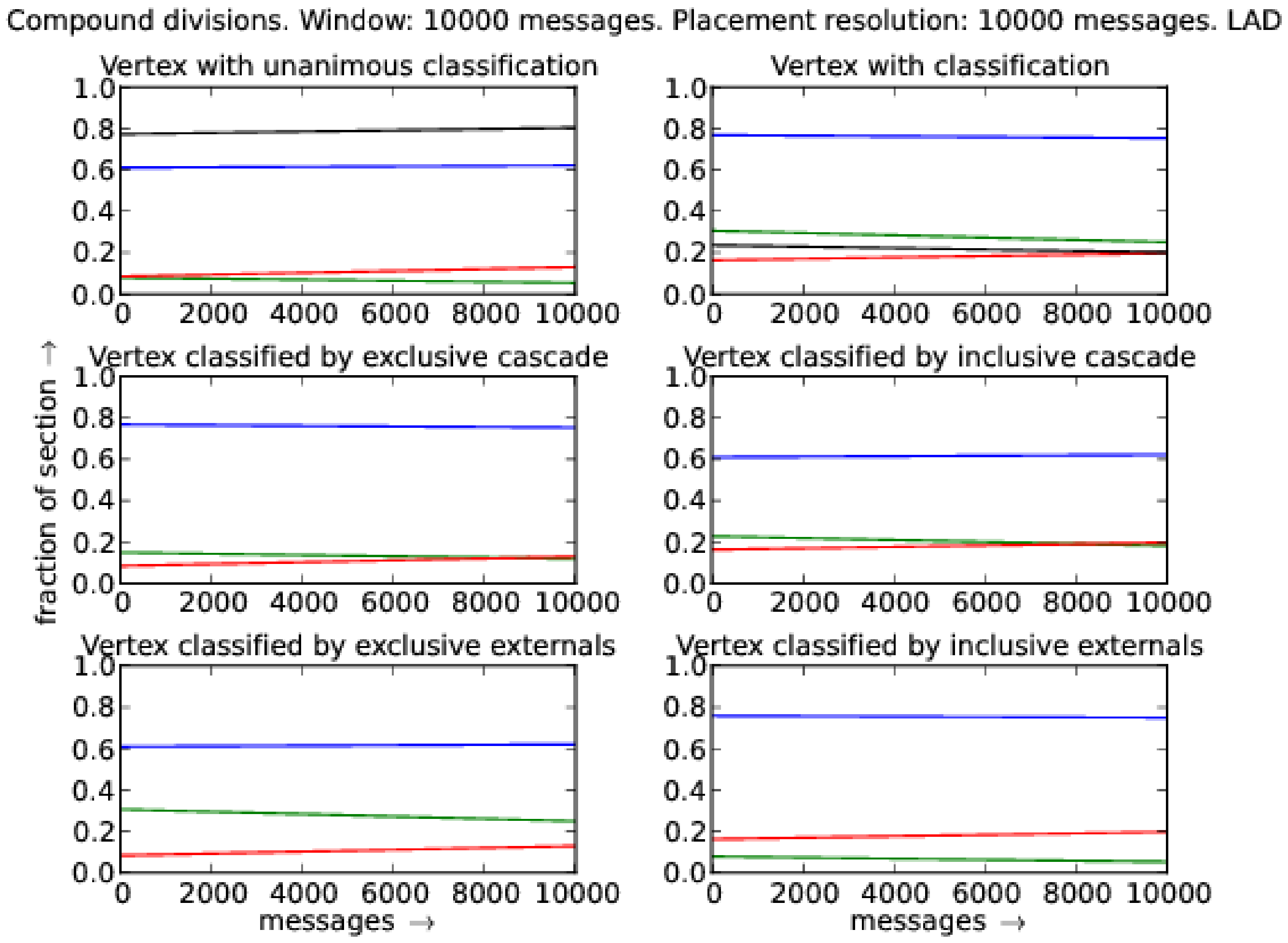

FIG. 21. Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection III B 1.

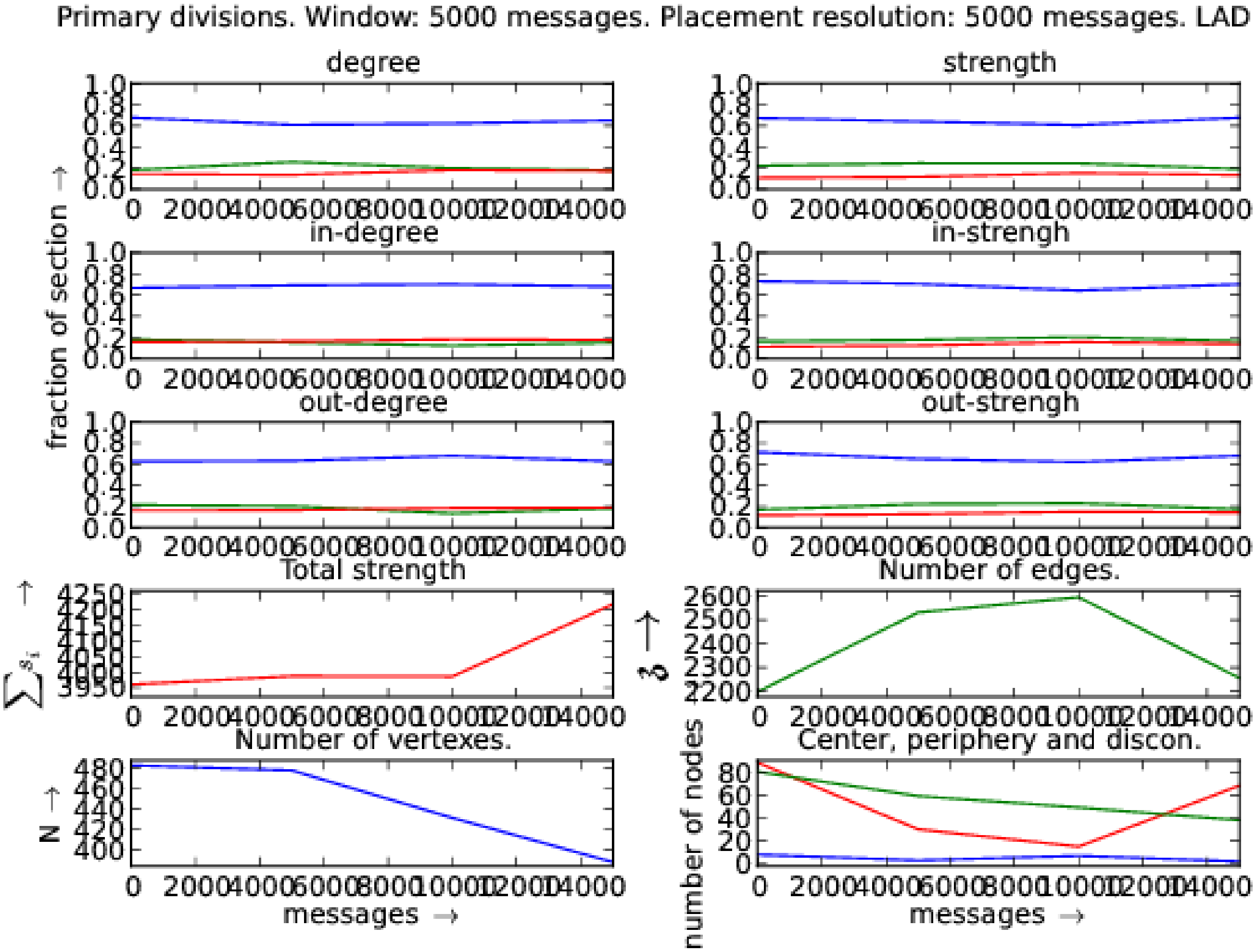

FIG. 22. Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.

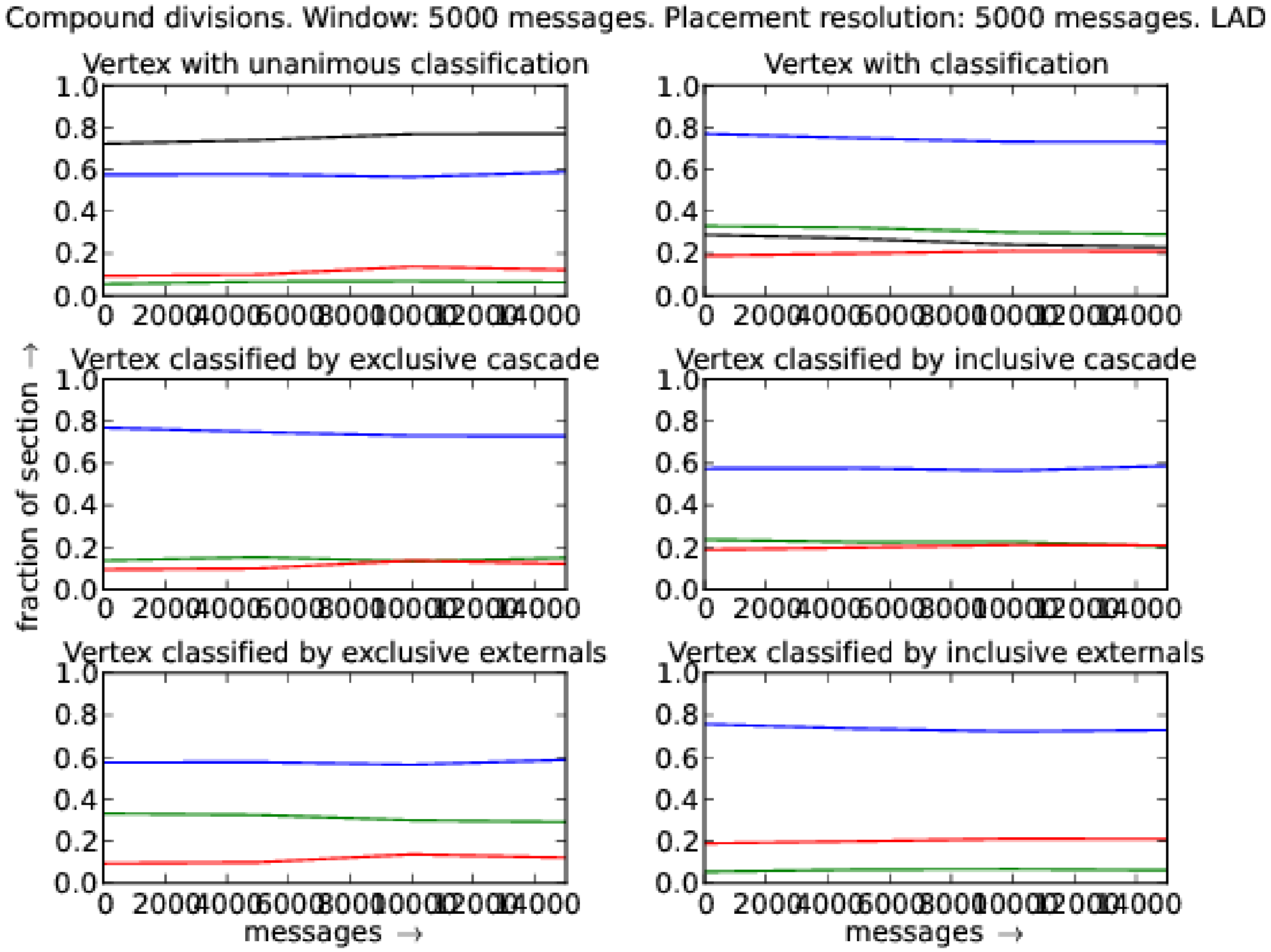

FIG. 23. Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection III B 1.

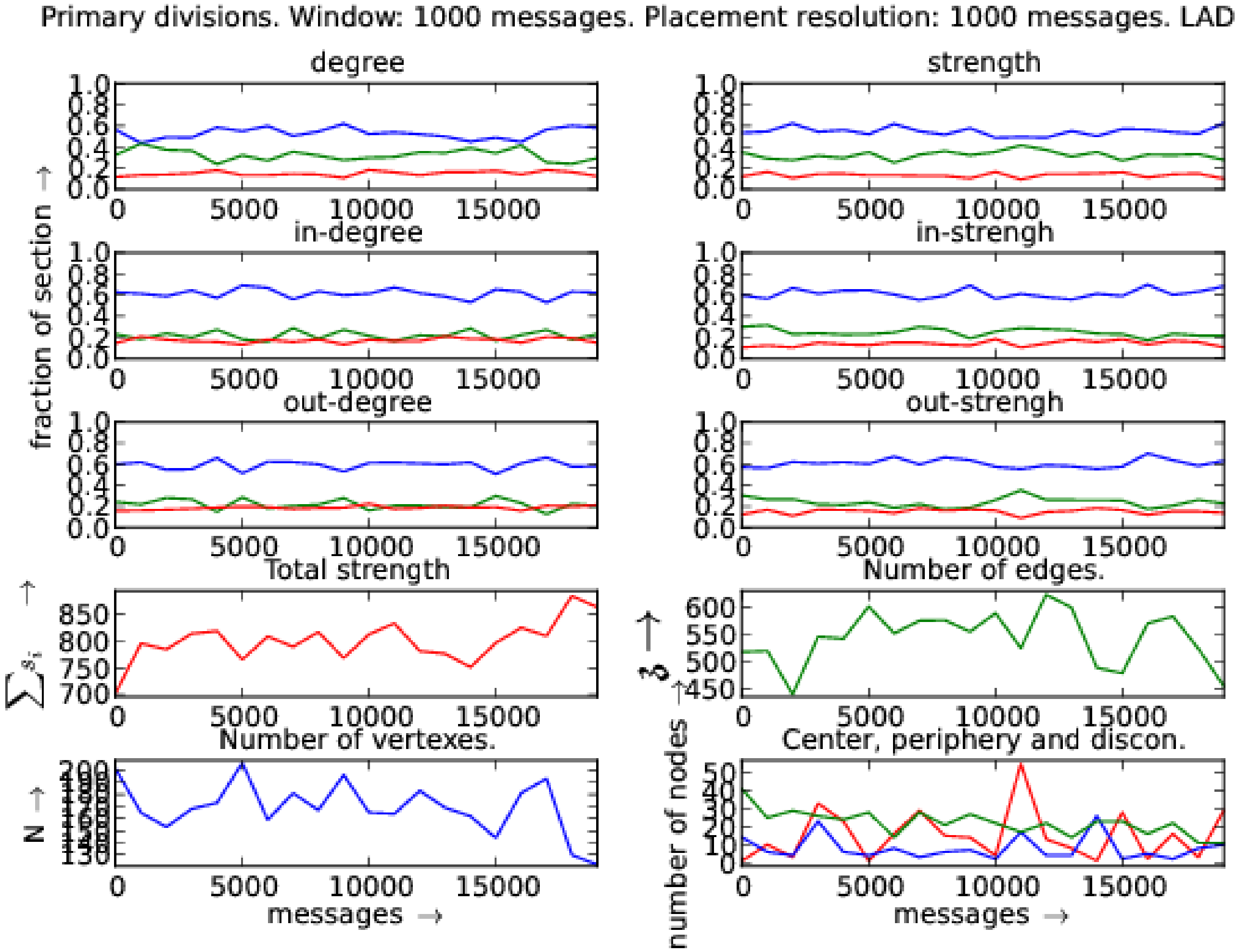

FIG. 24. Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.

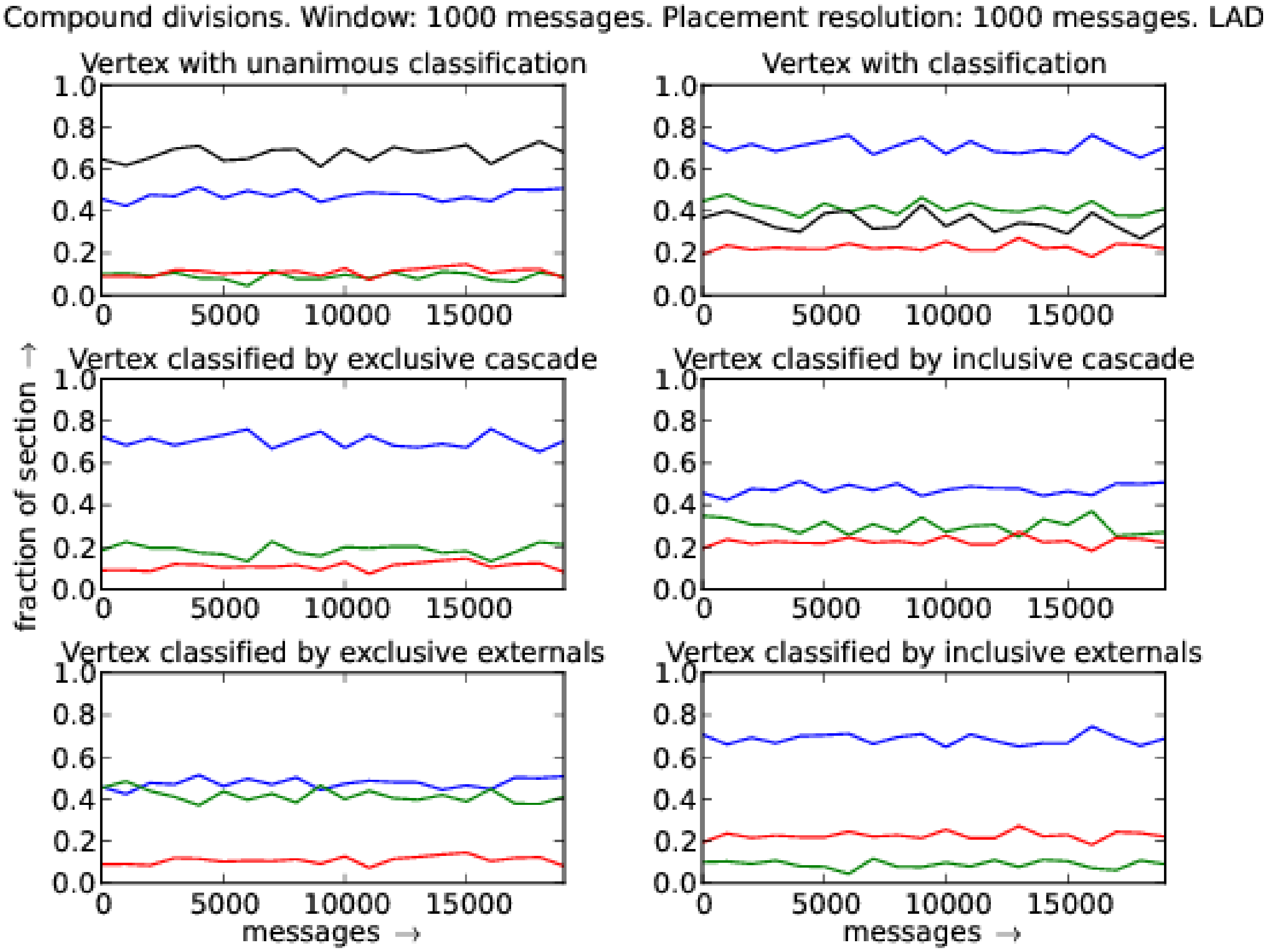

FIG. 25. Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection III B 1.

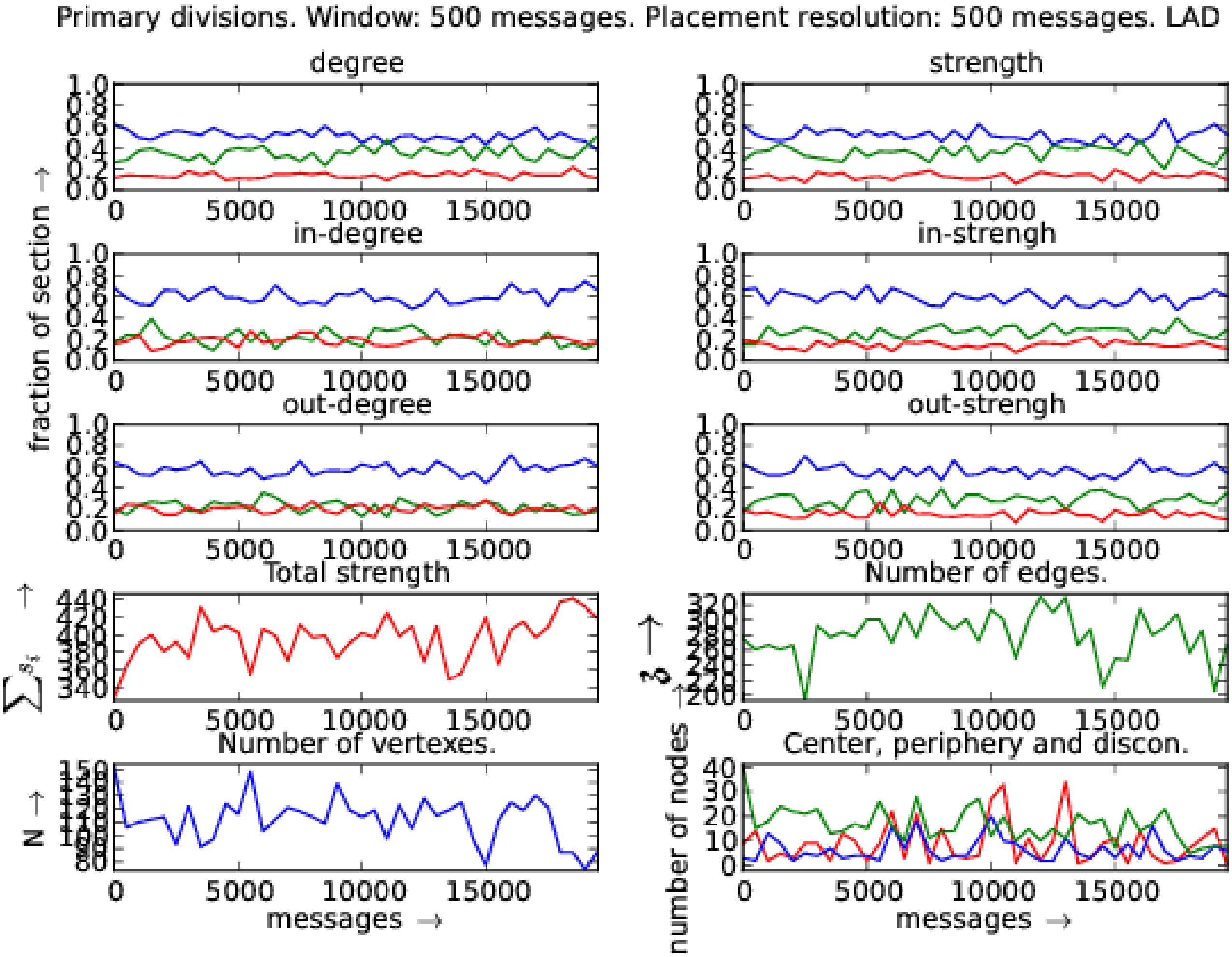

FIG. 26. Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.

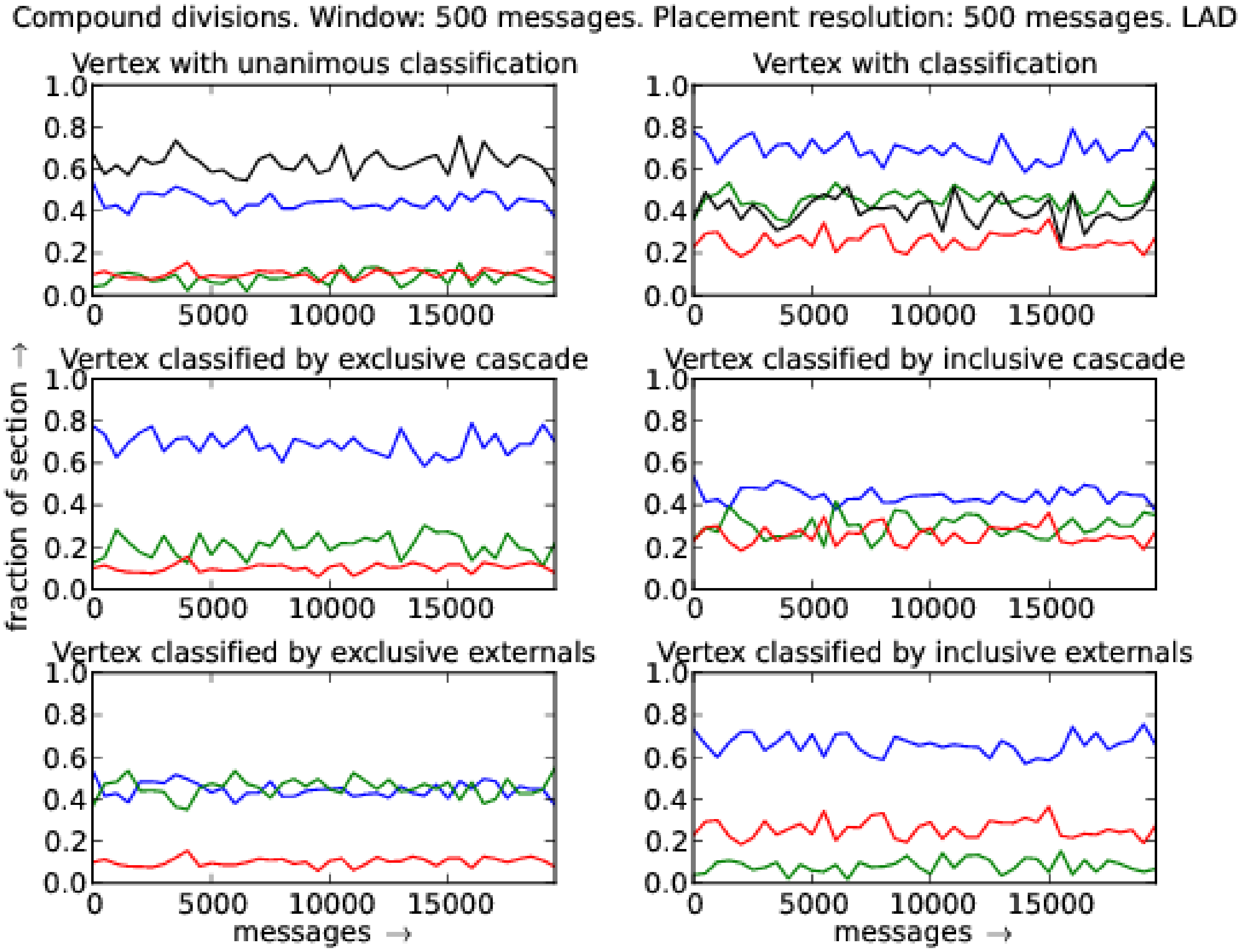

FIG. 27. Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection III B 1.

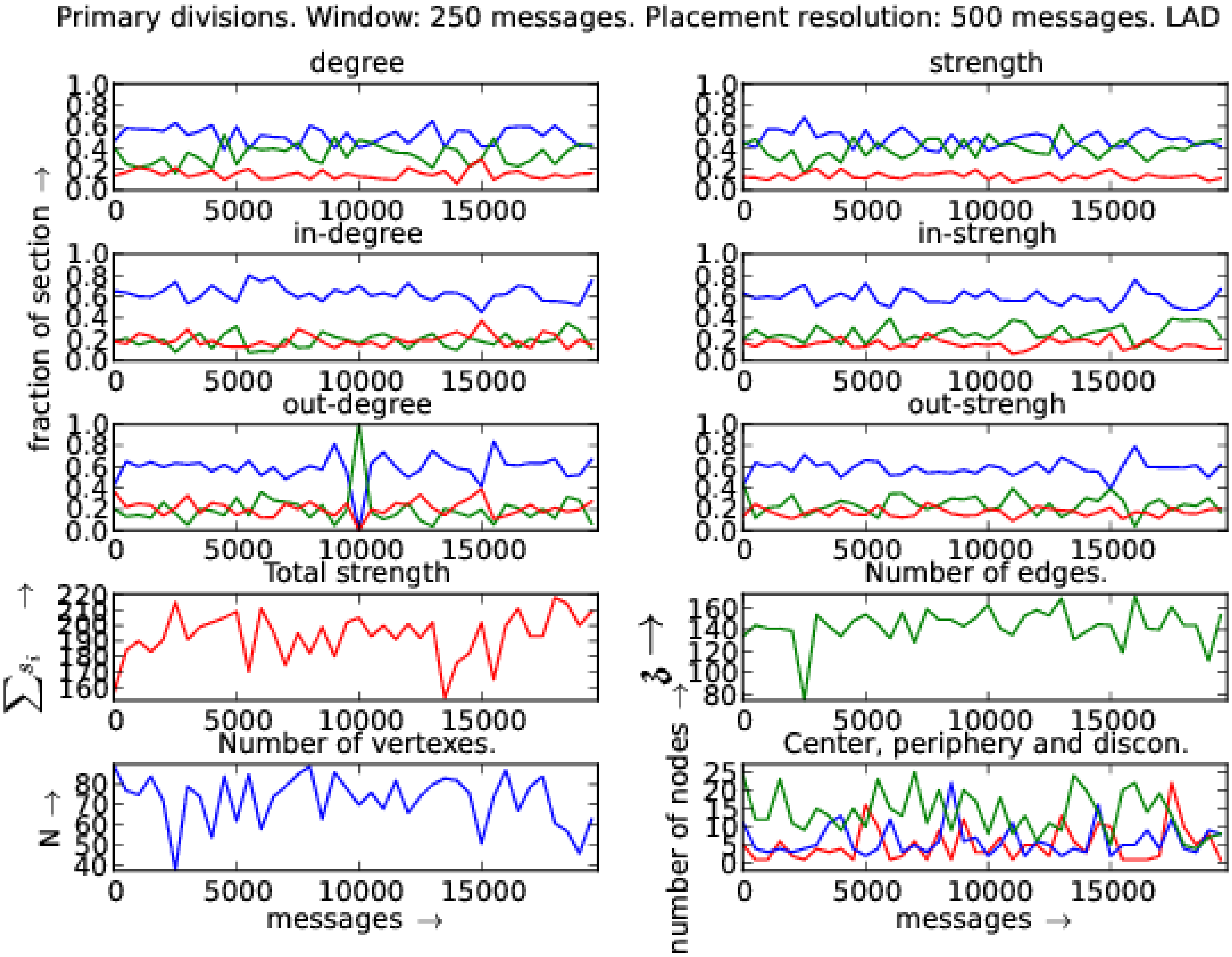

FIG. 28. Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.

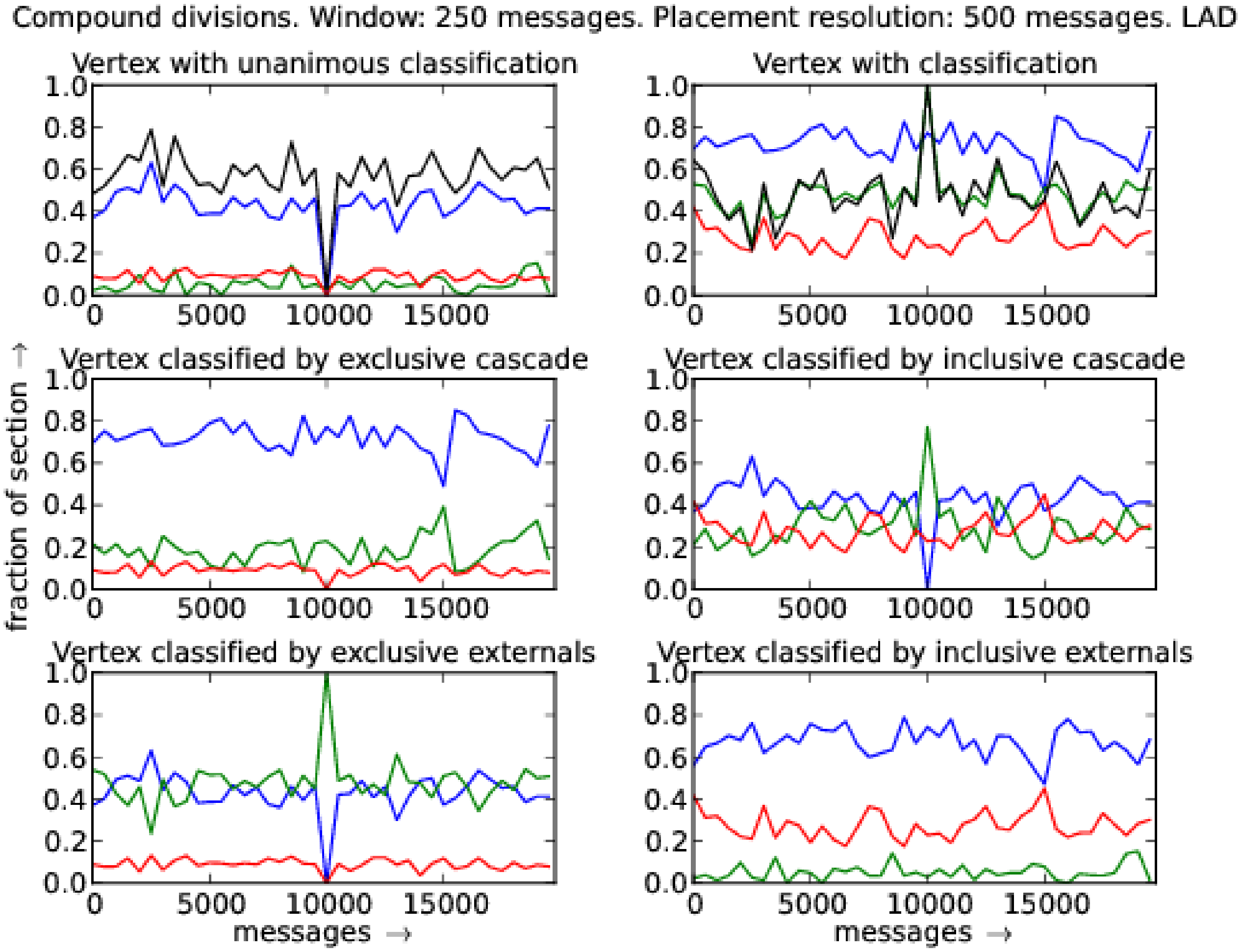

FIG. 29. Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection III B 1.

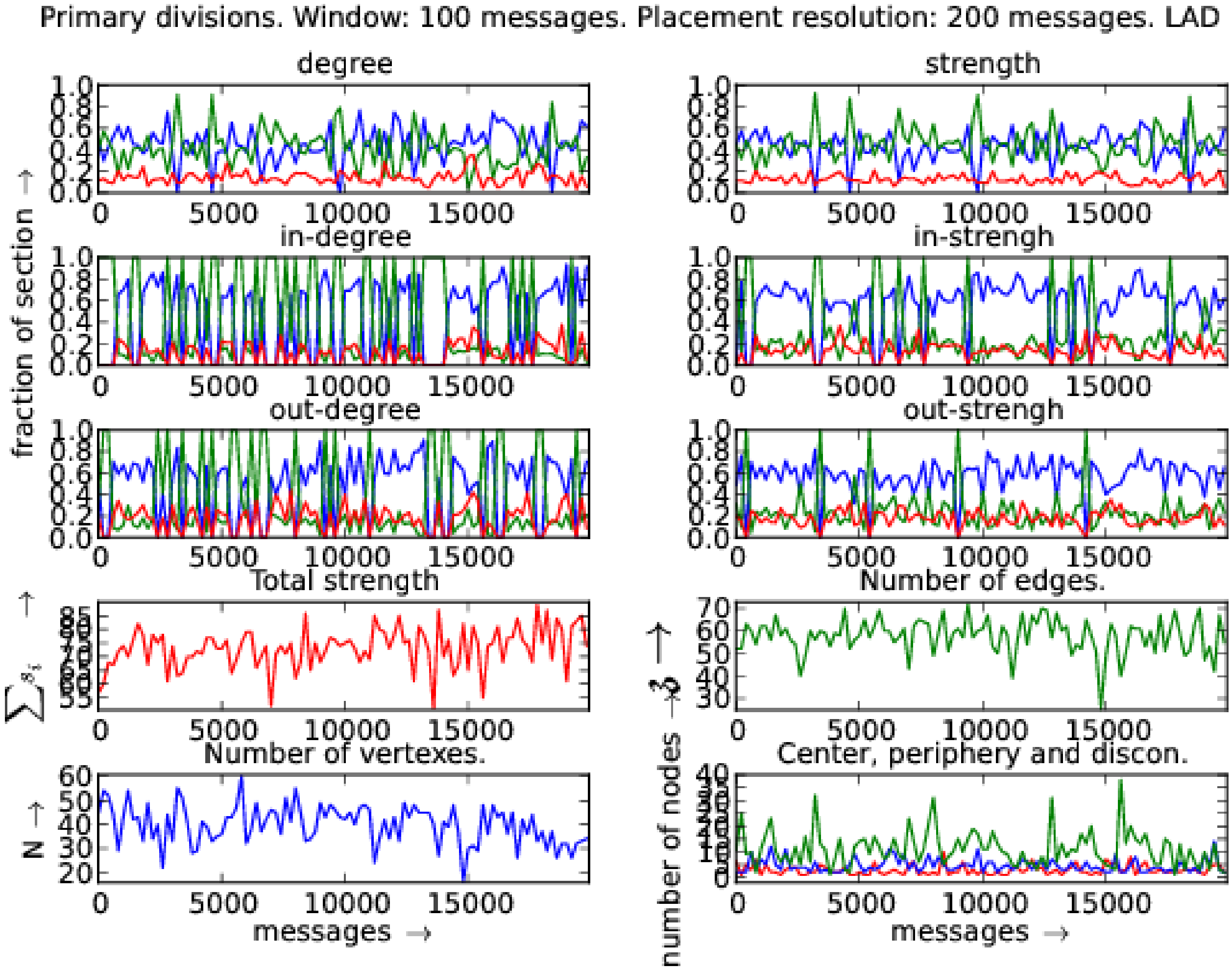

FIG. 30. Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.

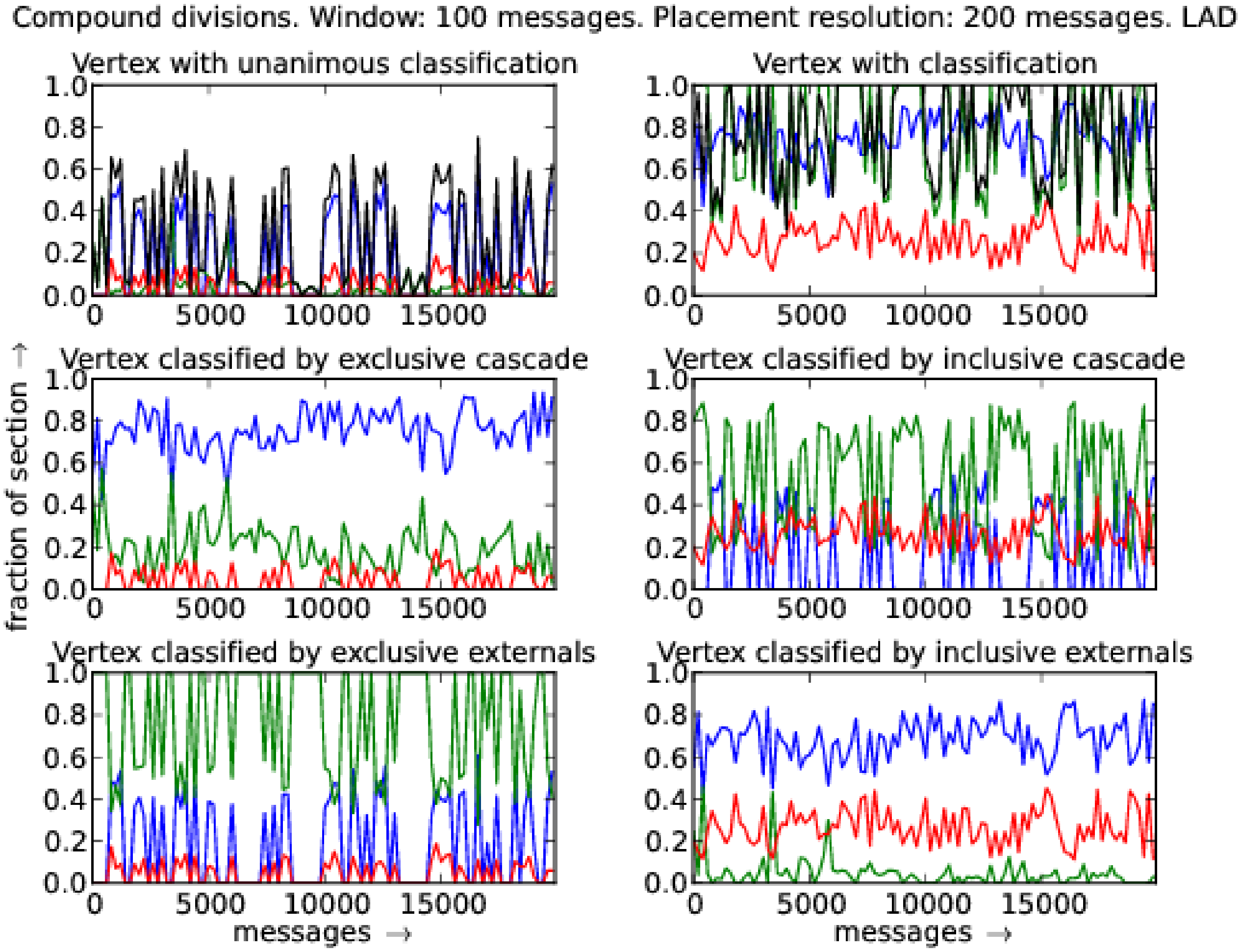

FIG. 31. Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection III B 1.

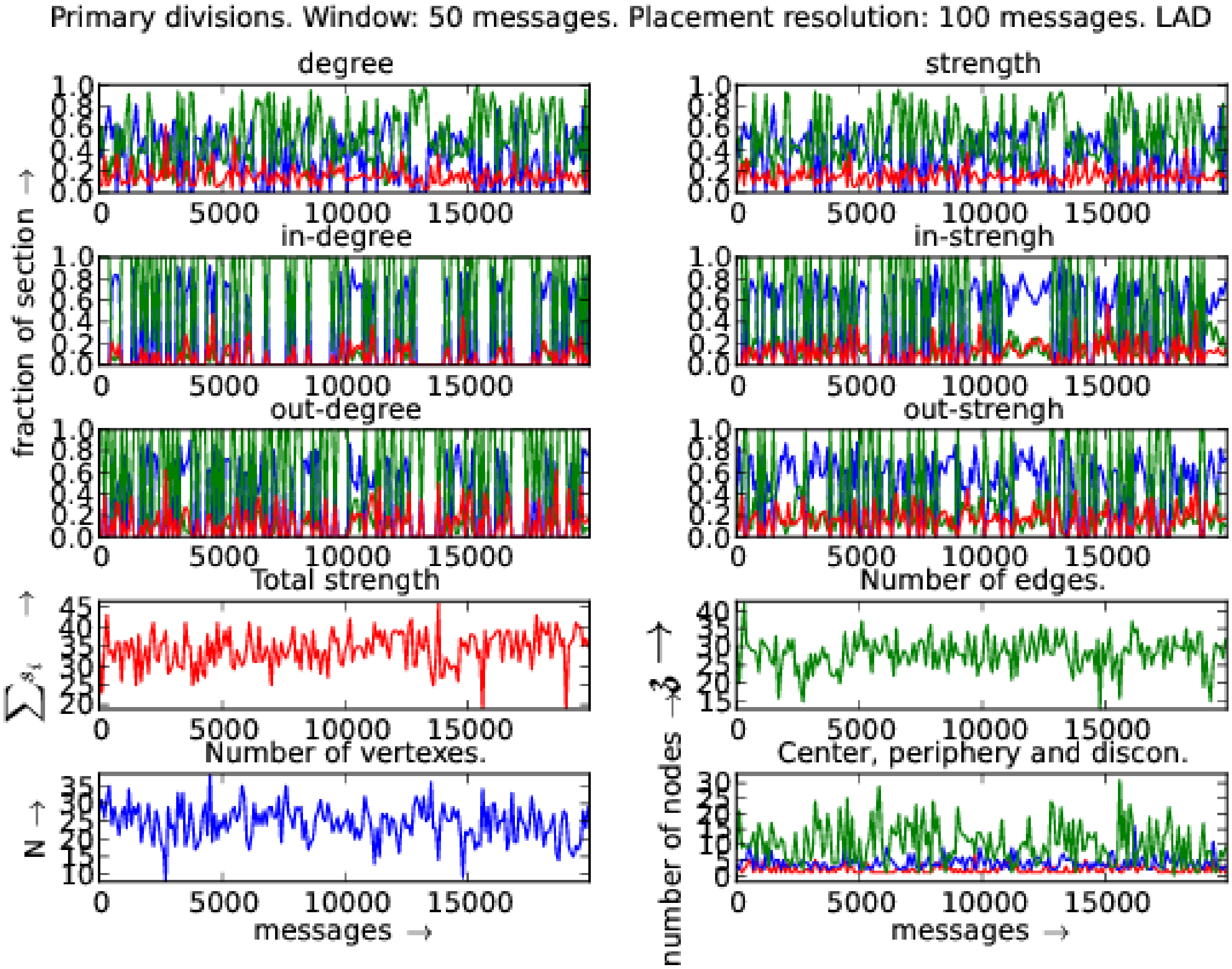

FIG. 32. Distribution of vertex with respect to each measure of activity: in and out degrees and strengths. CPP Std library official mailing list. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red is the center (maximum distance to another vertex in equal to radius), blue is periphery (maximum distance equals to diameter) of the giant component. On the same graph, green counts the disconnected vertexes.

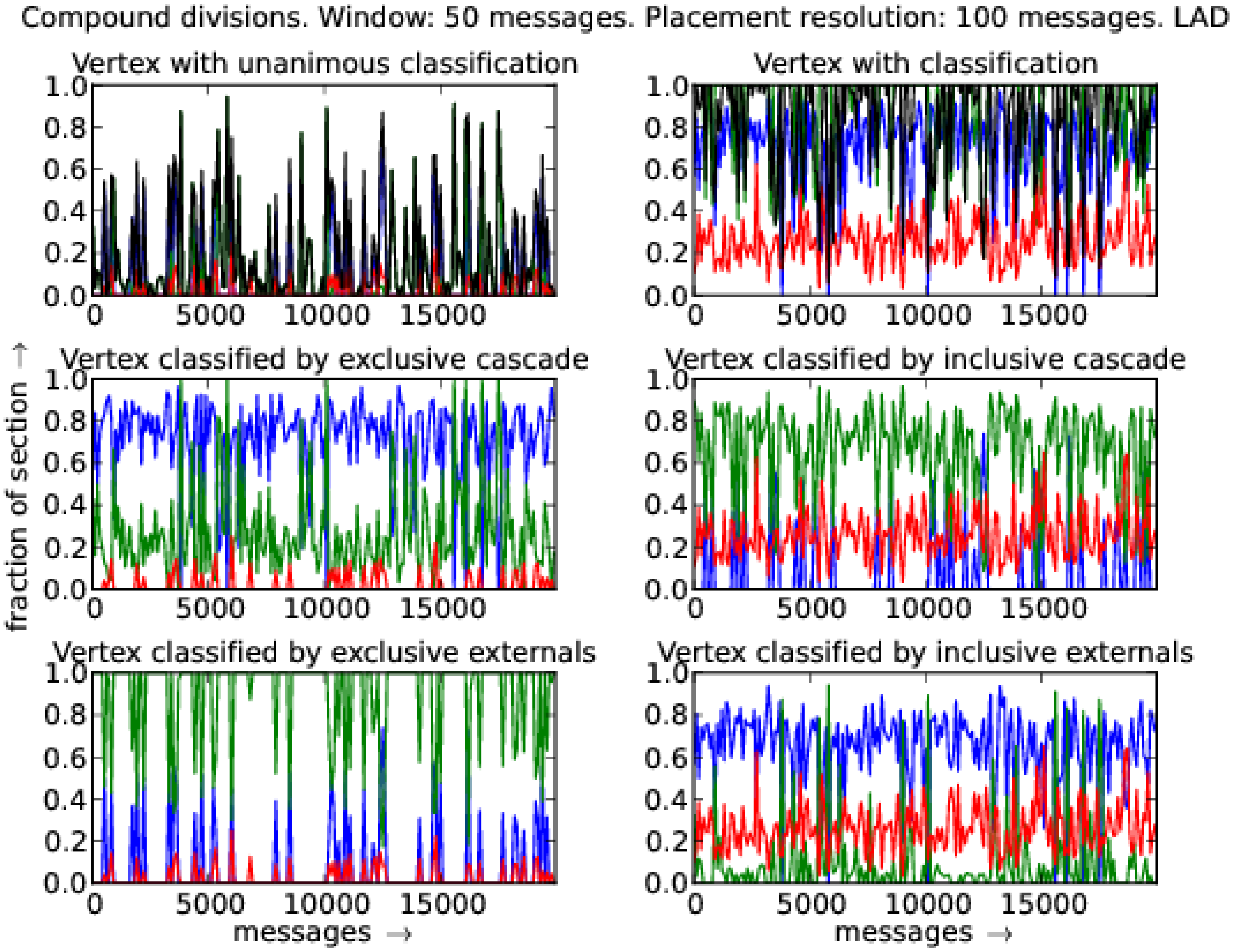

FIG. 33. Distribution of vertex with respect to each measure of activity: compound criteria. Red, green and blue are for hubs, intermediary and border (peripheral) vertex fractions. The first two plots picture classifications that are not functions. Thus, in the first plot, the fraction of vertexes with unique classification in plotted in black. On the second plot, black is used to depict how much a vertex was classified in more than section: $\frac{\text{number of classifications}-\text{number of nodes}}{\text{number of nodes}}$. Compound criteria described in subsection III B 1.