

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

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: 27 April 2015)

This is the supporting information of the article that 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.

PACS numbers: 89.75.Fb, 05.65.+b, 89.65.-s

Keywords: complex networks, social network analysis, pattern recognition, statistics, anthropological physics

These results were produced with the Gmane public domain data and an open python package designed for attaining these, and related, results. The interested reader should follow Appendix C to access both data and routines. Inline are results for 4 emails lists: LAD, LAU, MET and CPP, as described in Section II. Similar results can be reproduced for any number of (Gmane) email lists. To avoid repeating text of each table for each list, the text is given inline.

I. TIME TABLES IN DIFFERENT SCALES

Theory presented in Section III A and results exposed in Section IV A of the paper¹.

A. Time circular measures

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 $\frac{\max(incidence)}{\min(incidence)}$. Also, $\mu_{\frac{\max(incidence')}{\min(incidence')}}$ and $\sigma_{\frac{\max(incidence')}{\min(incidence')}}$ are given for 1000 uniform distribution simulations within the same number of bins and with the same number of samples. Section III A describes the theoretical background of directional (or circular) statistics.

^{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 Ceará (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)

TABLE I. LAU circular measures

scale	θ'_μ	$S(z)$	$Var(z)$	$\delta(z)$	$\frac{\max(incidence)}{\min(incidence)}$	$\mu_{\frac{\max(incidence')}{\min(incidence')}}$	$\sigma_{\frac{\max(incidence')}{\min(incidence')}}$
seconds	-//-	3.31	1.00	29337.65	1.27	1.29	0.04
minutes	-//-	3.13	0.99	8879.19	1.32	1.29	0.04
hours	-8.76	1.56	0.71	4.92	8.38	1.14	0.03
weekdays	-0.21	2.14	0.90	45.41	1.62	1.05	0.02
month days	-0.00	2.76	0.98	1021.59	2.11	1.17	0.03
months	3.55	2.30	0.93	94.53	1.57	1.09	0.02

TABLE II. LAD circular measures

scale	θ'_μ	$S(z)$	$Var(z)$	$\delta(z)$	$\frac{\max(incidence)}{\min(incidence)}$	$\mu_{\frac{\max(incidence')}{\min(incidence')}}$	$\sigma_{\frac{\max(incidence')}{\min(incidence')}}$
seconds	-//-	3.13	0.99	9070.17	1.28	1.29	0.05
minutes	-//-	3.60	1.00	205489.40	1.22	1.29	0.05
hours	-9.61	1.52	0.68	4.36	9.77	1.15	0.03
weekdays	-0.03	2.03	0.87	29.28	1.72	1.05	0.02
month days	-0.07	2.94	0.99	2754.16	2.21	1.17	0.03
months	-0.56	2.14	0.90	44.00	2.25	1.09	0.02

TABLE III. MET circular measures

scale	θ'_μ	$S(z)$	$Var(z)$	$\delta(z)$	$\frac{\max(incidence)}{\min(incidence)}$	$\mu_{\frac{\max(incidence')}{\min(incidence')}}$	$\sigma_{\frac{\max(incidence')}{\min(incidence')}}$
seconds	-//-	3.06	0.99	5910.47	1.27	1.29	0.05
minutes	-//-	3.14	0.99	9696.29	1.34	1.29	0.05
hours	-9.20	1.35	0.60	2.76	19.26	1.15	0.03
weekdays	-0.27	1.86	0.82	13.82	2.89	1.05	0.02
month days	0.14	2.49	0.95	237.27	1.93	1.17	0.03
months	-2.92	1.73	0.78	9.20	3.04	1.09	0.02

TABLE IV. CPP circular measures

scale	θ'_μ	$S(z)$	$Var(z)$	$\delta(z)$	$\frac{\max(incidence)}{\min(incidence)}$	$\mu_{\frac{\max(incidence')}{\min(incidence')}}$	$\sigma_{\frac{\max(incidence')}{\min(incidence')}}$
seconds	-//-	3.31	1.00	28205.46	1.26	1.29	0.04
minutes	-//-	3.18	0.99	12275.59	1.27	1.29	0.04
hours	-9.39	1.48	0.67	3.91	11.18	1.15	0.03
weekdays	-0.17	1.83	0.81	12.66	2.59	1.05	0.02
month days	-0.31	3.14	0.99	9541.67	1.93	1.17	0.03
months	0.15	2.34	0.93	115.49	1.50	1.09	0.02

B. Time histograms**C. Histograms of activity along the hours of the day**

Activity percentages along the 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.

TABLE V. LAU activity along the hours of the day

	1h	2h	3h	4h	6h	12h
0h	3.58					
1h	2.22	5.80	7.43	8.49	10.14	
2h	1.63	2.69				
3h	1.06					
4h	0.84	1.66	2.72			
5h	0.82			5.20		36.88
6h	1.17	3.54				
7h	2.37		7.07			
8h	3.53	9.57			26.74	
9h	6.04			23.20		
10h	6.83		19.67			
11h	6.79	13.62				
12h	6.11	12.36	18.75	24.68	35.66	
13h	6.26					
14h	6.38	12.31				
15h	5.93		16.91			
16h	5.52	10.98		20.73		63.12
17h	5.46					
18h	5.23	9.75				
19h	4.52		14.30		27.46	
20h	4.55	8.97				
21h	4.42			17.71		
22h	4.51		13.16			
23h	4.23	8.74				

TABLE VI. LAD activity along the hours of the day

	1h	2h	3h	4h	6h	12h
0h	4.01					
1h	2.52	6.53	8.32	9.37	10.78	
2h	1.79	2.84				
3h	1.06					
4h	0.75	1.40	2.46			
5h	0.66			3.81		33.11
6h	0.85	2.41				
7h	1.56		5.36			
8h	2.95	7.61			22.33	
9h	4.66			19.93		
10h	5.92		16.98			
11h	6.40	12.32				
12h	6.41	12.53	18.85	24.82	37.24	
13h	6.12					
14h	6.32	12.29				
15h	5.97		18.39			
16h	6.40	12.42		23.44		66.89
17h	6.02					
18h	5.99	11.02		15.65		
19h	5.03				29.65	
20h	4.63	9.22				
21h	4.59			18.63		
22h	4.88		14.00			
23h	4.53	9.41				

TABLE VII. MET activity along the hours of the day

	1h	2h	3h	4h	6h	12h
0h	2.87	4.64	5.67	6.31	7.15	
1h	1.77					
2h	1.04	1.67				
3h	0.64					
4h	0.47	0.85	1.48			
5h	0.38			2.89		29.33
6h	0.72	2.04				
7h	1.33		4.71			
8h	2.67	7.07			22.18	
9h	4.40			20.14		
10h	6.29	13.07	17.47			
11h	6.78					
12h	7.33	14.41	21.50	28.65	42.22	
13h	7.08					
14h	7.09	14.24				
15h	7.14		20.72			
16h	6.68	13.58		24.79		70.67
17h	6.89					
18h	5.99	11.22	16.19			
19h	5.23				28.44	
20h	4.98	9.34				
21h	4.37			17.22		
22h	4.24	7.88	12.25			
23h	3.64					

TABLE VIII. CPP activity along the hours of the day

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

D. Histograms of activity along the days of the week

Activity percentages along the days of the week. Higher activity was observed during weekdays, with a decrease of activity on weekends of at least one third and two thirds in extreme cases.

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

E. Histograms of activity along the days of the month

No significant variation of activity in the days along the month was observed. 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.

TABLE IX. LAU activity along the days of the month.

	1 day	5	10	15 days
1	5.49	18.50	35.66	52.08
2	3.48			
3	3.41			
4	2.91			
5	3.22	17.16	32.99	47.92
6	3.13			
7	3.97			
8	3.55			
9	3.43	16.42	31.35	16.37
10	3.09			
11	3.23			
12	3.39			
13	3.59	14.98	16.80	32.01
14	3.37			
15	2.84			
16	3.07			
17	3.09	16.57	15.21	32.71
18	3.35			
19	3.85			
20	3.21			
21	3.64	15.90	16.81	52.09
22	3.20			
23	2.61			
24	2.66			
25	2.88	17.75	17.53	35.28
26	2.92			
27	2.87			
28	3.66			
29	3.65	3.00	3.00	3.00
30	3.29			

TABLE X. LAD activity along the days of the month.

	1 day	5	10	15 days
1	5.39	17.53	35.28	52.09
2	2.87			
3	2.89			
4	3.53			
5	2.85	17.75	32.71	47.91
6	3.36			
7	3.03			
8	3.77			
9	4.04	16.81	15.90	32.01
10	3.55			
11	3.00			
12	3.13			
13	3.81	15.21	16.80	52.08
14	3.46			
15	3.40			
16	2.84			
17	2.95	17.53	17.75	35.28
18	3.51			
19	3.42			
20	3.17			
21	2.49	16.80	16.81	52.09
22	3.51			
23	3.18			
24	2.97			
25	3.06	17.75	17.53	35.28
26	3.84			
27	3.85			
28	3.37			
29	3.30	3.00	3.00	3.00
30	2.44			

TABLE XI. MET activity along the days of the month.

	1 day	5	10	15 days
1	5.31	20.06	37.31	52.41
2	3.54			
3	3.80			
4	3.66			
5	3.74	17.25	37.31	52.41
6	3.84			
7	3.24			
8	3.46			
9	3.42	15.09	30.63	47.59
10	3.29			
11	3.25			
12	3.07			
13	3.11	15.53	30.63	47.59
14	2.88			
15	2.77			
16	3.15			
17	3.16	15.52	32.06	47.59
18	3.53			
19	2.76			
20	2.93			
21	3.39	16.54	32.06	47.59
22	2.86			
23	3.52			
24	2.76			
25	2.98	16.54	32.06	47.59
26	3.09			
27	3.06			
28	3.54			
29	3.74	16.54	32.06	47.59
30	3.10			

TABLE XII. CPP activity along the days of the month.

	1 day	5	10	15 days
1	5.00	18.21	33.84	51.29
2	3.02			
3	3.65			
4	3.18			
5	3.37	15.63	33.84	51.29
6	3.22			
7	3.54			
8	2.59			
9	2.87	17.46	33.00	48.71
10	3.41			
11	3.42			
12	3.50			
13	3.30	15.54	33.00	48.71
14	3.48			
15	3.76			
16	3.37			
17	3.45	16.97	33.17	48.71
18	3.01			
19	2.97			
20	2.75			
21	3.26	16.20	33.17	48.71
22	3.70			
23	3.20			
24	3.22			
25	3.59	16.20	33.17	48.71
26	3.51			
27	3.12			
28	3.26			
29	3.70	16.20	33.17	48.71
30	2.62			

F. Histograms of activity along months of the year

Activity percentages of the months along the year from LAD list messages. Activity is concentrated in Jun-Aug for MET and LAD, and in Dec-Mar for CPP, LAU and LAD. These observations fit academic calendars, vacations and end-of-year holidays.

TABLE XIII. LAU activity along the months of the year.

	m.	b.	t.	q.	s.
Jan	10.22	19.56	28.24	35.09	49.16
Fev	9.34				
Mar	8.67	15.53	20.93	30.36	
Apr	6.86				
Mai	7.28	14.07	24.47	34.55	50.84
Jun	6.80				
Jul	8.97	16.29	26.36	22.80	
Ago	7.32				
Set	8.18	16.25	15.68	17.00	54.27
Out	8.06				
Nov	7.64	18.30	15.36	14.69	
Dez	10.66				

TABLE XIV. LAD activity along the months of the year.

	m.	b.	t.	q.	s.
Jan	11.24	18.51	26.46	36.07	57.96
Fev	7.26				
Mar	7.95	17.56	31.50	37.56	
Apr	9.61				
Mai	8.94	21.89	22.30	26.37	42.04
Jun	12.95				
Jul	9.03	15.67	19.74	14.69	
Ago	6.64				
Set	6.63	12.38	15.68	17.00	54.27
Out	5.75				
Nov	7.61	13.99	15.36	14.69	
Dez	6.38				

TABLE XV. MET activity along the months of the year.

	m.	b.	t.	q.	s.
Jan	4.87	11.00	16.89	23.30	47.70
Fev	6.13				
Mar	5.89	12.30	30.81	47.87	
Apr	6.41				
Mai	10.45	24.40	31.21	28.83	52.30
Jun	13.95				
Jul	13.24	23.47	21.09	4.59	
Ago	10.22				
Set	7.75	16.79	4.59	4.59	47.70
Out	9.04				
Nov	7.45	12.05	4.59	4.59	
Dez	4.59				

TABLE XVI. CPP activity along the months of the year.

	m.	b.	t.	q.	s.	
Jan	8.70	17.00	27.23	36.49	54.27	
Fev	8.29					
Mar	10.23	19.49	27.03	33.46		
Apr	9.26					
Mai	9.41	17.78	22.94	30.06	45.73	
Jun	8.37					
Jul	8.70	15.68	22.80	14.69		
Ago	6.98					
Set	7.26	15.36	14.69	4.59	47.70	
Out	8.10					
Nov	7.89	14.69	4.59	4.59		
Dez	6.81					

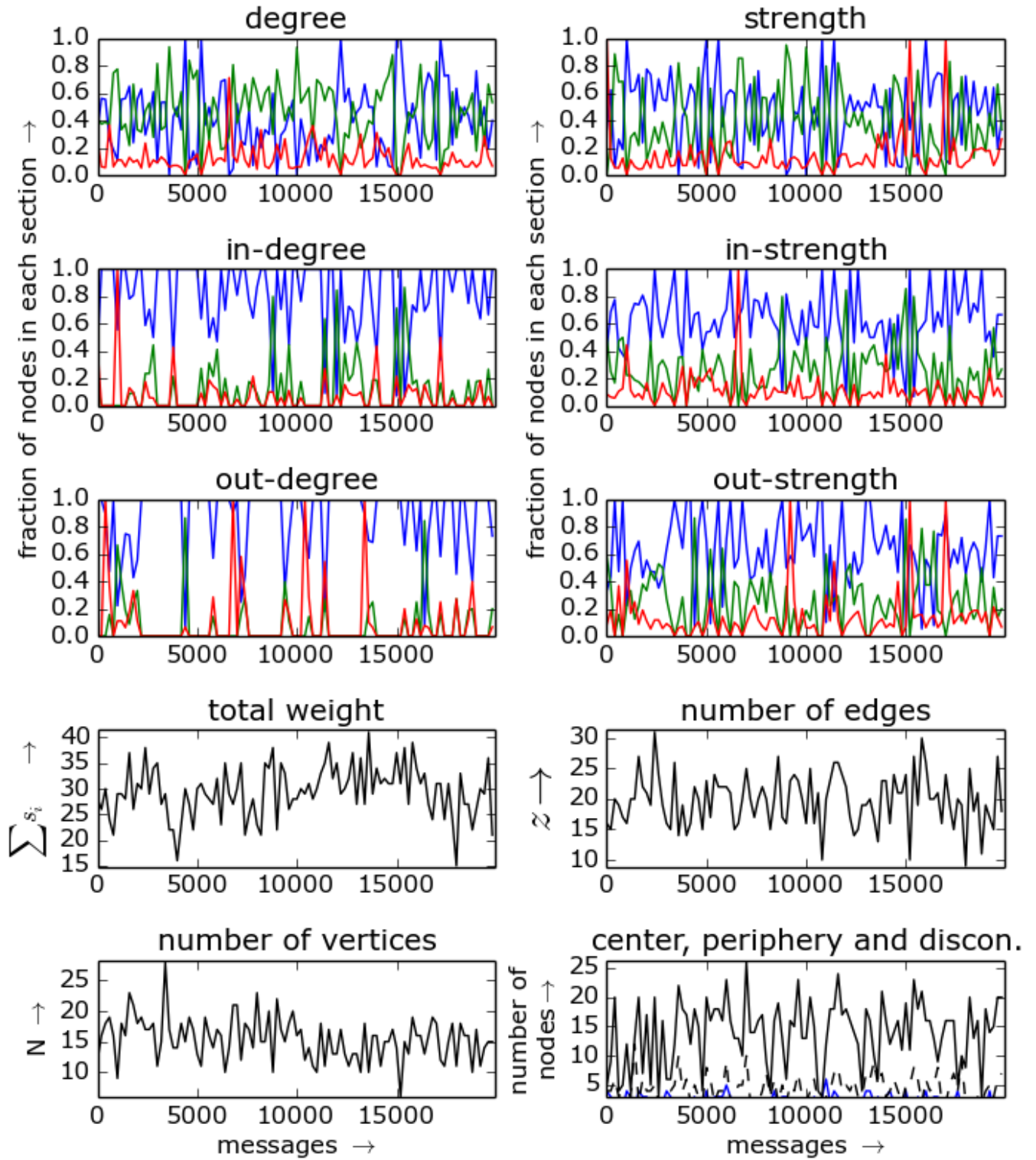
II. FRACTION OF PARTICIPANTS IN EACH ERDŐS SECTOR ALONG THE TIMELINE

Here we present the fraction of participants in each Erdős sector with respect to each criteria defined in Section III C. Step sizes of 50, 100, 250, 500, 1000 and 5000 are shown bellow, first for CPP, than for LAD list.

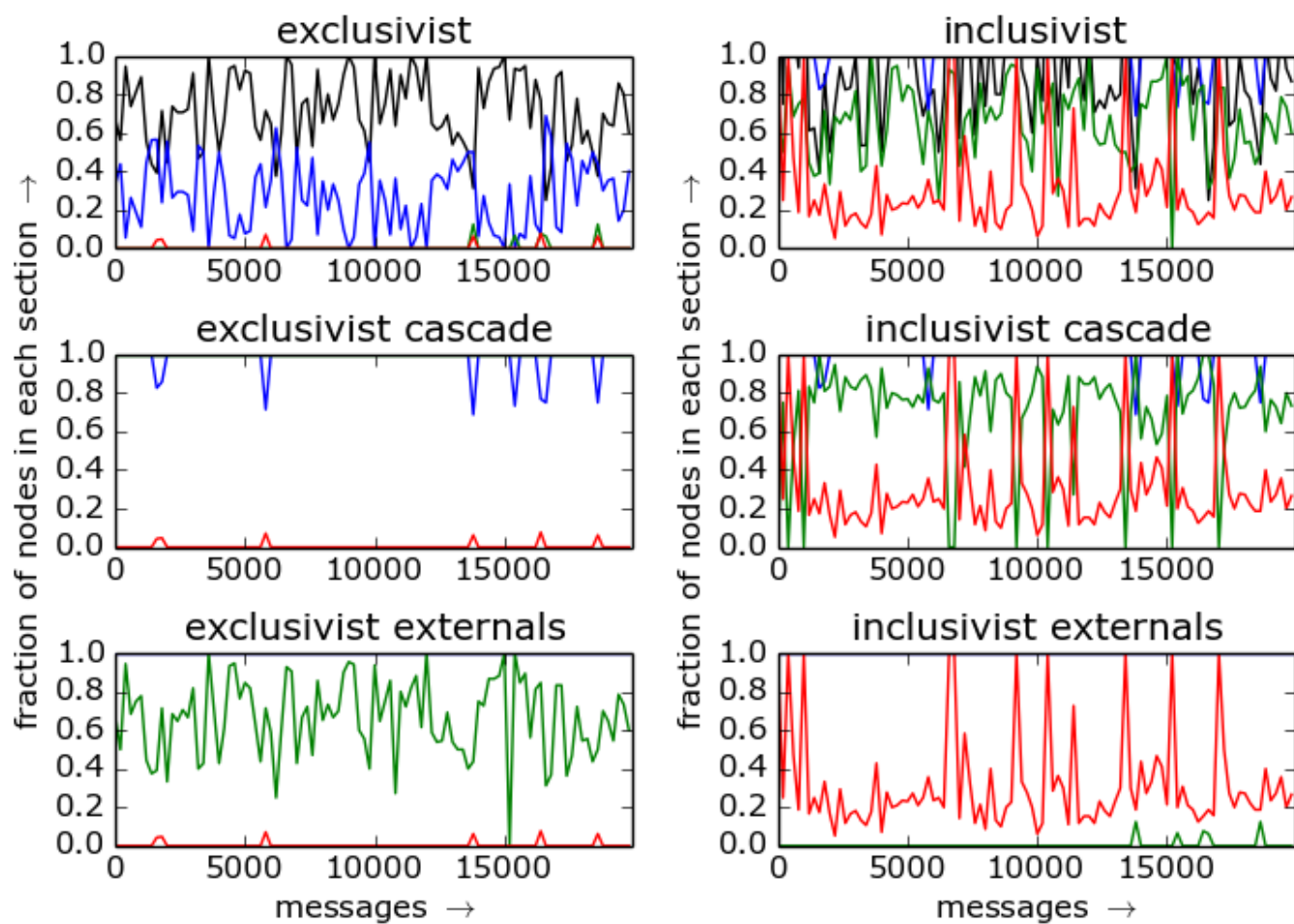
Each step size takes two pages of plot. On the first page, the criteria are each centrality measure observed separately: in and out degrees and strengths. In the first six plots, red is fraction of hubs, green is the fraction of intermediary and blue is for peripheral fraction. On the last plot, red 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.

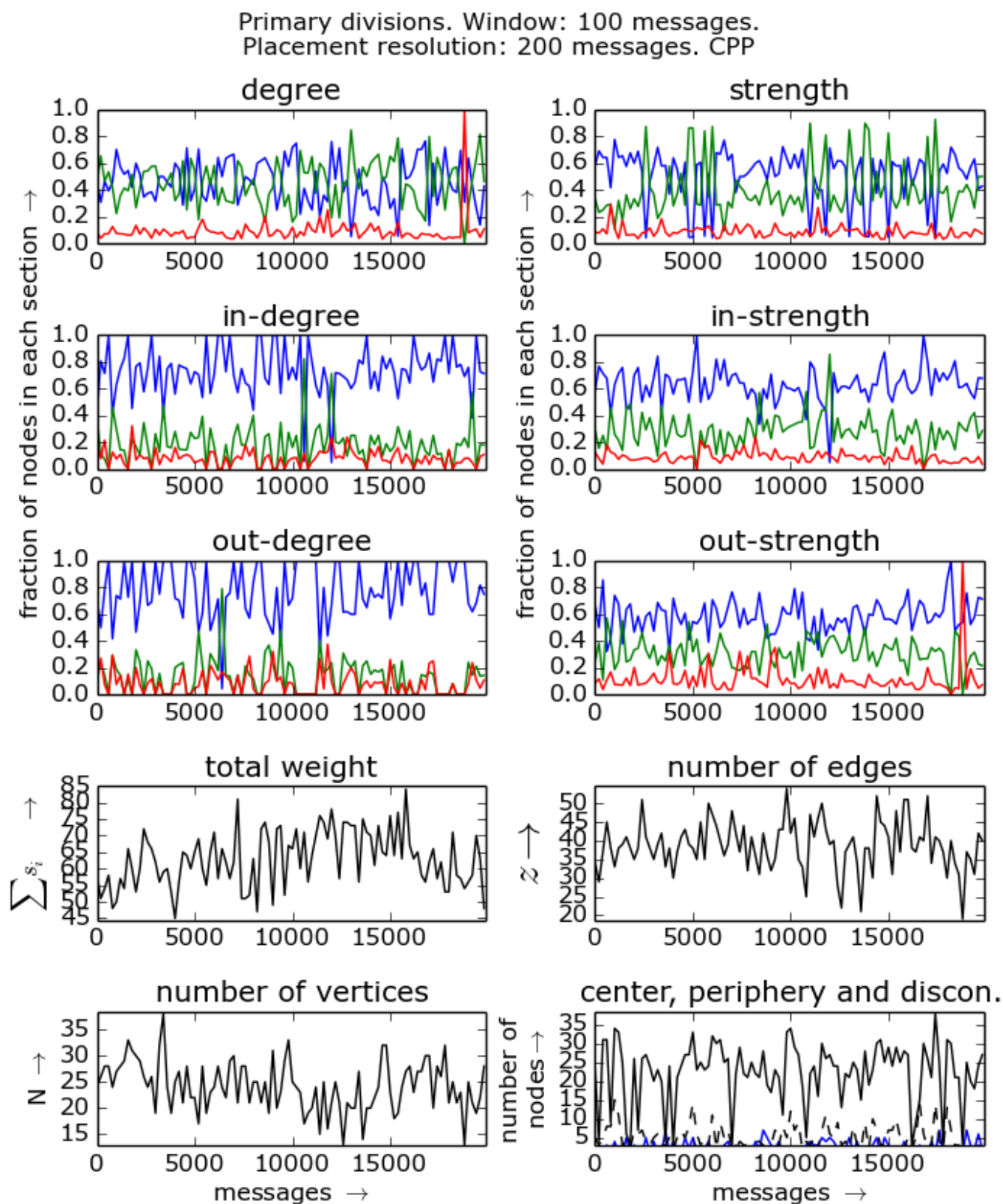
On the second page we show the fractions of participants with respect to each compound criteria for the Erdős sectioning. In the first plot, the fraction of vertices with unique classification is plotted in black: $\frac{\text{number of nodes uniquely classified}}{\text{number of nodes}}$. On the second plot, black represents the exceeding classifications for the given vertices : $\frac{\text{number of classifications} - \text{number of nodes}}{\text{number of nodes}}$.

Primary divisions. Window: 50 messages.
Placement resolution: 200 messages. CPP

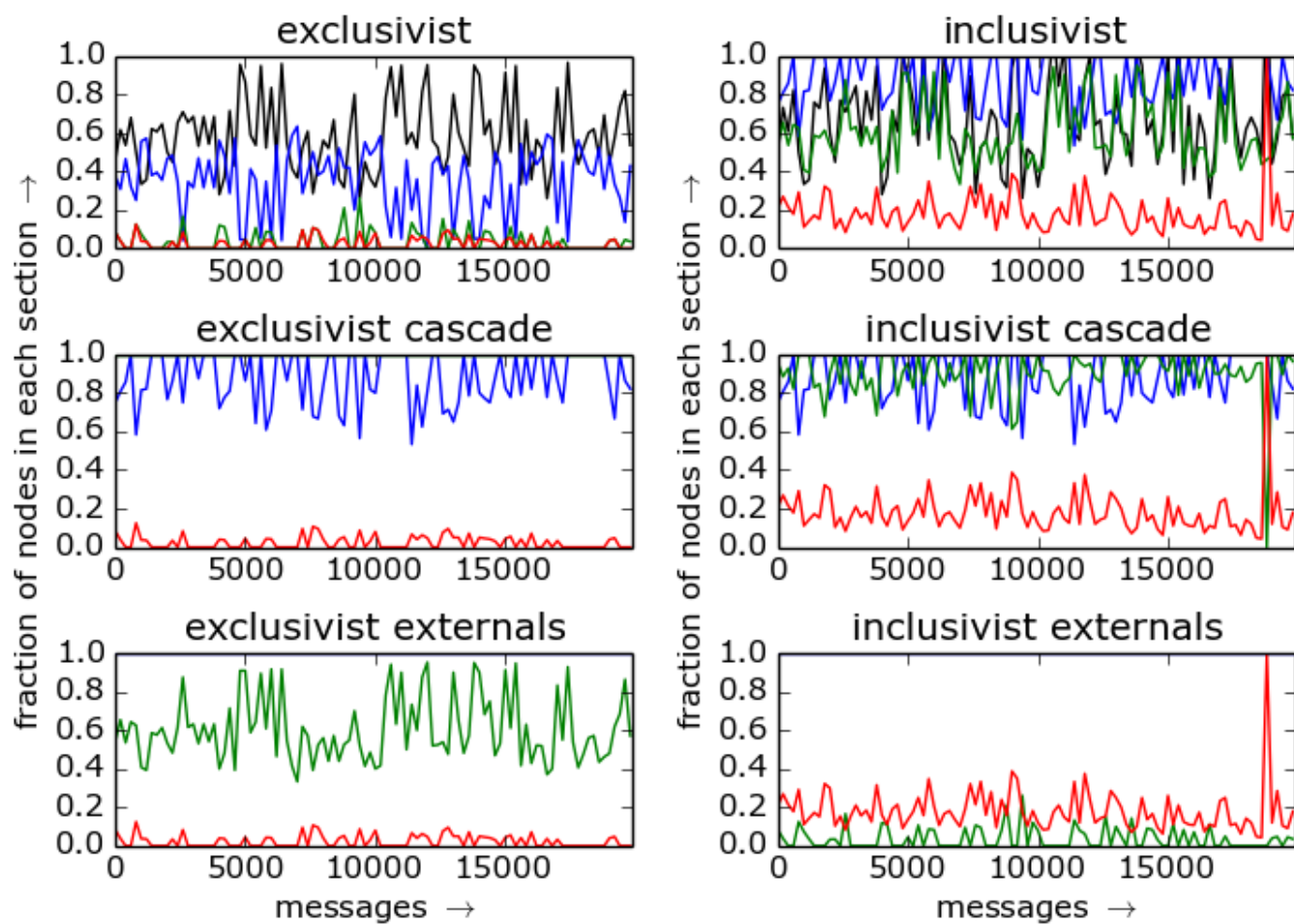


Compound divisions. Window: 50 messages.
Placement resolution: 200 messages. CPP

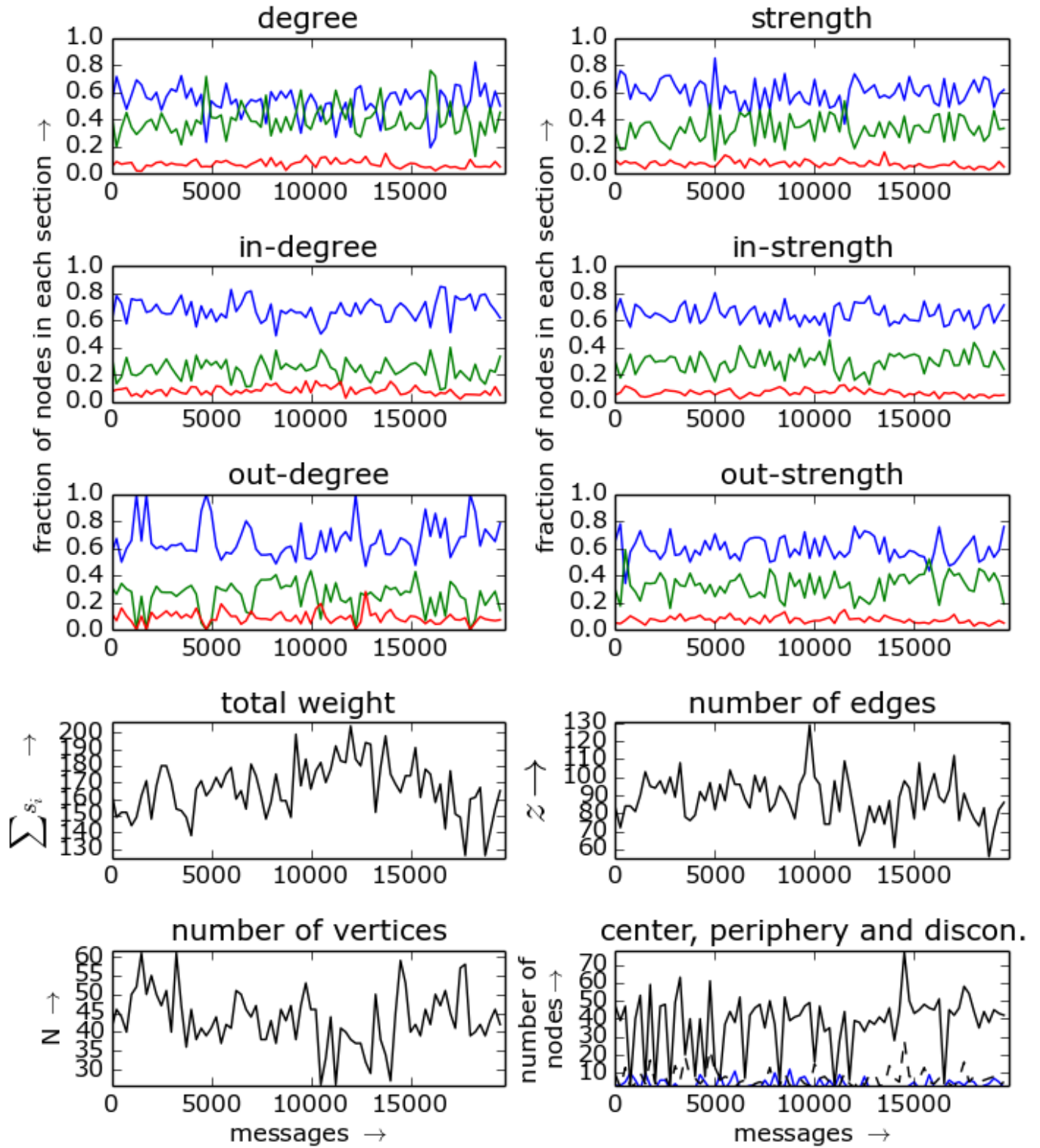




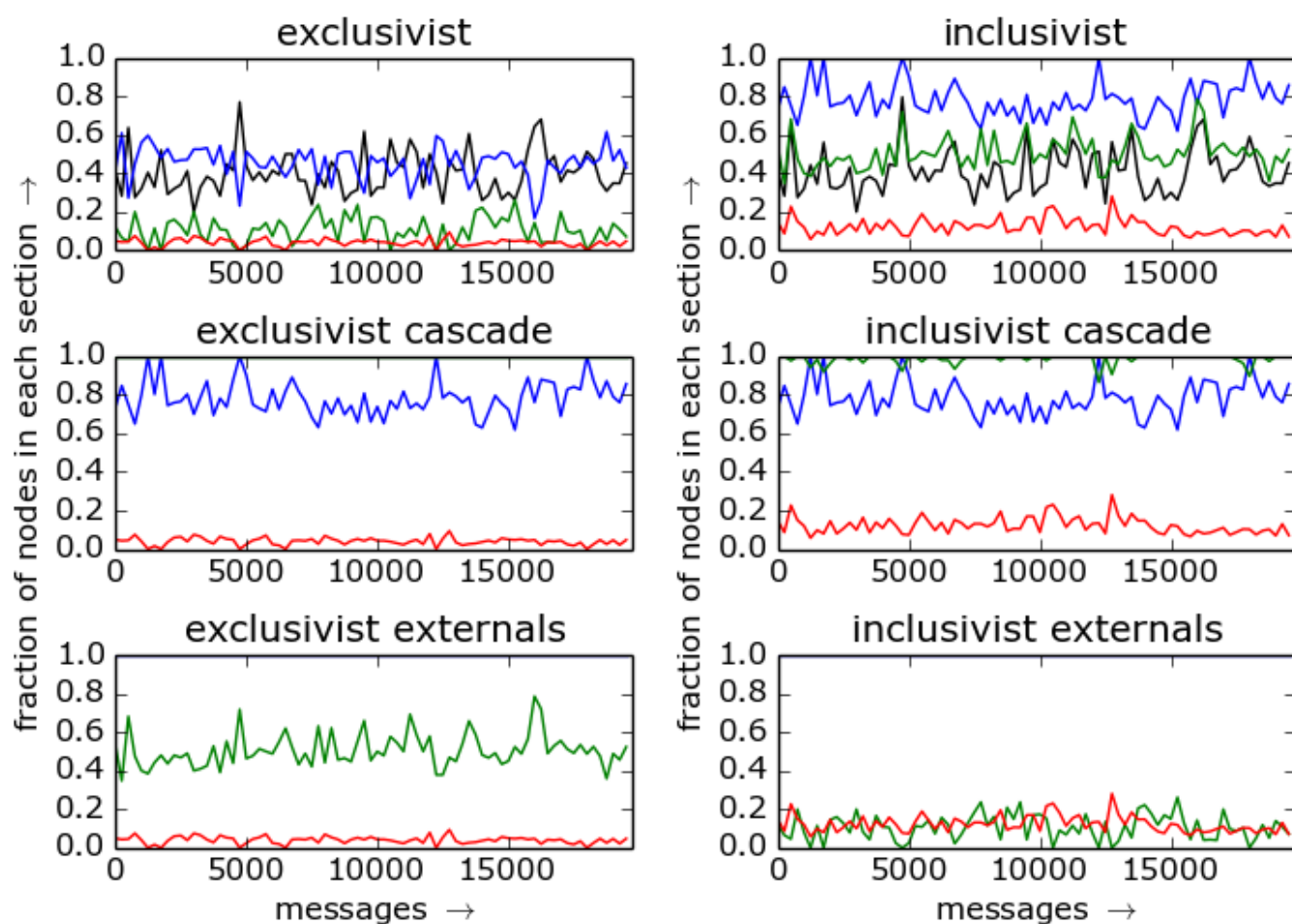
Compound divisions. Window: 100 messages.
Placement resolution: 200 messages. CPP

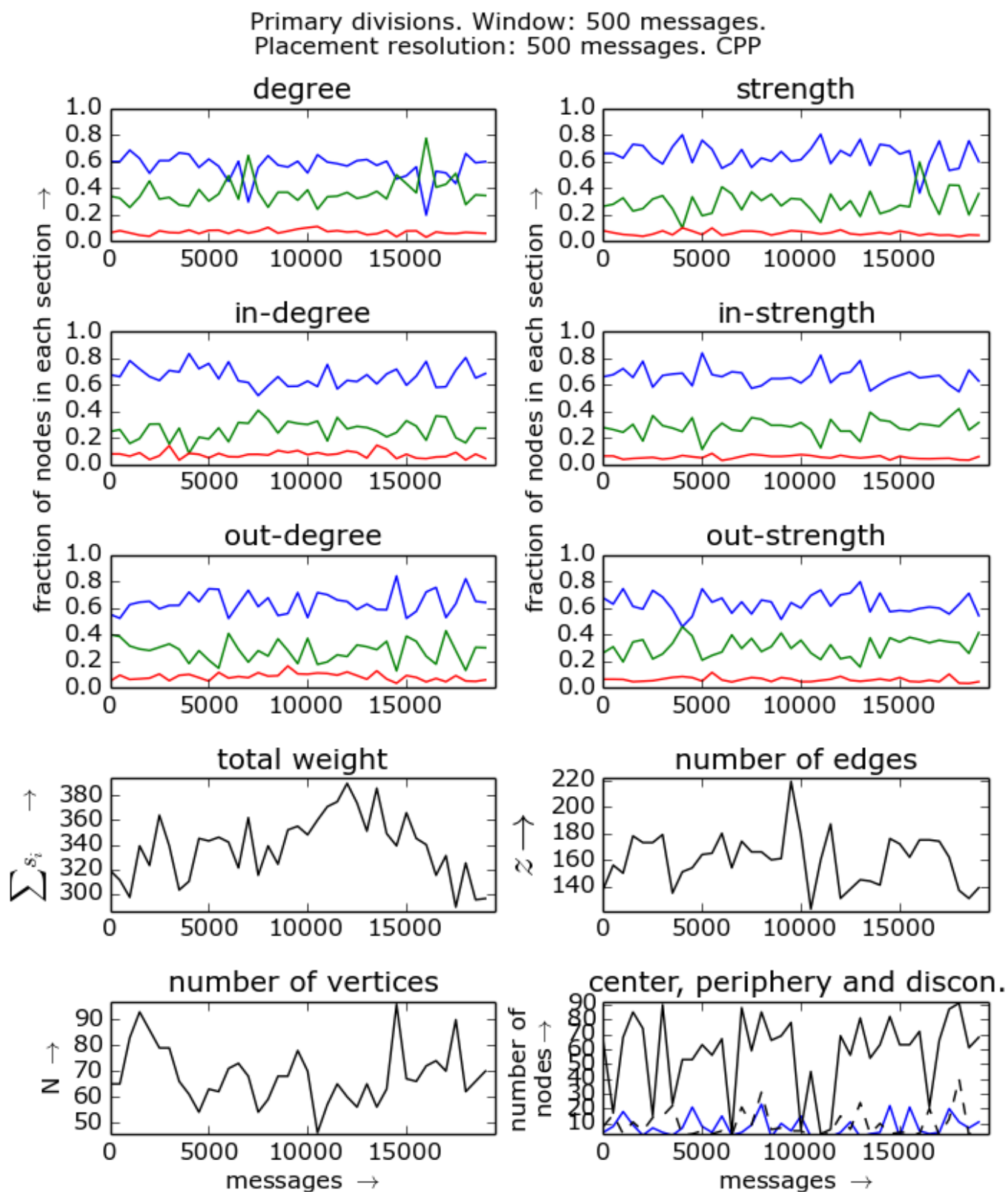


Primary divisions. Window: 250 messages.
Placement resolution: 250 messages. CPP

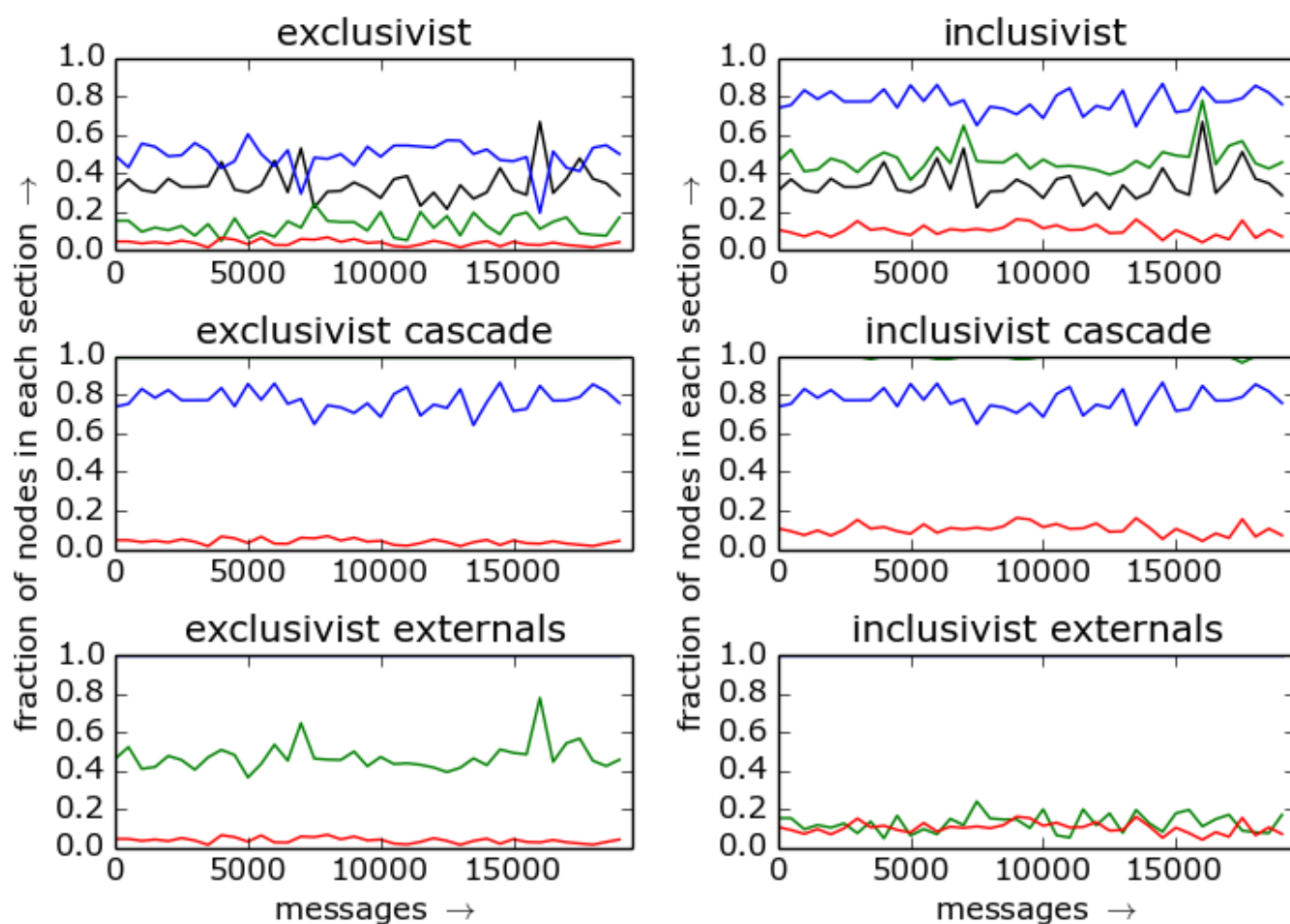


Compound divisions. Window: 250 messages.
Placement resolution: 250 messages. CPP

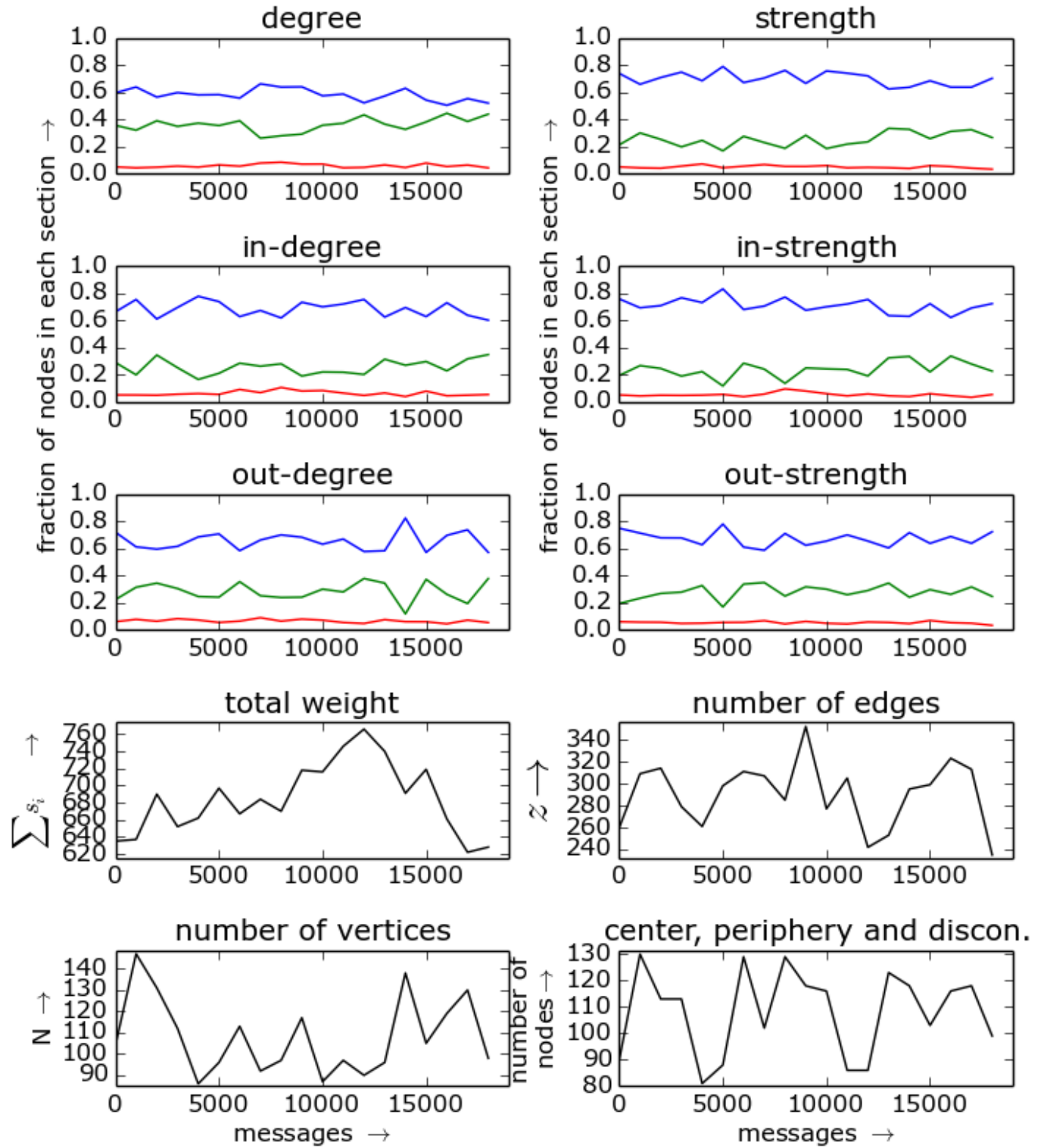




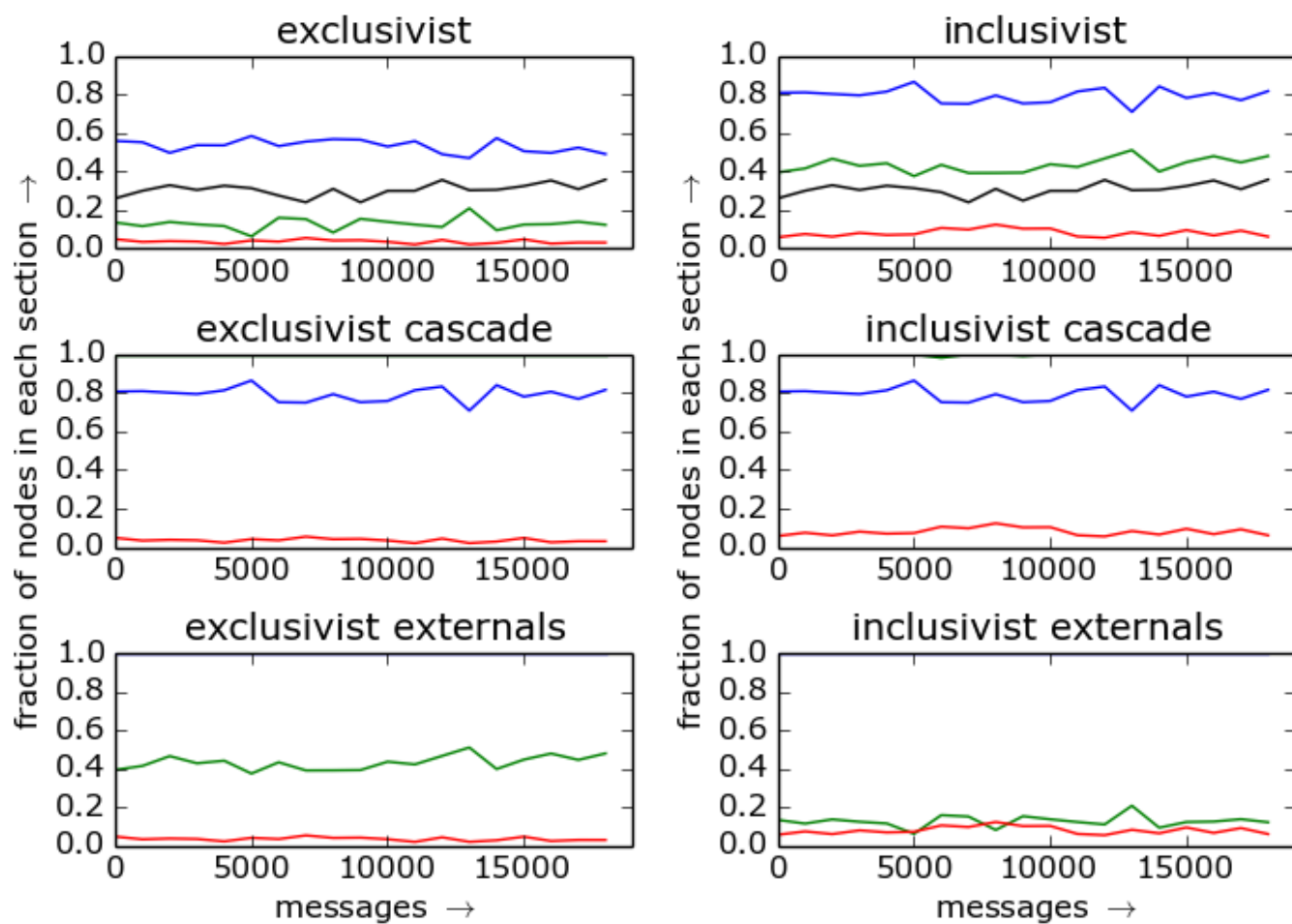
Compound divisions. Window: 500 messages.
Placement resolution: 500 messages. CPP



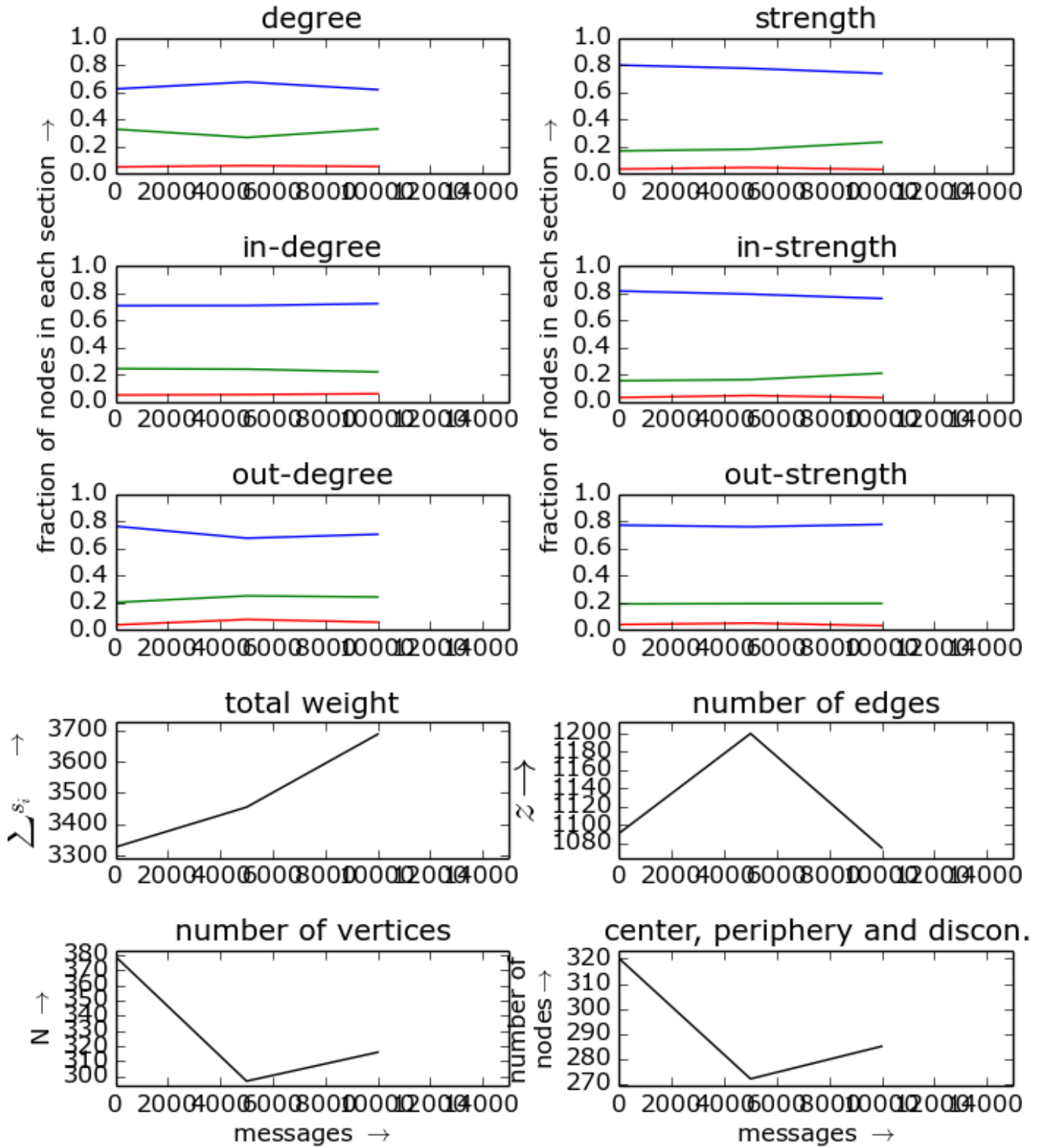
Primary divisions. Window: 1000 messages.
Placement resolution: 1000 messages. CPP



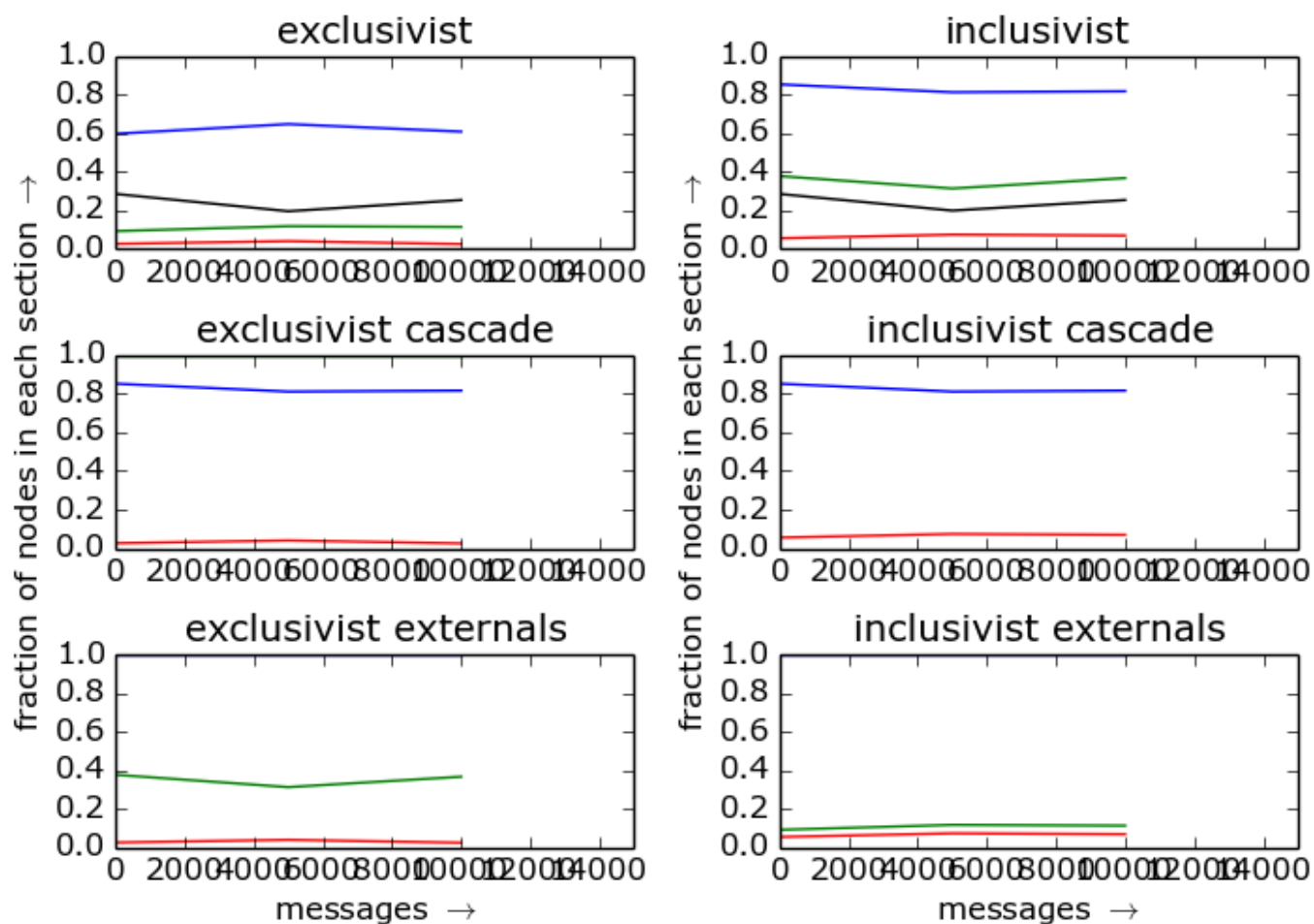
Compound divisions. Window: 1000 messages.
Placement resolution: 1000 messages. CPP



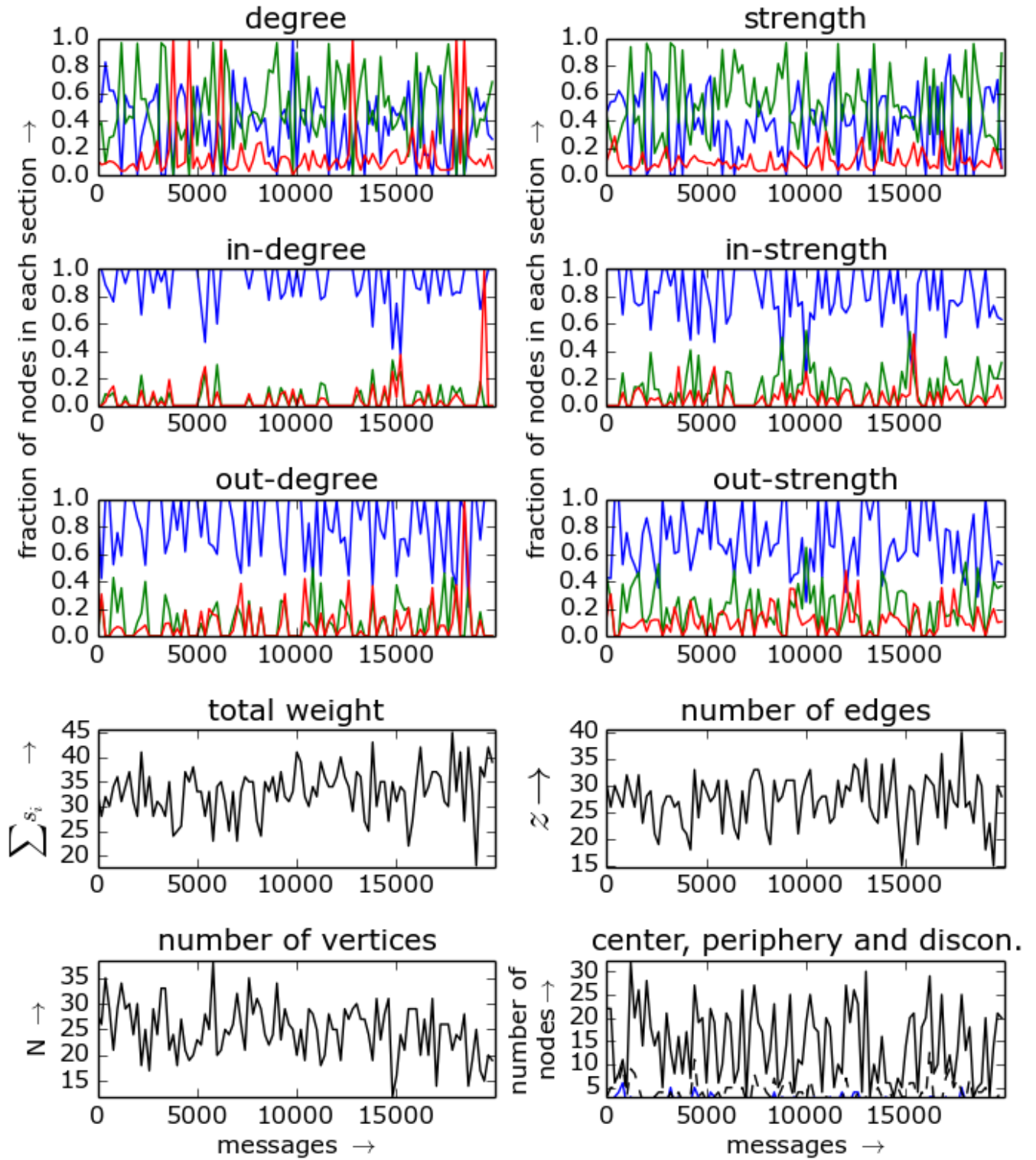
Primary divisions. Window: 5000 messages.
Placement resolution: 5000 messages. CPP



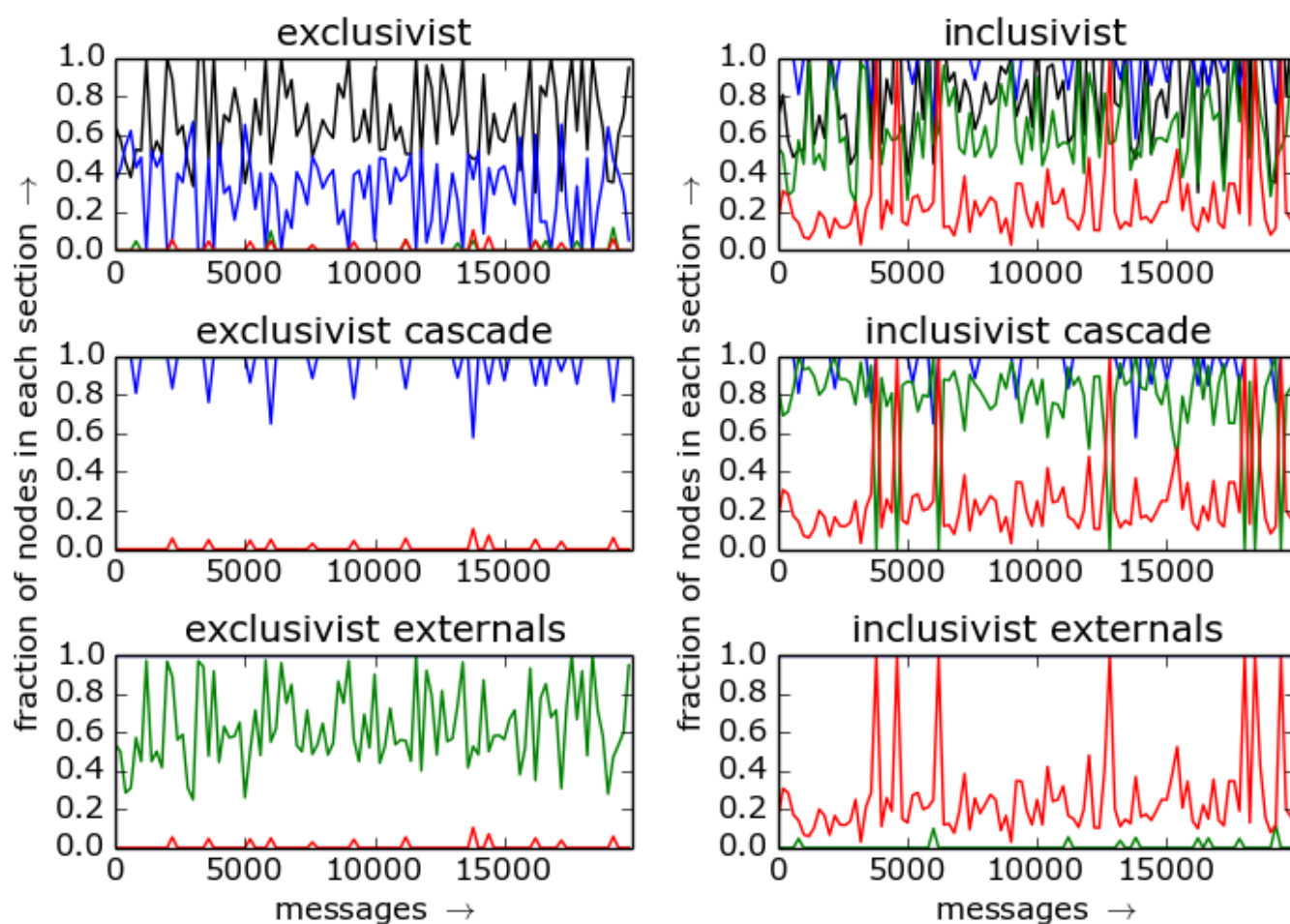
Compound divisions. Window: 5000 messages.
Placement resolution: 5000 messages. CPP



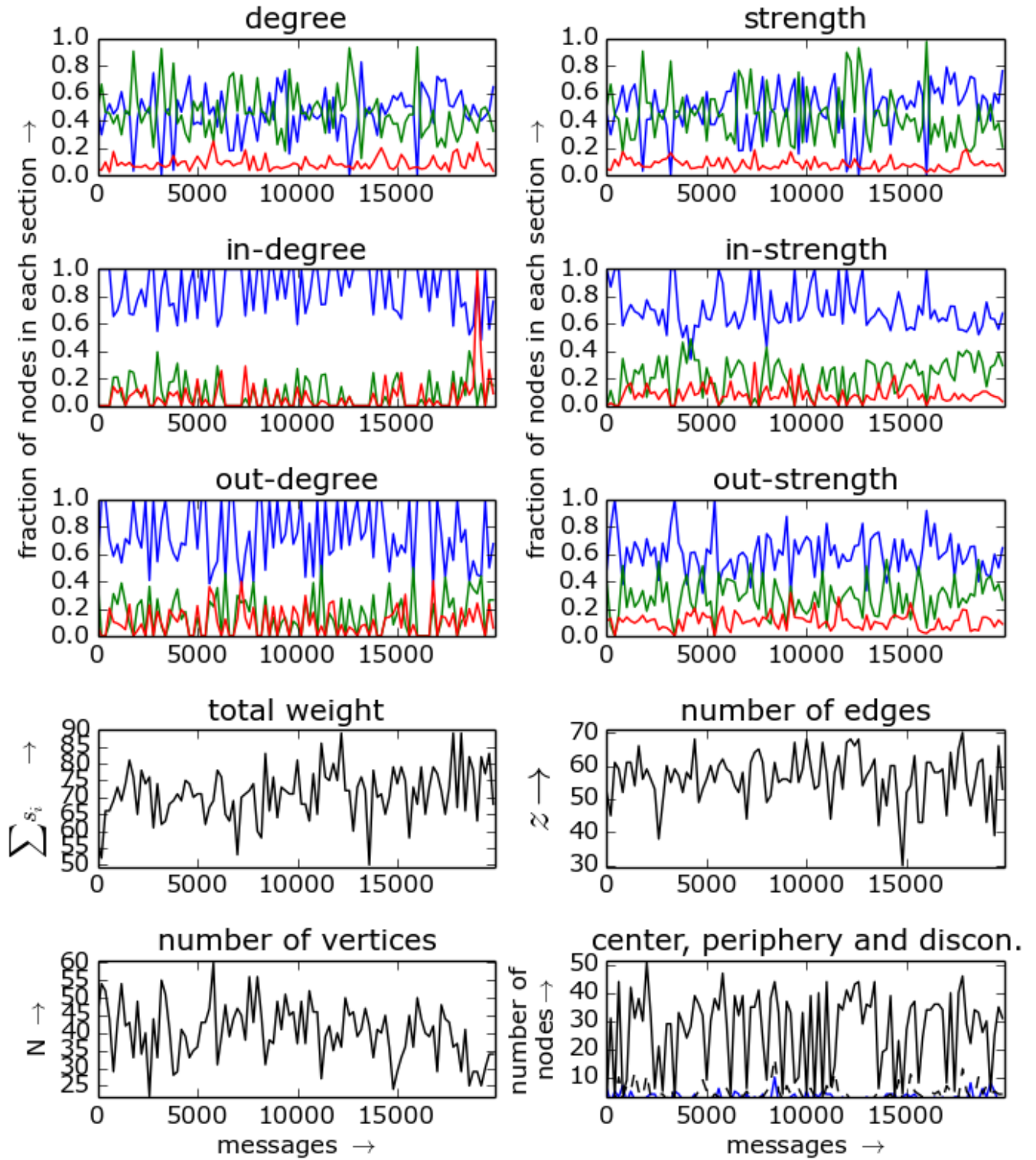
Primary divisions. Window: 50 messages.
Placement resolution: 200 messages. LAD



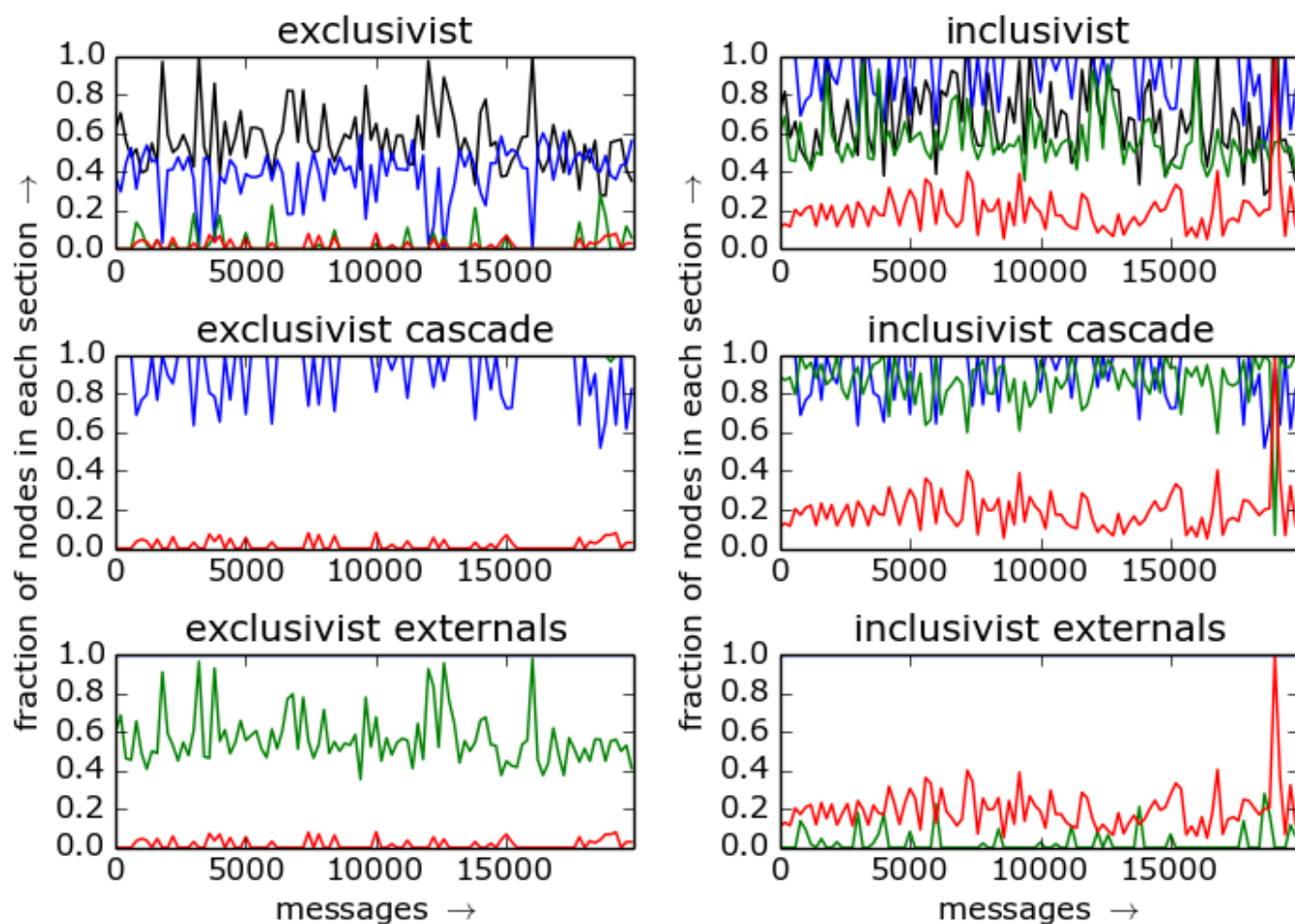
Compound divisions. Window: 50 messages.
Placement resolution: 200 messages. LAD



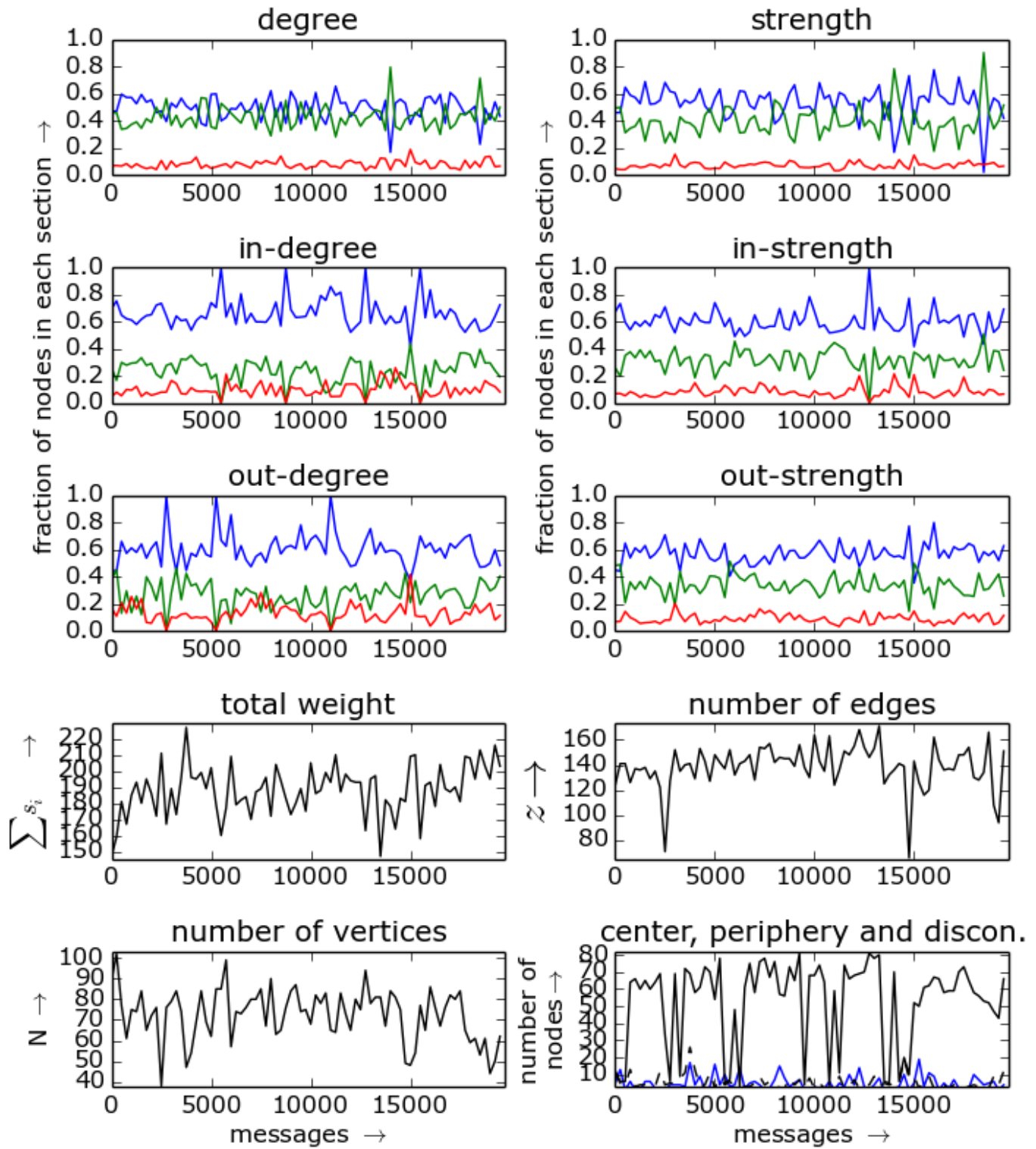
Primary divisions. Window: 100 messages.
Placement resolution: 200 messages. LAD



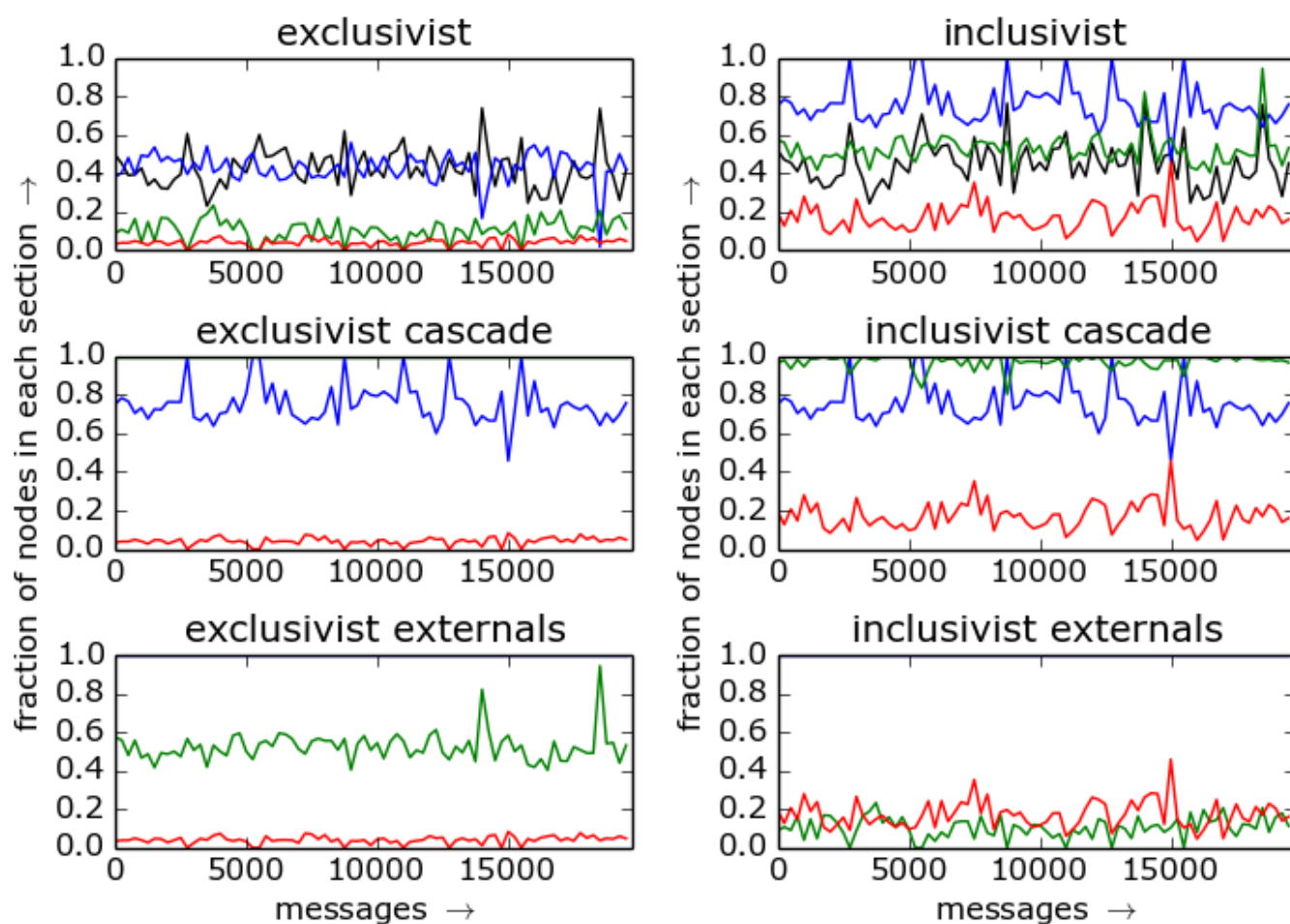
Compound divisions. Window: 100 messages.
Placement resolution: 200 messages. LAD



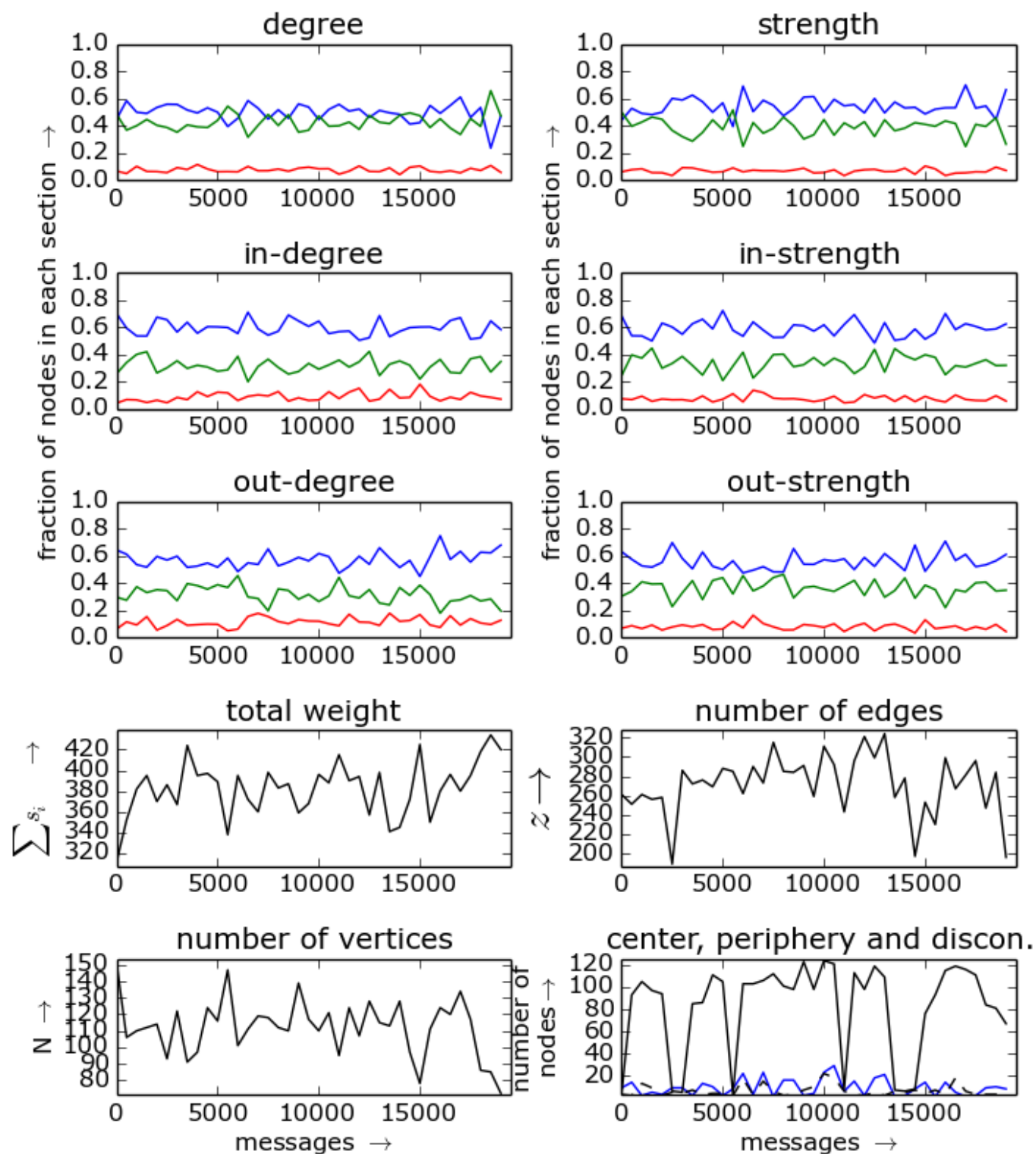
Primary divisions. Window: 250 messages.
Placement resolution: 250 messages. LAD



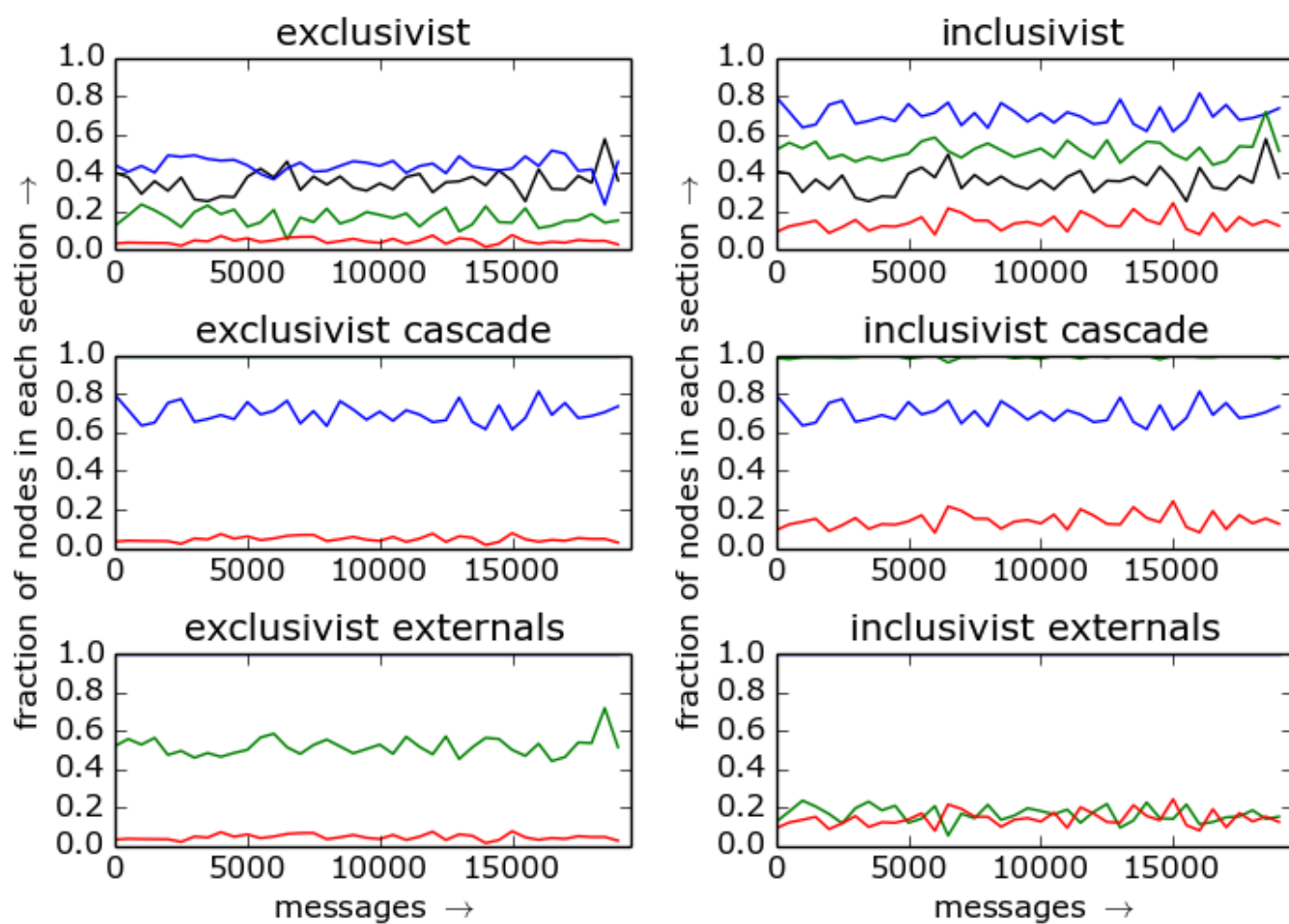
Compound divisions. Window: 250 messages.
Placement resolution: 250 messages. LAD



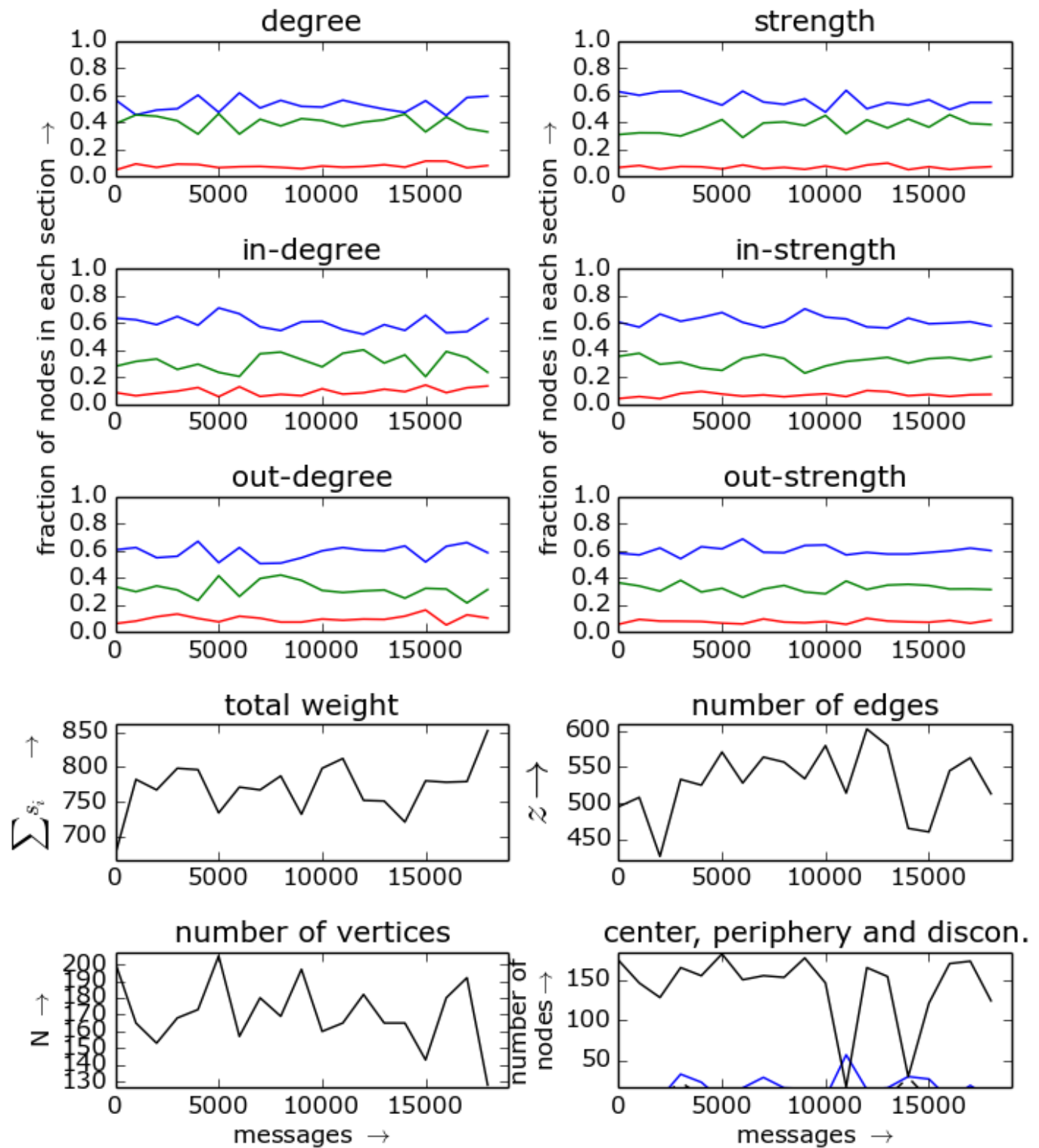
Primary divisions. Window: 500 messages.
Placement resolution: 500 messages. LAD



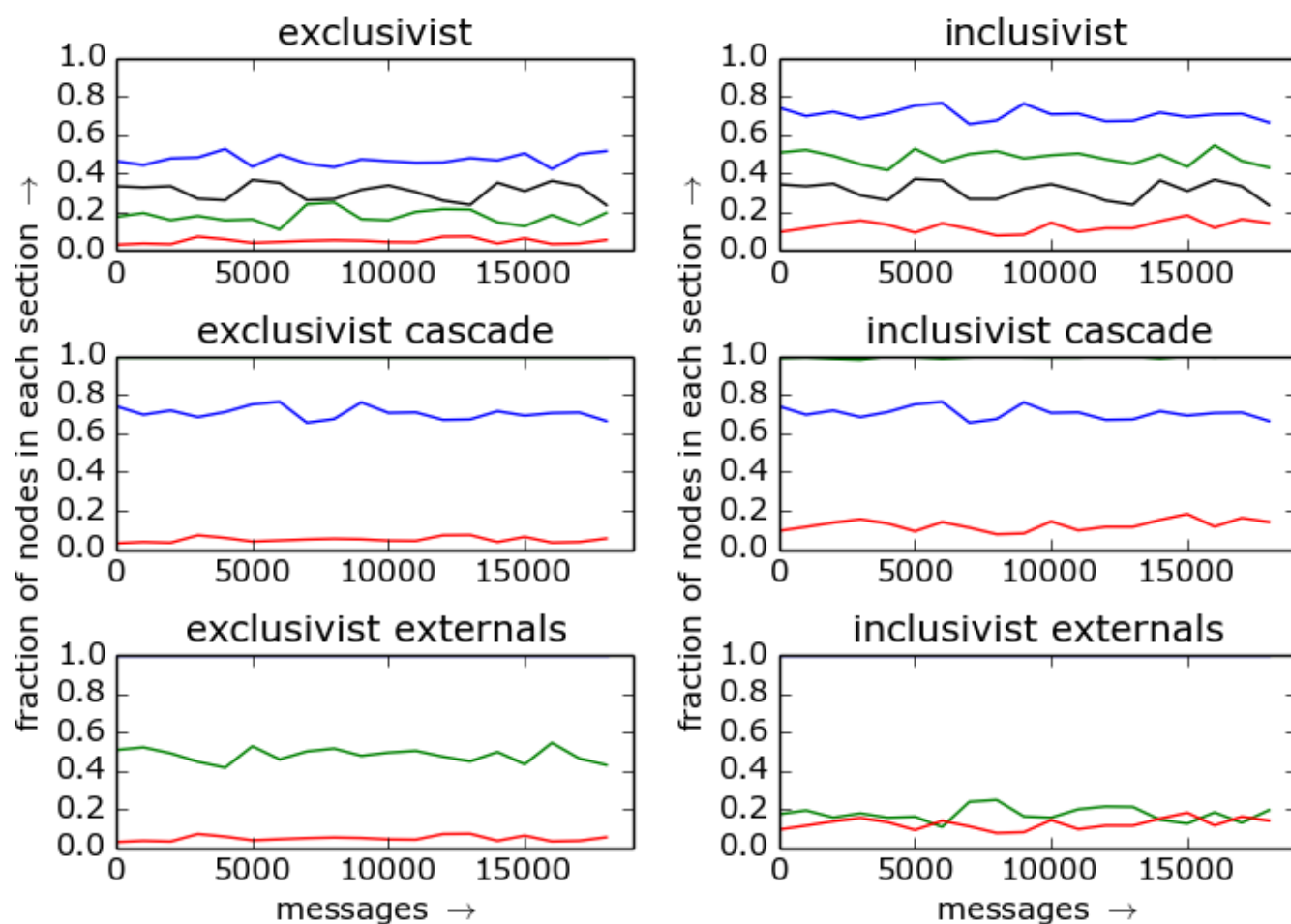
Compound divisions. Window: 500 messages.
Placement resolution: 500 messages. LAD



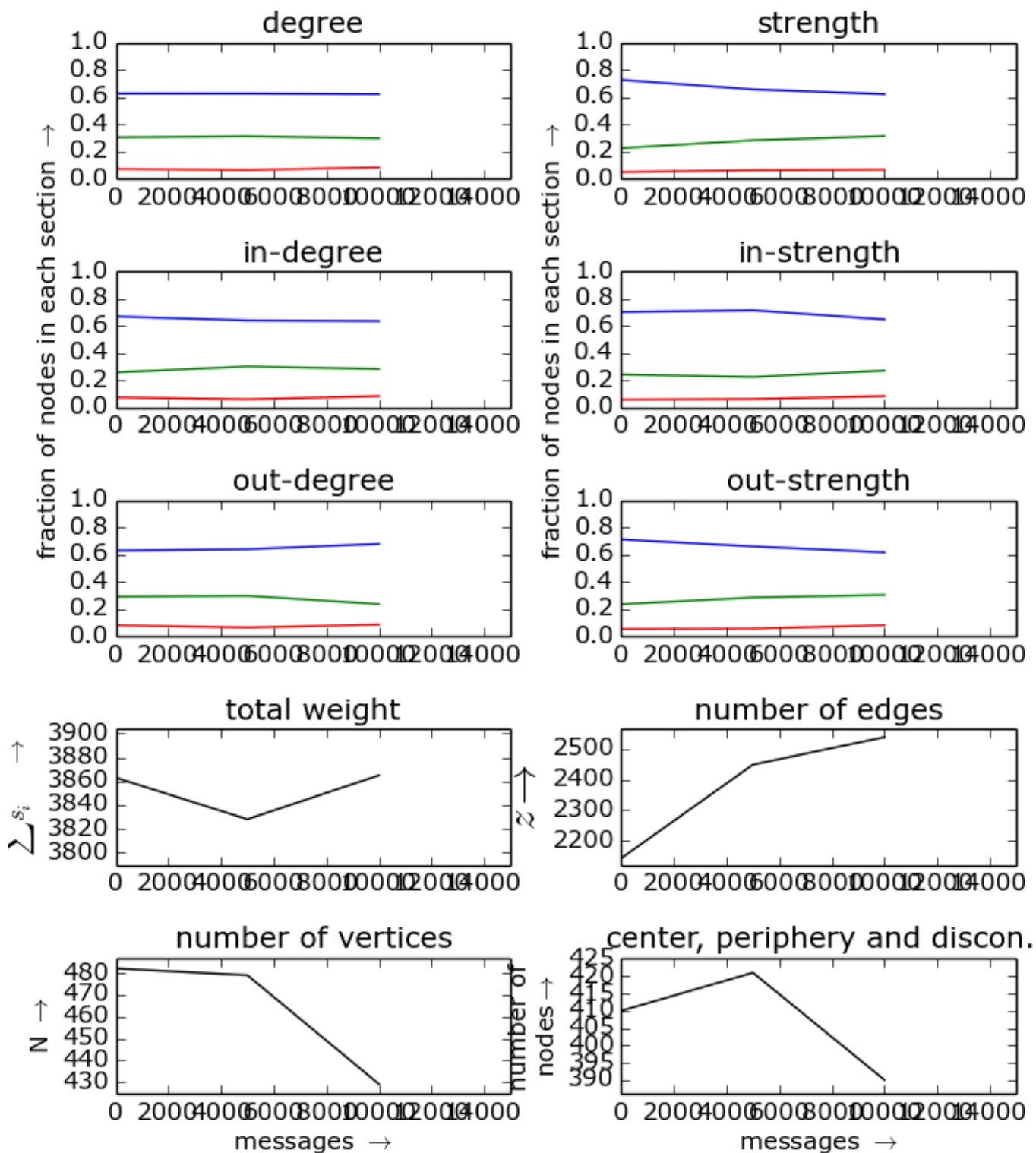
Primary divisions. Window: 1000 messages.
Placement resolution: 1000 messages. LAD



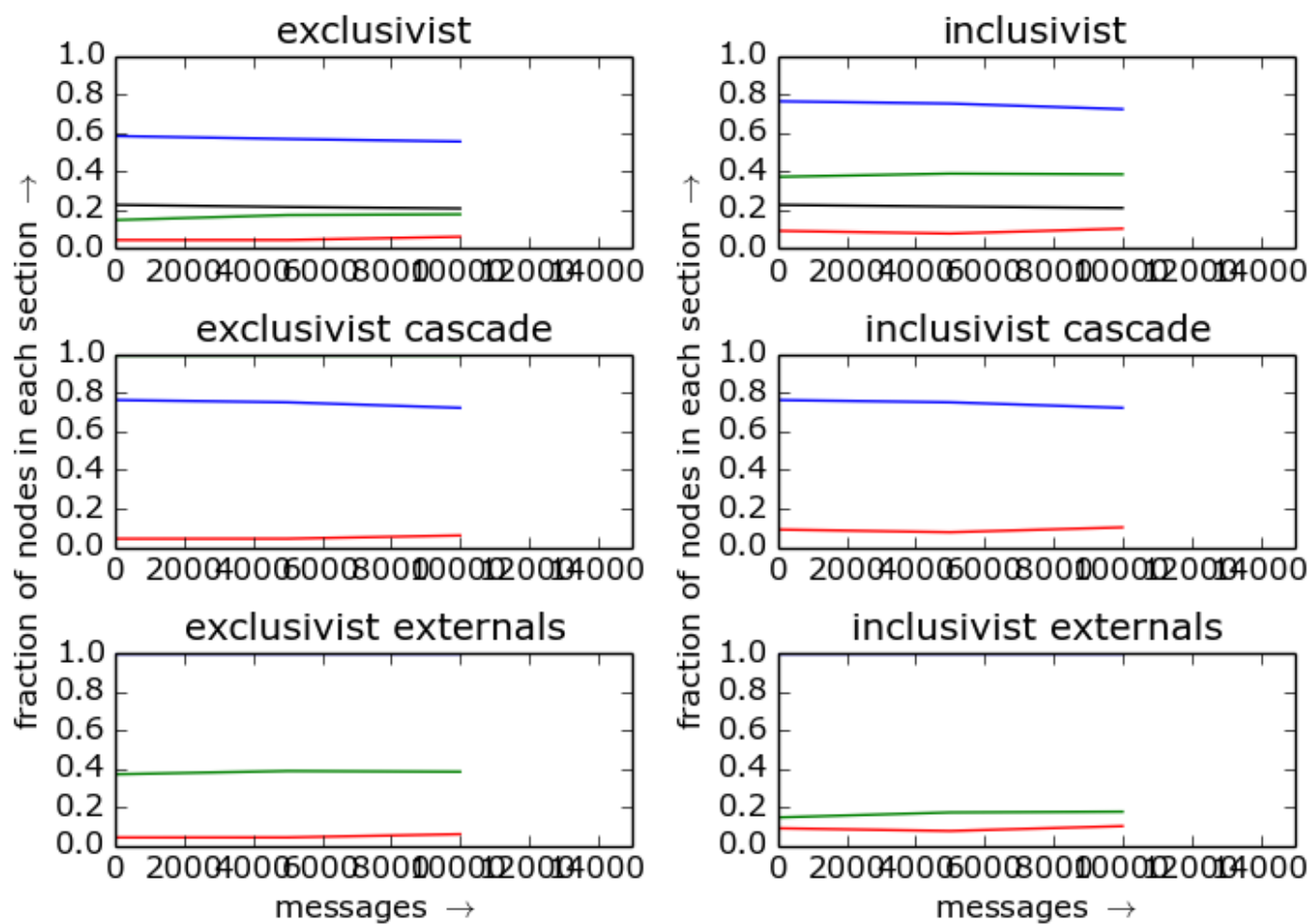
Compound divisions. Window: 1000 messages.
Placement resolution: 1000 messages. LAD



Primary divisions. Window: 5000 messages.
Placement resolution: 5000 messages. LAD



Compound divisions. Window: 5000 messages.
Placement resolution: 5000 messages. LAD



III. PCA OF MEASURES ALONG THE TIMELINE

A. Betweenness, clustering and degree

B. Betweenness, clustering, degrees and strengths

C. Betweenness, clustering, degrees, strengths and symmetry measures

¹R. Fabbri, “Stability in human interaction networks: vertices sectorialization, prominence of topological measures and time activity statistics,” arXiv preprint arXiv:1310.7769. <http://arxiv.org/abs/1310.7769>.