

Supplementary Information for “A Multi-Perturbation Stability Framework for Assessing Community Structure in Complex Networks”

Jinji Zhou¹, Wei Zhu^{2*}, Jinyin Bai¹, Xiangchen Wang¹,
Tianjin Ni¹, Zhaoyang Gu¹

¹Former College of Information and Communication, National University
of Defense Technology, Wuhan, 430034, Hubei, China.

^{2*}Hubei Provincial Defense Science and Technology Innovation Center for
Cyber & Information Game Confrontation and System Simulation,
Information Support Force Engineering University, Wuhan, 430019,
Hubei, China.

*Corresponding author(s). E-mail(s): zhuwei929@hotmail.com;

Abstract

This document contains Supplementary Figures S1–S8 and Supplementary
Tables S1–S10 supporting the main article. All symbols, perturbation operators,
and statistical conventions follow the Methods of the main article; empirical
p-values follow the add-one convention defined in the Methods of the main article.

Keywords: complex networks, community structure, perturbation stability, statistical
significance

1 Datasets and experiment configuration

Supplementary Table 1 lists the four real-world networks and the two control net-
works used in the main article, together with their basic statistics. Supplementary
Table 2 summarizes the complete parameter configuration of the real-world verification
experiment.

Table 1 Basic information of real-world and control networks.

Network	n	m	$\langle k \rangle$	Source / Description
Karate	34	78	4.59	Zachary’s karate club
Dolphins	62	159	5.13	Lusseau’s dolphin social network
Football	115	613	10.66	US college football schedule
Polbooks	105	441	8.40	Krebs’ political books co-purchasing
Strong planting $\mu = 0.05$	200	1000	10.00	SBM strong community structure, positive control
ER random graph	200	1000	10.00	$G(n, p)$, no community structure, negative control

Network sources as cited in the main article: Zachary (1977); Lusseau et al. (2003); Girvan & Newman (2002); Krebs (2002).

Table 2 Parameter configuration of the real-world network verification experiment.

Module	Parameter	Value	Description
Network data	Real networks	Karate, Dolphins, Football, Polbooks	Classic undirected social/biological networks
	Positive control	SBM ($N=200$, $C=10$, $\mu=0.10$, $\langle k \rangle=10$)	Strong community structure
	Negative control	ER random graph ($N=200$, $m=1000$)	No community structure
	Null model	Configuration model (Maslov–Sneppen)	Degree-sequence preserving
Random baseline	Edge swap multiplier	$4m$	Degree-preserving randomization
	Random networks R	500	Construction of the empirical null distribution

Detection settings follow the general protocol of the main article: Louvain detector with 5 independent runs (best partition selected), $B = 20$ Monte Carlo repetitions, perturbation grid $\alpha \in \{0, 0.05, 0.10, 0.20, 0.30, 0.50\}$, and 20 independent full repetitions with null set per repetition; significance by majority voting (at least 12 of 20 repetitions with $p < 0.05$).

2 CRITIC weight calibration, validation, and ablation

Supplementary Table 3 reports the CRITIC weight calibration on 50 SBM calibration graphs, including bootstrap stability. Supplementary Table 4 reports the weighting-scheme ablation, and Supplementary Fig. 1 shows in-distribution and out-of-distribution (LFR) validation of the SBM-calibrated weights.

3 SBM benchmark behaviour and the fair comparison with modularity

Supplementary Table 5 reports instance-averaged $S(G)$, Q , and the three per-operator AUC values across the full mixing-parameter gradient. Supplementary Table 6 reports the four-level fair null-calibrated comparison, and Supplementary Figs. 2 and 3 show the corresponding curves.

Table 3 CRITIC weight calibration results.

Perturbation type	σ	Conflict	Weight w	CV (%)	95% CI
REW	0.385	0.0215	0.465	1.44	[0.452, 0.478]
RER	0.393	0.0088	0.195	1.61	[0.189, 0.202]
RNR	0.390	0.0155	0.340	2.39	[0.323, 0.356]

Min-max positive normalization is applied to the three AUC metrics before calibration to eliminate dimensional differences. σ is the normalized standard deviation of the metric on the 50 SBM calibration networks. Conflict = $\sum_{k \neq j} (1 - r_{jk})$. Pairwise Pearson correlation coefficients: $r(\text{REW}, \text{RER}) = 0.9926$, $r(\text{REW}, \text{RNR}) = 0.9859$, $r(\text{RER}, \text{RNR}) = 0.9986$.

Table 4 CRITIC weighting ablation study.

Metric	$ \rho $ with u	DM (weak regime)
S_{CRITIC} (CRITIC weights)	1.0000	0.2982
S_{equal} (equal weights)	1.0000	0.3088
REW only	1.0000	0.2547
RER only	1.0000	0.3349
RNR only	1.0000	0.3367

u denotes the ground-truth strength ordering (with $1 - \mu$ as the strength proxy); $|\rho|$ is the absolute Spearman rank correlation with u over the full μ gradient; DM is the weak-community discriminability over $\mu \in [0.50, 0.70]$ (Methods of the main article). In-distribution validation uses 100 held-out SBM graphs; out-of-distribution testing on LFR networks ($N = 200$, $\tau_1 = 2.5$, $\tau_2 = 1.5$, $\langle k \rangle = 10$) yields $|\rho| = 0.9636$ for the SBM-calibrated weights.

4 Monte Carlo repetition sensitivity

Supplementary Table 7 and Supplementary Fig. 4 quantify the influence of the Monte Carlo repetition count B on estimation accuracy and runtime.

5 Cross-detector supplementary panels

Supplementary Fig. 5 shows weak-community discriminability across weighting configurations and the $\alpha = 0$ baseline comparison between re-running the stochastic detector and reusing the stored optimal baseline partition.

6 Scale sensitivity and z-score normalization

Supplementary Table 8 lists the parameter settings of the two network- and community-scaling conditions; Supplementary Table 9 reports the resulting $S(G)$ values; Supplementary Table 10 and Supplementary Fig. 6 summarize the effect of z-score normalization on cross-scale variation.

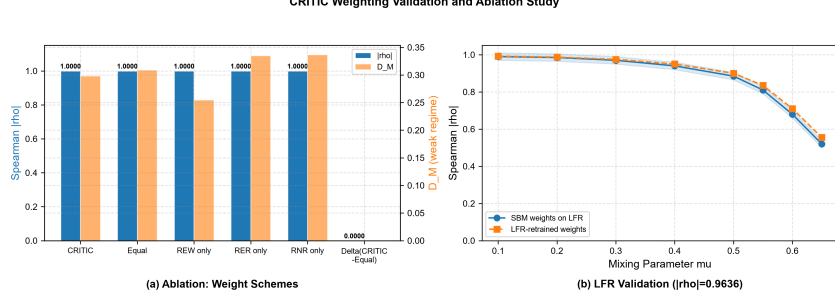


Fig. 1 CRITIC weighting validation: in-distribution and out-of-distribution (LFR) results.

Table 5 Comparison of $S(G)$ and modularity Q on SBM networks.

μ	$1 - \mu$	$S(G)$	Q	AUC(REW)	AUC(RER)	AUC(RNR)
0.05	0.95	0.9812 ± 0.0010	0.8444 ± 0.0057	0.9621 ± 0.0023	0.9979 ± 0.0004	0.9978 ± 0.0003
0.10	0.90	0.9760 ± 0.0027	0.7963 ± 0.0066	0.9537 ± 0.0034	0.9950 ± 0.0018	0.9955 ± 0.0023
0.20	0.80	0.9547 ± 0.0060	0.7076 ± 0.0028	0.9169 ± 0.0084	0.9847 ± 0.0041	0.9893 ± 0.0041
0.30	0.70	0.8984 ± 0.0126	0.6034 ± 0.0073	0.8351 ± 0.0145	0.9445 ± 0.0101	0.9585 ± 0.0117
0.40	0.60	0.7814 ± 0.0153	0.5106 ± 0.0166	0.6931 ± 0.0180	0.8329 ± 0.0182	0.8726 ± 0.0112
0.50	0.50	0.5604 ± 0.0253	0.4061 ± 0.0047	0.4730 ± 0.0232	0.6006 ± 0.0270	0.6569 ± 0.0272
0.55	0.45	0.4474 ± 0.0076	0.3567 ± 0.0028	0.3670 ± 0.0074	0.4753 ± 0.0098	0.5413 ± 0.0086
0.60	0.40	0.3405 ± 0.0155	0.3196 ± 0.0026	0.2785 ± 0.0160	0.3559 ± 0.0172	0.4164 ± 0.0158
0.65	0.35	0.3045 ± 0.0103	0.2917 ± 0.0041	0.2521 ± 0.0105	0.3099 ± 0.0120	0.3730 ± 0.0098
0.70	0.30	0.2830 ± 0.0249	0.2910 ± 0.0041	0.2340 ± 0.0236	0.2881 ± 0.0267	0.3472 ± 0.0261

SBM parameters $N = 200$, $C = 10$, $\langle k \rangle = 10$; Louvain detector. Entries are means $\pm$ standard deviations over the SBM instances generated for this validation experiment. The raw scores in Supplementary Table S5 come from the separate balanced null-calibration experiment (20 newly generated instances per μ , $R = 200$ null networks); because the two experiments use independently drawn instance sets, values in the near-threshold regime ($\mu = 0.40$ – 0.50) differ by up to 0.05 between the two tables, which does not affect any ranking, ROC-AUC, or significance conclusion.

7 Null-model validation and Type-I error control

Supplementary Fig. 7 shows the configuration-model null distribution of $S(G)$ for the representative SBM network at $\mu = 0.40$ ($S_{\text{obs}} = 0.7731$, $T_{0.95} = 0.2625$, $p = 0.0020$). Supplementary Fig. 8 shows Type-I error stability across 50 independent ER random graph realizations (empirical rate 4.0%, exact binomial 95% CI 0.5%–13.7%).

Table 6 Fair four-level comparison between $S(G)$ and modularity Q .

μ	S_{raw}	Q_{raw}	S_{norm}	Q_{norm}	p_S	p_Q
0.05	0.9811	0.8484	39.79	108.44	0.0050	0.0050
0.10	0.9732	0.7992	38.70	101.34	0.0050	0.0050
0.20	0.9408	0.6984	38.23	82.48	0.0050	0.0050
0.30	0.8769	0.5992	34.91	61.04	0.0050	0.0050
0.40	0.7542	0.4992	28.11	42.64	0.0050	0.0050
0.50	0.5078	0.3902	14.47	20.90	0.0050	0.0050
0.55	0.4049	0.3562	8.81	14.06	0.0050	0.0050
0.60	0.3189	0.3191	4.24	6.75	0.0313	0.0062
0.65	0.2739	0.3029	1.76	3.39	0.1386	0.0254
0.70	0.2590	0.2908	0.92	0.74	0.2639	0.2841

S_{raw} and Q_{raw} are computed on 20 newly generated instances per μ under the balanced null-calibration protocol ($R = 200$, $B = 20$) and are therefore not numerically identical to the instance-averaged $S(G)$ in Supplementary Table S4 (see the note thereto); qualitative conclusions, rankings, and significance results are unaffected. Both metrics receive identical null-model calibration.

Table 7 Influence of the Monte Carlo repetition count on $S(G)$ estimation accuracy.

B	$S(G)$	S_{std}	CV (%)	SE_{CV}	Time per run (s)
5	0.7506	4.089×10^{-3}	0.545	0.099	0.273
10	0.7512	4.001×10^{-3}	0.533	0.097	0.557
20	0.7518	2.198×10^{-3}	0.292	0.053	1.108
30	0.7519	2.457×10^{-3}	0.327	0.060	1.688
40	0.7504	2.033×10^{-3}	0.271	0.049	2.220
50	0.7514	1.721×10^{-3}	0.229	0.042	2.762
60	0.7517	1.534×10^{-3}	0.204	0.037	3.364
70	0.7520	1.398×10^{-3}	0.186	0.034	3.984
80	0.7514	1.291×10^{-3}	0.172	0.031	4.344
90	0.7518	1.325×10^{-3}	0.176	0.032	5.165
100	0.7516	1.375×10^{-3}	0.183	0.033	5.886

Evaluated on a representative SBM network at $\mu = 0.40$; S_{std} is the standard deviation of the $S(G)$ estimate, CV its coefficient of variation, and SE_{CV} the standard error of the CV across repetitions.

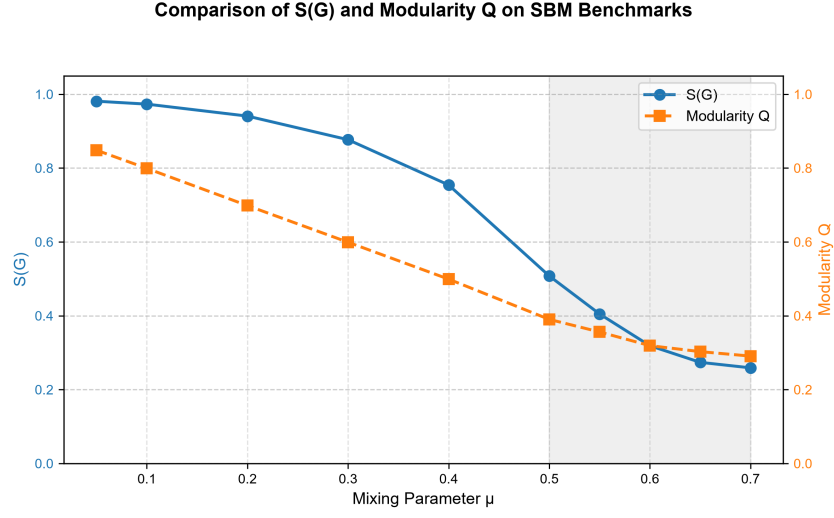


Fig. 2 Comparison of $S(G)$ and modularity Q across the SBM mixing parameter gradient ($N = 200$, $C = 10$, $\langle k \rangle = 10$, Louvain detector): both metrics decrease monotonically with increasing μ and intersect in the weak-community regime ($\mu \approx 0.60$).

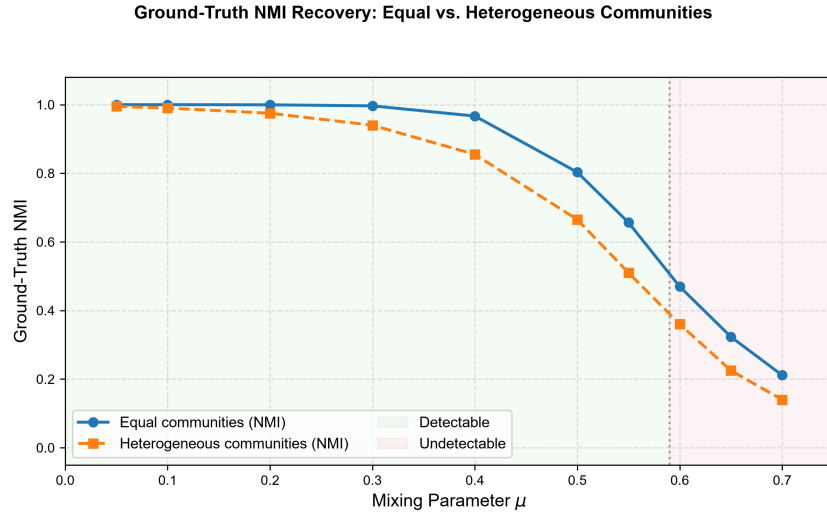


Fig. 3 Ground-truth NMI recovery for equal-size vs. heterogeneous-community SBM.

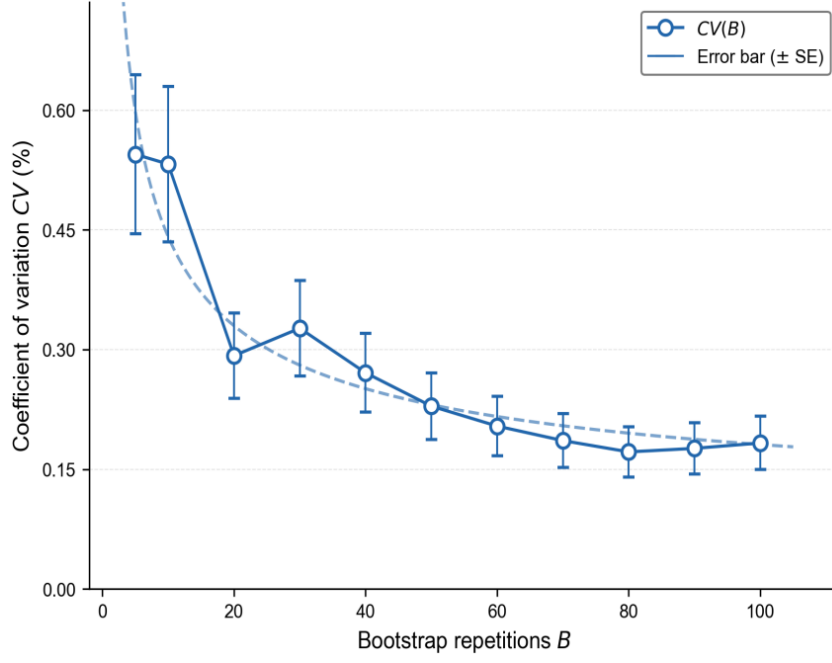


Fig. 4 Influence of the Monte Carlo repetition count B on $S(G)$ estimation accuracy.

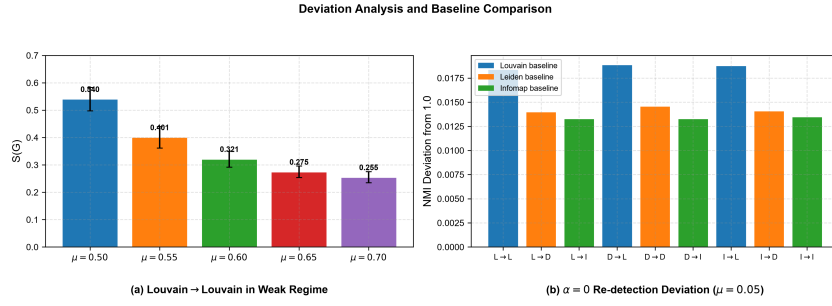


Fig. 5 Weak-community discriminability (left) and $\alpha = 0$ baseline comparison (right).

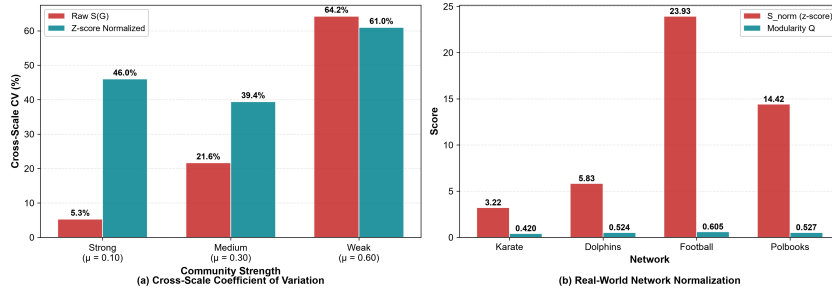


Fig. 6 Z-score scale normalization results. (a) Cross-scale coefficient of variation across three community strength levels. (b) Normalized scores for real-world networks.

Table 8 Parameter settings of the network- and community-scale sensitivity experiment.

Parameter	Condition (a)	Condition (b)
Number of communities C	10	4, 8, 12, 20, 30, 40
Community size n_c	20, 40, 60, 100, 150, 200	50
Total nodes N	200, 400, 600, 1000, 1500, 2000	200, 400, 600, 1000, 1500, 2000
Expected average degree $\langle k \rangle$	10	10
Mixing parameter μ	0.10, 0.30, 0.60	0.10, 0.30, 0.60
Instances n	5	5
Perturbation intensity α	0.05, 0.10, 0.20, 0.30, 0.50	0.05, 0.10, 0.20, 0.30, 0.50
Monte Carlo repetitions B	20	20
Community detection algorithm	Louvain (5 runs, best)	Louvain (5 runs, best)

Condition (a): fixed number of communities, growing community size. Condition (b): fixed community size, growing number of communities.

Table 9 Variation in $S(G)$ with network and community scaling.

N	Condition (a)			Condition (b)		
	$\mu = 0.10$	$\mu = 0.30$	$\mu = 0.60$	$\mu = 0.10$	$\mu = 0.30$	$\mu = 0.60$
200	0.9686	0.8662	0.3139	0.9003	0.5555	0.2630
400	0.9500	0.8001	0.2155	0.9334	0.7407	0.1802
600	0.9308	0.7385	0.1430	0.9387	0.7732	0.1562
1000	0.9036	0.6443	0.1073	0.9452	0.7976	0.1372
1500	0.8701	0.5289	0.0822	0.9460	0.8071	0.1333
2000	0.8350	0.4664	0.0620	0.9497	0.8148	0.1214

Entries are means over 5 instances per configuration.

Table 10 Cross-scale coefficient of variation before and after z-score normalization.

Community strength	Raw $S(G)$ CV	Z-score normalized CV
Strong ($\mu = 0.10$)	5.3%	46.0%
Medium ($\mu = 0.30$)	21.6%	39.4%
Weak ($\mu = 0.60$)	64.2%	61.0%

The cross-scale CV is computed over the six network scales of Supplementary Table S8 at each strength level.

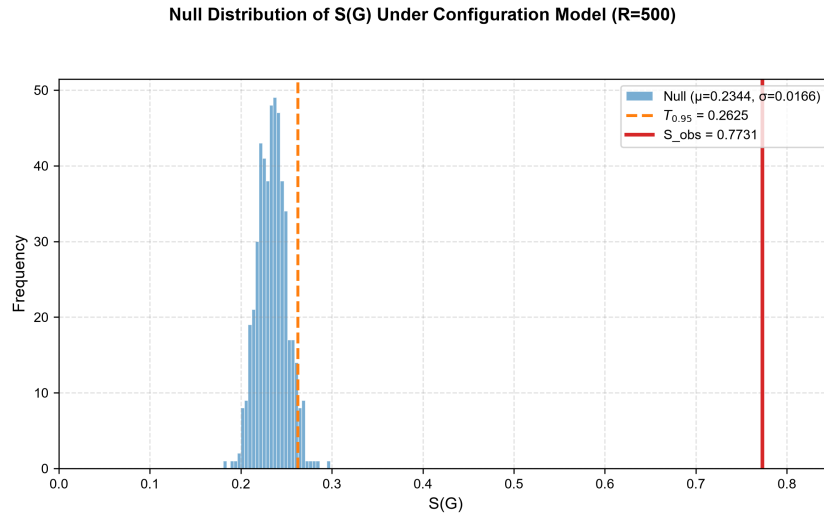


Fig. 7 Null distribution of $S(G)$ for configuration-model random networks.

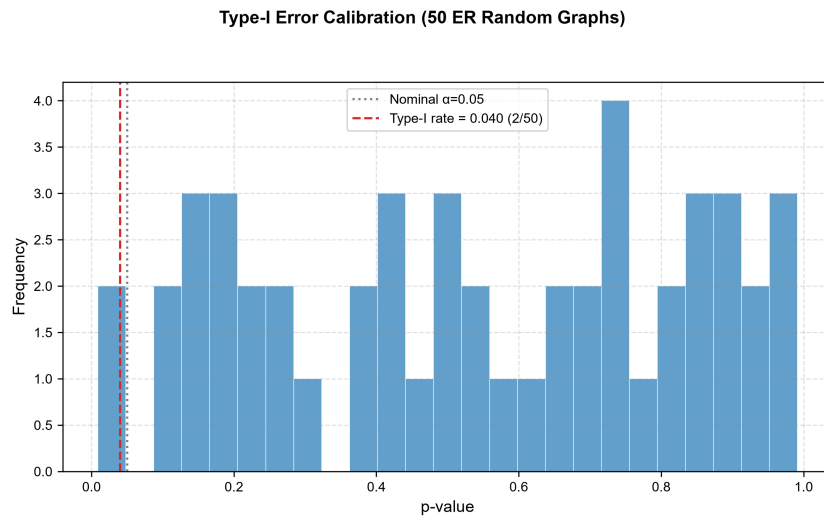


Fig. 8 Type-I error stability across ER random graph realizations.