



Graph perturbations and corresponding spectral changes in Internet topologies



Bo Jiao^{a,*}, Jian-mai Shi^b

^a Luoyang Electronic Equipment Test Center, Luoyang 471003, China

^b College of Information System and Management, Changsha 410073, China

ARTICLE INFO

Article history:

Received 13 March 2015

Revised 28 November 2015

Accepted 28 November 2015

Available online 4 December 2015

Keywords:

Weighted spectral distribution

Evolving topology

Autonomous system (AS)

Internet topology

Normalized Laplacian spectrum

ABSTRACT

The normalized Laplacian spectrum (NLS) is a powerful tool for comparing graphs with different sizes. Recently, we showed that two NLS features, namely the weighted spectral distribution (WSD) and the multiplicity of the eigenvalue 1 (ME1), are particularly relevant to the Internet topology at the inter-domain level. In this paper, we examine the physical meaning of the two metrics for the Internet. We show that the WSD reflects the transformation from single-homed nodes to multi-homed nodes for better fault-tolerance and that the ME1 quantifies the initial star-based structure associated with node classification, both of which are critical to the robustness of the Internet structure. We then investigate the relation between the metrics and graph perturbations (i.e., small changes in a graph). We show that these two NLS metrics can be a good choice for study on the Internet optimization. Our work reveals novel insights into the Internet structure and provides useful knowledge for statistical analysis on complex networks.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Graph spectra (i.e., eigenvalue sets of several matrices, including adjacency, Laplacian and normalized Laplacian matrices) contain a large amount of information on graph structure [1,2]. In contrast to the spectral radii (i.e., maximum eigenvalues) of adjacency and Laplacian matrices, Cetinkaya et al. demonstrated that the normalized Laplacian spectral radius is a good indicator of the connectivity when comparing networks with different sizes [3,4]. The Internet topology evolves over time [5–7]; i.e., metrics weakly related to network size are more important.

The Positive-Feedback Preference (PFP) model [8] is an Internet topology generator based on the Interactive growth mechanism whose inputs are not related to network size and only correspond to mechanisms for the addition of nodes and links. Through comparisons with other commonly used metrics, Jiao et al. [9] found that only the WSD and the ME1 of the NLS are independent of network size and provide sensitive discriminations for each input of the generator PFP; i.e., the two metrics are suitable for capturing evolving features of Internet topologies. Therefore, it is of value to study the relation between graph perturbations in Internet topologies and corresponding changes of the two spectral metrics.

The eigenvectors associated with the largest eigenvalues of the adjacency spectrum can be used to analyze clustering features of the Internet [10], and the eigenvalue power law of the spectrum in the Internet is a direct consequence of the degree distribution of a star-based structure [11]. A metric based on the adjacency spectrum is the natural connectivity [12], which represents the weighted sum of closed walks of all lengths. The second smallest eigenvalue of the Laplacian spectrum, also known as the algebraic connectivity [13], reflects how well connected the overall graph is. Compared with the metrics based on the adjacency and Laplacian spectra, the metrics based on the NLS are more suitable for the comparative analysis of networks with different sizes [3,4,9]. The WSD (i.e., a metric of the NLS) counts the number of N -cycles normalized by the degree of each node in the cycle, which is strongly associated with the distribution of random walk cycles in a network [14]. Fay et al. used the WSD to distinguish different types of graphs, e.g., graphs from network topology generators, Internet application graphs and dK-random graphs [15]. Additionally, Jiao et al. designed an algorithm using the 4-cycle graph structure to calculate the WSD [16]. The NLS eigenvalues lie in the range of zero to two [14] and are quasi-symmetrical about 1 [3,4], while the ranges of the adjacency and Laplacian eigenvalues are not limited [17]. The largest NLS eigenvalues obey a power law [18], and the ME1 of the NLS remains stable over time in spite of the explosive growth of the Internet [19,20]. AS graphs simulated by the Internet topology generator Inet [21] and real topologies were compared and analyzed by Vukadinovic et al. [19,20]. Additionally, they observed that there is a significant difference in the ME1 of the NLS

* Corresponding author. Tel. +86 15138758790.

E-mail address: jiaoboleetc@outlook.com (B. Jiao).

and related the lower bound of the ME1 to a structural classification of nodes in AS graphs [19,20]. Haddadi et al. [22] used the WSD to tune input parameters of Internet topology generators. However, the relation between graph perturbations in AS-level Internet topologies and corresponding changes of the NLS has not been further investigated.

The existing Internet topology generators (e.g., Waxman [23], BA [24], GLP [25], Inet-3.0 [26] and PFP [8]) were compared using the metrics of degree distribution, average neighbor connectivity, clustering coefficient, rich club connectivity, shortest path distribution and WSD [27]. Based on the work of Haddadi et al. [27], Inet-3.0 and PFP outperform other generators because Waxman and BA are too simplistic in the assumptions they make about the connectivity of graphs and only Inet-3.0 and PFP generate topologies with a sufficiently dense core. Additionally, the existing topology generators fail to capture local connectivity, and the metrics such as clustering and WSD should be focused on for tuning and optimizing the topology generators [27].

In this paper, our contributions focus on the relation between graph perturbations in AS-level Internet topologies and corresponding changes of the WSD and the ME1 of the NLS. According to the relation, we can further understand structural characteristics of the Internet topology, and the spectral features of synthetic Internet AS graphs can be tuned and optimized.

2. Normalized Laplacian spectrum

Denote by $G = (V, E)$ an undirected and simple graph, where V is the node set and E is the edge set. Let $\|V\| = n$, $\|E\| = m$, and d_v be the degree of node v in graph G . The normalized Laplacian of graph G is the matrix $L(G)$ defined as follows:

$$L(G)(u, v) = \begin{cases} 1 & \text{if } u = v \text{ and } d_v \neq 0 \\ -\frac{1}{\sqrt{d_u \cdot d_v}} & \text{if } u \text{ and } v \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

If A is the adjacency matrix of graph G (where $a_{ij} = 1$ if there is an edge between v_i and v_j and $a_{ij} = 0$ otherwise) and D is a diagonal matrix with $d_{ii} = d_{v_i}$, $L(G) = D^{-1/2}(D - A)D^{-1/2}$. The NLS is the eigenvalue set of $L(G)$: $0 = \lambda_1 \leq \dots \leq \lambda_n \leq 2$. To find a lower bound for the ME1, Vukadinovic et al. classified nodes in AS graphs into $P(G) = \{v \in V | d_v = 1\}$ called “pendants”, $Q(G) = \{v \in V | \exists w, (v, w) \in E, w \in P(G)\}$ called “quasi-pendants” and $R(G) = V \setminus (Q(G) \cup P(G))$ called “inners” [19,20]. Let p, q, r respectively be the cardinalities of the sets $P(G), Q(G)$ and $R(G)$. Denote by $Inner(G)$ the subgraph of G induced by $R(G)$. Let inn be the number of isolated nodes in $Inner(G)$. Then, Vukadinovic et al. demonstrated the lower bound of the ME1 [19,20]:

$$ME1 \geq p - q + inn. \quad (2)$$

Let $Inner(G) = (V_I, E_I)$ and $d_I(v)$ be the degree of node v in subgraph $Inner(G)$. Jiao et al. [9] further divided $Inner(G)$ into $PI = \{v \in V_I | d_I(v) = 1 \wedge \forall (v, w) \in E_I, d_I(w) > 1\}$ called “inner pendants”, $QI = \{v \in V_I | \exists w, (v, w) \in E_I, w \in PI\}$ called “inner quasi-pendants”, $RI = \{v \in V_I | d_I(v) \geq 2 \wedge \forall (v, w) \in E_I, w \in QI\}$ called “inner restricted nodes”, $BI = \{v \in V_I | d_I(v) = 1 \wedge \forall (v, w) \in E_I, d_I(w) = 1\}$ called “inner binate nodes”, $II = \{v \in V_I | d_I(v) = 0\}$ called “inner isolated nodes” and $OI = V_I \setminus SI$ called “inner other nodes” where $SI = PI \cup QI \cup RI \cup BI \cup II$ called “inner summation nodes”. Let $pi, qi, ri, bi, ii, si, oi$ respectively be the cardinalities of the sets PI, QI, RI, BI, II, SI and OI . Then, Jiao et al. confirmed the following approximate equation relationship [9]:

$$r \approx si, \quad (3)$$

$$ME1 \approx p - q + pi + ri - qi + ii. \quad (4)$$

To find a collective metric of the NLS, Fay and Haddadi et al. defined the WSD as the sum over all N -cycles $C = u_1 u_2 \dots u_N$ in graph G [14,15,22,27]:

$$W(G, N) = \sum_{i=1,2,\dots,n} (1 - \lambda_i)^N = \sum_C \frac{1}{d_{u_1} d_{u_2} \dots d_{u_N}}. \quad (5)$$

Calculating the eigenvalues of a large (even sparse) matrix is computationally expensive. Therefore, Fay and Haddadi et al. [14,15,22,27] estimated the distribution of eigenvalues $f(\lambda = \theta)$ using Sylvester’s Law of Inertia [28] to calculate the percentage of eigenvalues that fall in a given interval, and the WSD can be transformed into:

$$W(G, N) \approx \sum_{\theta \in \Omega} (1 - \theta)^N \cdot f(\lambda = \theta). \quad (6)$$

where Ω denotes equally spaced bins in $[0, 2]$. According to Eqs. (5) and (6), the WSD is weakly related to the ME1. In general, the sum of 4-cycles with the normalization of node degrees represents the quasi-randomness of a graph in the non-regular case [15]. Thus, Fay and Haddadi et al. [14,15] considered four to be the suitable value of parameter N .

3. Geographical decomposition of the large-scale topology

An intuitive application of the relation studied in this paper is to optimize the NLS features in synthetic Internet AS graphs. The core problem of the optimization is to select an appropriate cost function (e.g., spectral radius [29], algebraic connectivity [30] and weighted spectrum [22]) that reflects those aspects of the graph that are important to the user. Such a selection process is inherently subjective; i.e., there is no “best” cost function in general [22]. Once a suitable cost function is selected, the real meaning of “network optimization” is to tune synthetic graphs to produce output that optimally matches said cost function with minimum changes.

To verify the correctness of our contributions, we consider the WSD and the ME1 of the NLS as the cost function to optimize synthetic AS graphs. Please note that the main theme of this paper is to study the graph perturbation properties of the WSD and the ME1. The “network optimization” is one example for the application of our contributions. Our observed real topology dataset comes from the ITDK of CAIDA Skitter project [31,32], which is obtained by running traceroutes over a large range of IP addresses and mapping the prefixes to AS numbers using RouteViews BGP data. Because the Skitter data represents paths that have actually been traversed by packets to their destinations, rather than paths calculated and propagated by the BGP system, it is more likely to faithfully represent the AS topology than the BGP data alone.

To compare two AS graphs, the WSD and the ME1 only need to be calculated two times, while they have to be calculated many times to lead the optimization of the NLS. The ITDK provides mappings from routers to regions and mappings from routers to AS nodes. Thus, the AS subgraphs of different regions can be constructed where the connection relationships between AS nodes are built by the links between routers located in a certain region. AS nodes may span across regions; i.e., some AS nodes may be allocated into different AS subgraphs. The Chinese AS subgraph with 84 nodes has been obtained by Zhou et al. [33], who compared the small-scale graph with the global topology and found that the Chinese AS-level topology preserves well structural characteristics of the global Internet. The connection patterns between routers are strongly related to geographical distance, and all AS nodes exceeding a certain threshold in size are maximally dispersed geographically [34], which is an important evidence for the study of Zhou et al. [33]. Additionally, the Chinese AS graph with 84 nodes has been utilized for comparisons between the existing topology generators [27]. Therefore, the geographical decomposition of the global Internet topology is an effective method to decrease the time for calculating the WSD and the ME1.

Table 1
Some AS subgraphs obtained by ITDK201110–ITDK201304.

Geographic location	Month/year	Node number	Edge number	WSD	WSD ^f
USA (US)	Oct. 2011	6331	13147	0.02812	0.01424
USA (US)	Jul. 2012	10552	20023	0.02585	0.01310
USA (US)	Apr. 2013	10918	20730	0.02502	0.01270
USA region (US. 6765)	Apr. 2013	2511	6689	0.01779	0.00890
USA region (US. 4242)	Apr. 2013	2412	6127	0.00953	0.00456
Russian (RU)	Apr. 2013	3110	5225	0.06617	0.03403
Deutschland (DE)	Apr. 2013	1469	2321	0.04619	0.02343
Brazil (BR)	Apr. 2013	1282	2340	0.05080	0.02605
Ukraine (UA)	Apr. 2013	1200	1637	0.09888	0.05037

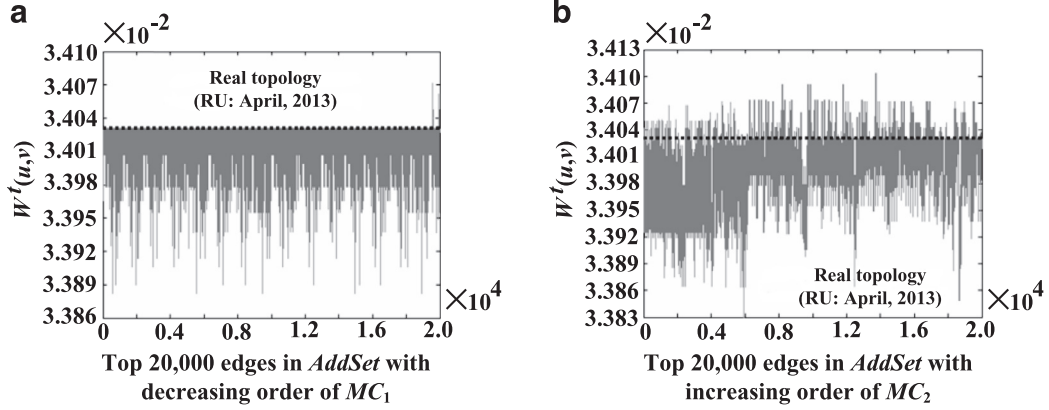


Fig. 1. Effects of adding an edge in the real AS subgraph of RU in the snapshot of April 2013. (a) $W^t(u,v)$ with decreasing order of MC_1 . (b) $W^t(u,v)$ with increasing order of MC_2 .

The NLS eigenvalues are quasi-symmetrical about 1 [3,4] and $(1-\theta)^N$ tends to zero with θ close to 1 and increasing N . Thus, the WSD is strongly related to the WSD^f derived from Eq. (6), where $\Omega \in \{(2(i-1)/k, 2i/k)\}_{i=1}^t$, $\theta = (2i-1)/k$ and $t < k/2$. The parameter k represents the number of equally spaced bins in $[0, 2]$, which depends on the granularity required.

Table 1 lists some subgraphs with the largest number of nodes and shows that the graph of the US in the snapshot of April 2013 can be further decomposed into various US regions. If $k = 30$ and $t = 10$, the WSD is approximately equal to $2WSD^f$, as shown in Table 1. Thus, the WSD can be replaced by WSD^f for faster calculation [9]. According to the quasi-symmetry of eigenvalues about 1 [3,4], the eigenvalues from 0 to 1 are sufficient for the description of the WSD [14,22,27], and we can determine that $ME1 \approx n - 2 \cdot (1 - eps)^{-1}$ [9] where $eps = 10^{-6}$ and $(1 - eps)^{-1}$ is the number of eigenvalues smaller than $1 - eps$ calculated by Sylvester's Law of Inertia [28].

4. Graph perturbations with WSD^f

The WSD is weakly related to the ME1; i.e., graph perturbations with WSD^f should be under the constraint that the ME1 is unaltered. Based on Eq. (4), the constraint means that $stavalue = p - q + pi + ri - qi + ii$ should be constant. Let d_i and d_j be the degrees of nodes i and j , respectively, and $f(d_i)$ and $f(d_j)$ be the frequencies of those degrees. Winick [26] defined the weighted value of d_j with respect to d_i as

$$M(i, j) = \max \left(1, \sqrt{\left(\log \frac{d_i}{d_j} \right)^2 + \left(\log \frac{f(d_i)}{f(d_j)} \right)^2} \right) \cdot d_j. \quad (7)$$

which has been used for calculating the probability linking node i to node j in the topology generator Inet-3.0 [26]. In subgraph $Inner(G)$, let $Nr(v)$ be the set of nodes attached to node v . According to the notations defined in Section 2, we divide PI into $PI_1 = \{v \in PI | \exists w \in QI, v = \arg_{u \in Nr(w) \cap PI} M(u, w)\}$ and $PI_2 = PI \setminus PI_1$; i.e., $Card(PI_1) =$

qi (denote by $Card(\cdot)$ the cardinality of a set). In subgraph $Inner(G) = (V_I, E_I)$, QI will not be altered if the connection relationships of nodes in PI_1 have not been changed.

Theorem 1. For an Internet AS graph $G = (V, E)$, $stavalue = p - q + pi + ri - qi + ii$ is constant when (see Appendix A for the proof)

- (i) Add an edge $(u, v) \notin E$, where $u \neq v$, $u \in Q(G)$ and $v \in Q(G) \cup V_I$.
- (ii) Delete an edge $(u, v) \in E$, where $u \neq v$, $u \in Q(G)$, $d_u > 2$, $v \in Q(G) \cup V_I$, $d_v > 2$ if $v \in Q(G)$, $Card(\{(w, v) \in E | w \in Q(G)\}) > 1$ if $v \in PI$, and $Card(\{(w, v) \in E | w \in Q(G)\}) > 2$ if $v \in II$.
- (iii) Add an edge $(u, v) \notin E$, where $u \neq v$, $u \in QI$ and $v \in PI_2 \cup RI \cup II$.
- (iv) Delete an edge $(u, v) \in E$, where $u \neq v$, $u \in QI$, $d_I(u) > 2$, $v \in PI_2 \cup RI$, $Card(\{(w, v) \in E | w \in Q(G)\}) > 1$ if $v \in PI_2$, and $Card(\{(w, v) | w \in Q(G)\}) > 0$ if $v \in RI \wedge Card(\{(w, v) \in E | w \in QI\}) = 2$ (denote by $d_I(u)$ the degree of node u in $Inner(G)$). \square

Denote by $AddSet$ the edge set to be added associated with the above operators (i) and (iii) and by $DelSet$ the edge set to be deleted associated with the above operators (ii) and (iv). Based on Eq. (4), the ME1 can be controlled if graph perturbations with WSD^f are restricted in $AddSet$ and $DelSet$. The real AS subgraph of RU listed in Table 1 is used for the following illustration.

Let $W^t(u, v)$ be the WSD^f of a graph induced by adding or deleting an edge $(u, v) \in AddSet \cup DelSet$ for AS graph G . We sort edges $(u, v) \in AddSet$ using two metrics, $MC_1 = |d_u - d_v|$ and $MC_2 = \max(d_u, d_v)$. The $W^t(u, v)$ curves of the two metrics are shown in Fig. 1.

Edges with top largest MC_1 link two nodes with high-small degrees, and edges with top smallest MC_2 link two nodes with small-small degrees. According to Table 1 and Fig. 1, connecting two nodes with high-small degrees can steadily decrease WSD^f . Because k -degree node u corresponds to at least k^2 4-cycles (i.e., $u-v_i-u-v_j-u$ with $ij = 1, 2, \dots, k$) where v_1, v_2, \dots, v_k are all nodes attached to node u , the 4-cycle number of a high-degree node is extremely large compared to that of a small-degree node. Based on Eq. (5), WSD^f is

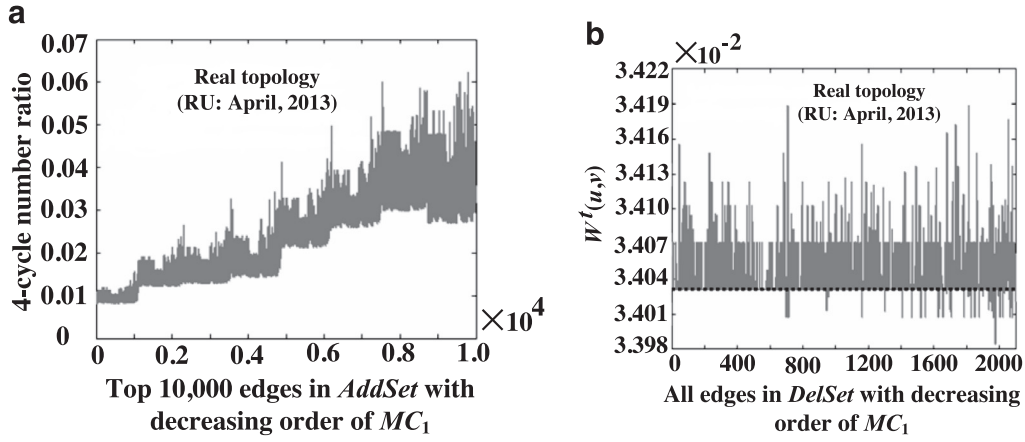


Fig. 2. 4-cycle number ratio and edge deletion effect. (a) Ratio of the number of newly increased 4-cycles induced by adding an edge to the original 4-cycle number of the two connected nodes. (b) Effect of deleting an edge: $W^t(u,v)$ of all edges in $DelSet$ with decreasing order of MC_1 .

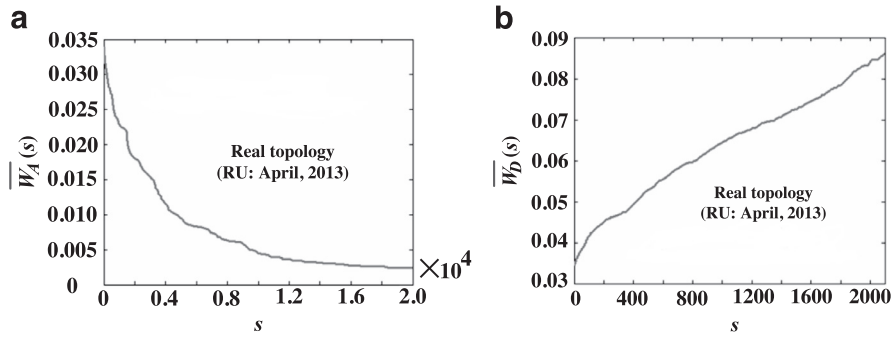


Fig. 3. $\bar{W}_A(s)$ and $\bar{W}_D(s)$ curves in the real AS subgraph of RU in the snapshot of April 2013. (a) $\bar{W}_A(s)$ with the consecutive addition of edges. (b) $\bar{W}_D(s)$ with the consecutive deletion of edges.

decided by the 4-cycle number and node degrees. For a high-degree node, the degree increasing by 1 extremely decreases the sum over all 4-cycles (normalized by the node degrees) of that node. For edges with high-small degrees, Fig. 2(a) shows the ratio of the number of newly increased 4-cycles induced by adding an edge to the original 4-cycle number of the two connected nodes. The very small ratio shown in Fig. 2(a) provides the evidence for Fig. 1(a). For an edge with small-small degrees, both the number of newly increased 4-cycles induced by adding the edge and the original 4-cycle number of the two connected nodes are small, which induces the uncertainty of WSD^t , as shown in Fig. 1(b).

Deleting an edge is an inverse process of adding an edge; i.e., shearing an edge with high-small degrees tends to increase WSD^t . The number of edges with small-small degrees in $DelSet$ is small because small-degree nodes tend to link high-degree nodes based on the attachment models of Internet topologies [26]. Thus, all edges in $DelSet$ are included in Fig. 2(b).

Denote by $AddSet(T_A)$ the top T_A edges in $AddSet$ with decreasing order of MC_1 . Sort the edges with increasing order of $W^t(u, v)$, and let $AddSet(T_A) = \{e_i^a\}_{i=1}^{T_A}$ where $W^t(e_1^a) \leq W^t(e_2^a) \leq \dots \leq W^t(e_{T_A}^a)$. Simultaneously, denote by $DelSet(T_D)$ the top T_D edges in $DelSet$ with decreasing order of MC_1 . Let $DelSet(T_D) = \{e_i^d\}_{i=1}^{T_D}$ where $W^t(e_1^d) \geq W^t(e_2^d) \geq \dots \geq W^t(e_{T_D}^d)$. Denote by $\bar{W}_A(s)$ the WSD^t of a graph induced by adding a series of edges $\{e_i^a\}_{i=1}^s$ to AS graph G and by $\bar{W}_D(s)$ the WSD^t of a graph induced by deleting a series of edges $\{e_i^d\}_{i=1}^s$ from AS graph G . Please note that the consecutive deletion should not alter nodes in sets $P(G)$ and QI ; otherwise, the $stavalue$ value may not be constant. If the deletion of edge e_s^d alters nodes in $P(G)$ and QI , add the edge to AS graph G again and let $\bar{W}_D(s) = \bar{W}_D(s-1)$. The $\bar{W}_A(s)$ and $\bar{W}_D(s)$ curves in Fig. 3 show that WSD^t monotonically decreases or increases.

Denote by $\bar{M}_A(s_A)$ the ME1 of a graph induced by adding a series of edges $\{e_i^a\}_{i=1}^{s_A}$ and by $\bar{M}_D(s_D)$ the ME1 of a graph induced by deleting a series of edges $\{e_i^d\}_{i=1}^{s_D}$. The consecutive deletion should not alter nodes in sets $P(G)$ and QI . The statistical values of $\bar{M}_A(s_A)$ and $\bar{M}_D(s_D)$, where $s_A \in \{1, 2, \dots, Card(AddSet)\}$ and $s_D \in \{1, 2, \dots, Card(DelSet)\}$, are listed in Table 2, which verifies the invariability of the ME1.

In general, it is not necessary to decrease the WSD^t of an AS graph by an extreme amount; i.e., the parameter T_A may be obviously smaller compared to the scenario in Fig. 3(a). According to the analyses of Fig. 4, a lower bound T_{low} (e.g., 5000 for RU and 1000 for DE) should be given for the T_A value; i.e., with a smaller value of $T_A (\geq T_{low})$, the $\bar{W}_A(s)$ curve corresponds to a faster descent speed (because edges with large values of MC_1 are sorted ahead in the edge set $\{e_i^a\}_{i=1}^{T_A}$). A faster descent speed of the curve implies that graph perturbations of AS graph G may be much smaller with a given WSD^t .

5. An algorithm of graph perturbations with a given WSD^t

Inputs: An AS graph $G = (V, E)$, a given $WSD^t (W_{obj})$ and two given thresholds (T_{low}, T_{upper}).

Step 1. Let W_{org} be the WSD^t value of AS graph G . If $W_{obj} > W_{org}$, calculate the edge set $DelSet(T_D) = \{e_i^d\}_{i=1}^{T_D}$, where $W^t(e_1^d) \geq \dots \geq W^t(e_{T_D}^d)$ and $T_D = Card(DelSet)$, and use the bisection method (because $\bar{W}_D(s)$ monotonically increases) to calculate

$$s_{obj} = \arg_{s=1, \dots, T_D} \min |\bar{W}_D(s) - W_{obj}|. \quad (8)$$

Then, delete the edges $\{e_i^d\}_{i=1}^{s_{obj}}$ from AS graph G and conclude the algorithm.

Table 2

Statistical values of the ME1 with the consecutive addition (deletion) of edges in *AddSet* (*DelSet*).

Geographic location	$Card(AddSet)$	Mean value	Variance	$Card(DelSet)$	Mean value	Variance
RU	586015	1908	0	2098	1908	0.2
US. 6765	344873	1805	0	4044	1805	5.9
DE	88153	1029	0	791	1029	0.1

Table 3

Statistical analyses for graph perturbations in the real AS subgraph of RU, where W_{obj} corresponds to 9 given WSD^t . Denote by W_{per} the WSD^t of the perturbed graph, by R^{WSD} the ratio of $|W_{per} - W_{org}|$ to W_{org} , and by R^{Edges} the ratio of m_{alter} to m_{org} , where m_{alter} is the number of added (or deleted) edges, and W_{org} and m_{org} are the WSD^t and the edge number of the real AS subgraph of RU, respectively.

$W_{obj} \times 10^2$	1.0	1.5	2.0	2.5	3.0	4.0	5.0	6.0	7.0
$W_{per} \times 10^2$	1.0003	1.5001	2.0014	2.4982	3.0002	3.9981	5.0007	5.9998	7.0024
R^{WSD}	0.7061	0.5592	0.4119	0.2659	0.1184	0.1749	0.4695	0.7631	1.0577
R^{Edges}	0.7474	0.3411	0.1667	0.0789	0.0260	0.0147	0.0731	0.1340	0.2193
$T_A(T_D)$	5000	5000	5000	5000	5000	2098(T_D)	2098(T_D)	2098(T_D)	2098(T_D)
ME1	1908	1908	1908	1908	1908	1908	1908	1908	1913

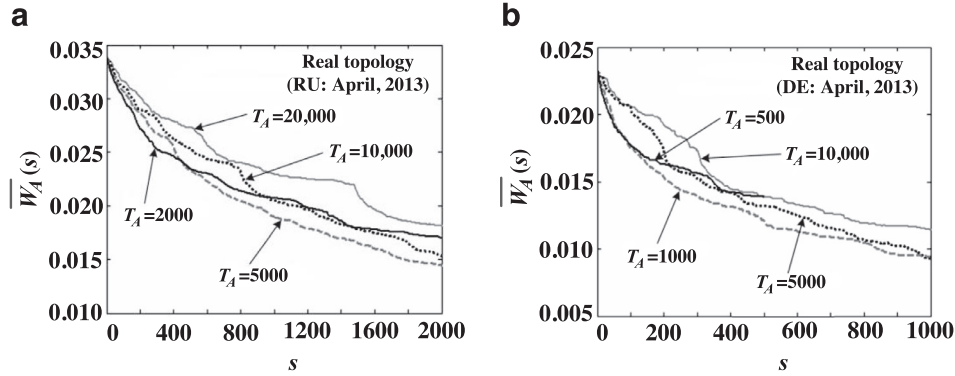


Fig. 4. $\bar{W}_A(s)$ curves with different values of T_A . (a) RU in April 2013. (b) DE in April 2013.

If $W_{obj} < W_{org}$, calculate the edge set *AddSet*, let $T_A := T_{low}$, and go to **Step 2**.

Step 2. Calculate $AddSet(T_A) = \{e_i^{T_A}\}_{i=1}^{T_A}$, where $W^t(e_1^{T_A}) \leq \dots \leq W^t(e_{T_A}^{T_A})$. If $\bar{W}_A(T_A) < W_{obj}$, use the bisection method (because $\bar{W}_A(s)$ monotonically decreases) to calculate

$$s_{obj} = \arg \min_{s=1, \dots, T_A} |\bar{W}_A(s) - W_{obj}|. \quad (9)$$

Then, add the edges $\{e_i^{s_{obj}}\}_{i=1}^{s_{obj}}$ to AS graph G , and conclude the algorithm.

If $\bar{W}_A(T_A) \geq W_{obj}$ and $T_A \leq T_{upper}$, let $T_A := T_A + 1000$ and go to **Step 2**. If $T_A > T_{upper}$, conclude the algorithm and AS graph G cannot be perturbed with W_{obj} . \square

The time complexity of the algorithm depends on the calculation of $\{W^t(e_i^{T_A})\}_{i=1}^{T_A}$ (or $\{W^t(e_i^{T_D})\}_{i=1}^{T_D}$), which is proportional to the parameter T_A (or T_D). The bisection methods used in Eqs. (8) and (9) only correspond to $\bar{W}_D(s)$ and $\bar{W}_A(s)$ calculated $\log_2(T_D)$ and $\log_2(T_A)$ times.

The statistical values of the real AS subgraph RU are listed in Table 3, which implies that there is an extremely small influence on the ME1 for the above algorithm, and the perturbations with a given WSD^t are effective and can control the alteration of connection relationships between nodes.

Perturbation effects of four other metrics with different W_{obj} values are shown in Fig. 5: with descending W_{obj} , the curves in Fig. 5(a)–(c) ascend and the curve in Fig. 5(d) moves toward the y-axis. The phenomena are induced by the increase of edges to decrease W_{obj} .

6. Physical interpretation of the spectral metrics

To explain the eigenvalue power law of Internet AS graphs, Francesc et al. [11] constructed a deterministic star-based model of Internet topologies, as shown in Fig. 6(a). The model has two components: the tree component that contains all nodes that belong exclusively to trees having a depth of one (i.e., star subgraphs) and the core component that contains all roots of the trees [36].

Francesc's model is closer to Internet AS topologies before 1999 because most of the tree-nodes have a degree of one [36]. However, increasingly more single-homed nodes want to become multi-homed for better fault-tolerance [37]. Based on the extension of Francesc's model shown in Fig. 6(b), we illustrate the physical meaning of the WSD and the ME1.

In Fig. 6(b), dotted lines connect various single-homed nodes to nodes in the core component, which provides better fault-tolerance because of the increased number of multi-homed nodes. More multi-homed nodes means that more edges linking high-small degree nodes are added into the AS graph. According to Sections 4 and 5, the consecutive addition of these edges can steadily decrease the WSD. Thus, the WSD plausibly represents the occupancy of multi-homed nodes in AS graphs.

The ME1 shows the classification of nodes in $P(G)$, $Q(G)$, PI , QI , RI , BI and II defined in Section 2. In Fig. 6(b), for a root node in the core component, if all of its child nodes in the tree component are transformed into multi-homed nodes, it tends to fall in QI and BI , such as nodes 18, 19 and 20 in QI and node 17 in BI ; otherwise, it tends to fall in $Q(G)$, such as nodes 14, 15 and 16. Thus,

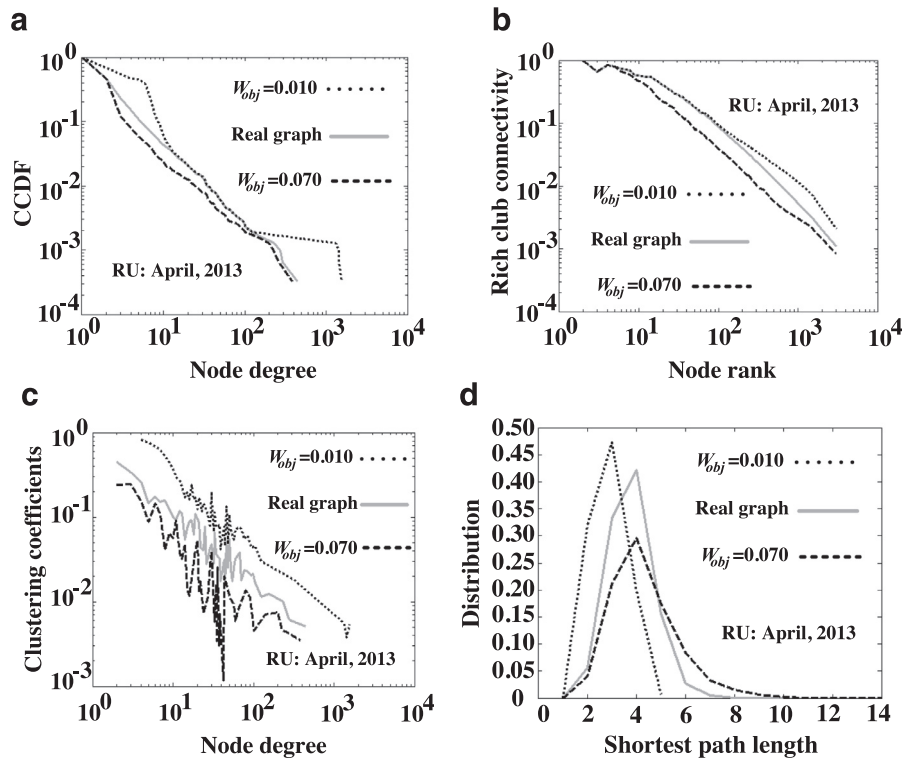


Fig. 5. Perturbation effects with different W_{obj} values. (a) Complementary cumulative distribution function (CCDF) of the node degree [26]: in the log–log scale, the x-axis represents the node degree d , and the y-axis represents the ratio of the number of nodes with degree larger than d to the number of all nodes. (b) Rich club connectivity [8]: in the log–log scale, the x-axis represents the node rank i with decreasing order of degree and the y-axis represents the ratio of the number of edges in the subgraph induced by the top i largest-degree nodes to the maximum possible edges $i(i-1)/2$. (c) Clustering coefficients [35]: in the log–log scale, the x-axis represents the node degree d , and the y-axis represents $C(d) = 2m(d)/(d^2 - d)$, where $m(d)$ is the average number of edges between the neighbors of d -degree nodes. (d) Shortest path length [35]: the x-axis represents the shortest path length l between two nodes, and the y-axis represents the distribution probability of node pairs with distance l .

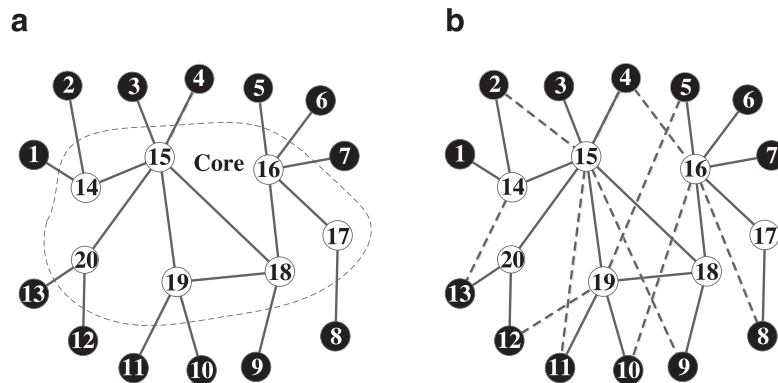


Fig. 6. Star-based model and its extension. (a) Francese's Internet topology model: deleting the connection relationships between nodes in the core component results in a non-connected graph constituted by the union of star subgraphs. (b) Extension of Francese's model with single-homed nodes becoming multi-homed nodes.

the core of an Internet AS graph is composed of nodes in $Q(G) \cup QI$ and part of the nodes in BI . In Fig. 6(b), nodes 9, 10, 11 and 13 fall in PI because they are attached to nodes in $Q(G)$ by dotted lines, while node 12 falls in RI because it is attached to a node in QI by a dotted line. In the real RU AS subgraph in the snapshot of April 2013, there are 155 nodes in PI , which is an extremely larger amount than the 4 nodes in RI because the highest-degree nodes of AS graphs mainly fall in $Q(G)$ and these nodes can attract more nodes to connect with them. In Fig. 6(b), nodes 2 and 4 are attached to nodes in $Q(G)$ by dotted lines, so they fall in II (node 5 falls in PI). The highest-degree nodes in $Q(G)$ induce the number of nodes in II to be large; for example, 599 nodes of the real RU AS subgraph (April 2013) fall in II .

According to the above analyses, the structural classification of the ME1 plausibly reflects the initial star-based structure of the Internet

and the tendency that increasingly more single-homed nodes want to become multi-homed for better fault-tolerance.

Additionally, we use the stochastic model PFP [8] to illustrate the structural classification. PFP has three inputs, where p^* and q^* are probabilities for selecting mechanisms to add 1-degree nodes, $1 - p^* - q^*$ is a probability for selecting one mechanism to add 2-degree nodes, and δ is used to calculate the nonlinear preferential attachment probability. In the evolving Internet topology, most of the newly added small-degree nodes are linked by other nodes with a very small probability. Thus, the newly added 1-degree nodes are the main source of $P(G)$, and the newly added 2-degree nodes are the main source of subgraph $Inner(G)$. The newly added 2-degree nodes tend to link two (or one) quasi-pendants (with high degree and linked by pendants in $P(G)$); i.e., after removing nodes in $P(G)$ and $Q(G)$, the

Table 4

Statistical analyses of six real Internet topologies and the corresponding 6-by-100 synthetic and optimized graphs. The index *ratio* is defined by $\text{Card}((E_{\text{Inet}} \setminus E_{\text{Our}}) \cup (E_{\text{Our}} \setminus E_{\text{Inet}})) / \text{Card}(E_{\text{Inet}})$, where E_{Inet} is the edge set of a synthetic graph generated by Inet-3.0 and E_{Our} is the edge set of the graph with optimized NLS. Both mean value (upper) and variance (bottom) are used to illustrate each statistical value associated with graphs generated by Inet-3.0 and graphs with optimized NLS.

Region		US Oct 2011	US.6765 Apr 2013	RU Apr 2013	DE Apr 2013	BR Apr 2013	UA Apr 2013
Real topologies	Node number	6331	2511	3110	1469	1282	1200
	Edge number	13147	6689	5225	2321	2340	1637
	$m_G(1)$	4497	1805	1908	1029	836	722
Graphs generated by Inet-3.0	Edge number	12103	6218	5352	2296	2293	1683
		6049	2011	2311	549.36	167.62	32.150
	$m_G(1)$	4715	1789	1941	1082	863.08	757.62
		3706	1095	414.354	240.098	221.634	83.0156
	$D^f(G, N) \times 10^3$	2.8340	3.6558	10.1860	4.6443	7.5557	12.7620
Graphs with optimized NLS	Edge number	12218	6229	5167	2318	13243	2.1121
		9083	1958	9135	795.94	1322	618.32
	$m_G(1)$	4497	1805	1908	1029	836	722
		0	0	0	0	0	0
	$D^f(G, N) \times 10^3$	1.5680	2.2189	5.6256	2.8133	5.1026	9.2711
		0.1823	0.2668	1.3106	1.4194	1.6862	2.6282
	<i>Ratio</i>	0.0127	0.0067	0.0605	0.0160	0.0223	0.0255
	0.00003	0.00003	0.0015	0.0001	0.0006	0.0004	

degree of nodes in subgraph $Inner(G)$ corresponds to zero or one with a very high probability, which can explain the structural classification of the ME1.

7. Graph perturbations with the ME1

Based on deterministic and stochastic models of the Internet topology, we have explained the generation process of the structural classification of the ME1. The right-hand side of Eq. (4), i.e., $stavalue = p - q + pi + ri - qi + ii$, can be adjusted using the alteration of connection relationships. Thus, the left-hand side of Eq. (4), i.e., the ME1, can be perturbed in a controllable way. In general, parameters p and q depend on the fraction of 1-degree nodes and the preferential attachment model, which have been paid more attention by existing topology generators, e.g., PFP and Inet-3.0. Therefore, we only consider the adjustment in subgraph $Inner(G)$.

In subgraph $Inner(G)$, divide OI into $OI_1 = \{v \in OI | \text{Card}(Nr(v) \setminus QI) = 1\}$ and $OI_2 = OI \setminus OI_1$, where $Nr(v)$ includes all nodes attached to node v .

The notations used in this section are defined in Sections 2 and 4.

Theorem 2. $stavalue := stavalue - 2$ if adding an edge (v_1, v_2) , where $v_1, v_2 \in II$.

Theorem 3. $stavalue := stavalue - 2$ if adding an edge (v_1, v_2) , where $v_1 \in II, v_2 \in Pl_2$.

Theorem 4. $stavalue := stavalue + 1 - 1$ if adding an edge (v_1, v_2) , where $v_1 \in II$ and $v_2 \in \{v \in OI | \text{Card}(Nr(v) \cap OI_1) = 1\}$.

Theorem 5. $stavalue := stavalue - 1$ if adding an edge (v_1, v_2) , where $v_1 \in Pl_2, v_2 \in OI$.

Theorem 6. $stavalue := stavalue + 2$ induced by the following three operators: (i) delete an edge (v_1, v_2) where $v_1, v_2 \in BI$ and $v_2 = Nr(v_1)$; (ii) if $\text{Card}(\{(v_1, w) \in E | w \in Q(G)\}) = 1$, add an edge (v_1, w_1) where $(v_1, w_1) \notin E$ and $w_1 \in Q(G)$; and (iii) if $\text{Card}(\{(v_2, w) \in E | w \in Q(G)\}) = 1$, add an edge (v_2, w_2) where $(v_2, w_2) \notin E$ and $w_2 \in Q(G)$.

Theorem 7. $stavalue \geq stavalue + 1$ induced by the following four operators: i) let $v_1 \in OI$ and delete all edges $\{(v_1, v_2)\}$, where $v_2 \in Nr(v_1) \setminus QI$; ii) if v_1 is transformed into PI induced by i) and $\text{Card}(\{(v_1, w) \in E | w \in Q(G)\}) = 0$, add an edge (v_1, w_1) where $(v_1, w_1) \notin E$ and $w_1 \in Q(G)$; iii) if v_1 is transformed into II induced by i) and $\text{Card}(\{(v_1, w) \in E | w \in Q(G)\}) < 2$, add one or two edges to ensure $\text{Card}(\{(v_1, w) \in E | w \in Q(G)\}) = 2$; and iv) if v_2 is transformed

into $PI \cup BI$ induced by i) and $\text{Card}(\{(v_2, w) \in E | w \in Q(G)\}) = 0$, add an edge (v_2, w_2) where $(v_2, w_2) \notin E$ and $w_2 \in Q(G)$. (see Appendix A for the proofs)

According to the analyses of Section 6, nodes in $Q(G)$, QI and part of the nodes in BI generate the core of Internet AS graphs; i.e., in Francesc's deterministic model shown in Fig. 6, the periphery of AS graphs is composed of nodes in $P(G)$, PI , RI , II and the rest of the nodes in BI . In Fig. 6, the nodes in BI originate from the star subgraph having only two nodes (one root and one leaf), e.g., nodes 17 and 8. Thus, the core and peripheral nodes plausibly represent equal portions in BI ; i.e., the number of core and peripheral nodes are $q + qi + bi/2$ and $p + pi + ri + ii + bi/2$, respectively. Therefore, $p - q + pi + ri - qi + ii$ can be considered as the peripheral node number minus the core node number, which is a further explanation for the physical meaning of the ME1. Theorems 2–7 exhibit some intuitive methods to adjust the core and peripheral node percentages, which effectively perturbs the ME1 according to our requirement.

8. An application of graph perturbations in the NLS

In this section, our contributions are used for the optimization of synthetic Internet AS graphs generated by Inet-3.0. The core problem of "optimization" is to select a cost function that reflects those aspects of the graph that are important to the user [22,29,30]. We consider the WSD and the ME1 of the NLS as the cost function. Then, the real meaning of "network optimization" is to tune synthetic AS graphs to optimally match the cost function with minimum changes. We use the generator Inet-3.0 because its inputs (i.e., node degree distribution and fraction of 1-degree nodes) are easy to evaluate according to real Internet AS subgraphs obtained in Section 3.

The WSD is a collective metric of the NLS, but Eq. (6) provides a two-dimensional display of the eigenvalue distribution with different weights $(1 - \theta)^N$. To compare two AS graphs G_1 and G_2 , based on Eq. (6), Fay et al. [14] defined a new metric as follows:

$$D(G, N) = \sum_{\theta \in \Omega} (1 - \theta)^N \cdot |f_1(\lambda = \theta) - f_2(\lambda = \theta)| \quad (10)$$

where f_1 and f_2 are the eigenvalue distributions of G_1 and G_2 , respectively, and Ω denotes equally spaced bins in $[0, 2]$. Based on the fact that the WSD can be replaced by WSD^f for faster calculation, $D(G, N)$ can be replaced by $D^f(G, N)$, where $\Omega \in \{(2(i-1)/k, 2i/k)\}_{i=1}^t$, $\theta = (2i-1)/k$, $k = 30$ and $t = 10$. Specifically, $D(G, N) \approx 2 \cdot D^f(G, N)$.

In contrast to WSD^f , $D^f(G, N)$ contains more information. Thus, the $D^f(G, N)$ metric, which measures the difference between a synthetic

graph and the corresponding real Internet AS subgraph, can be used for more accurate optimization of the WSD.

According to the studies of Sections 4–7, we can design the ME1 optimization algorithm (see Appendix B) and the WSD optimization algorithm (see Appendix C). To verify the correctness of our contributions, we use the algorithms to optimize 6-by-100 synthetic AS graphs generated by Inet-3.0 according to the NLS features of six real Internet AS subgraphs. The experimental results of these graphs are shown in Table 4 and Fig. 7, which verify that the optimization of the NLS is effective and that the corresponding perturbations on AS graphs are very small.

The WSD represents the local connectivity of AS graphs [27], which is potentially critical when studying properties of routing protocols. Additionally, based on Sections 6 and 7, we have further explained the physical meaning of the WSD and the ME1: the WSD measures the fault-tolerance of AS graphs because the consecutive addition of edges with high-small degrees increases the multi-homed node percentage and steadily decreases the WSD; the ME1 reflects the decomposition of core and peripheral nodes that is generated by the initial star-based structure of the Internet and the tendency that increasingly more single-homed nodes want to become multi-homed.

Moreover, the WSD and the ME1 of the NLS remain stable in spite of the explosive growth of the Internet [9]. Thus, it is worth selecting the two spectral metrics as the cost function to tune synthetic AS graphs. There is no "best" cost function in general [22,29,30]; i.e., other features of the Internet may not be optimized by the work of this paper. However, our contributions are useful explorations for the study of complex features embedded in the Internet AS topology.

Research to date has analyzed the Internet AS-level topology at a worldwide level of detail, which is useful when the Internet is analyzed at a very coarse level. However, it may be misleading if the analysis is more focused on a specific geographical region [38]. Currently, increasingly more literature has focused on region-level Internet topologies [39,40] because the region competitiveness is reflected in the AS growth pattern. The structural similarities between region graphs and the global AS-level topology have been analyzed in Section 3. Therefore, the subgraphs based on geographical decomposition are sufficient for the research of graph perturbations with the NLS. Additionally, the average time to optimize synthetic AS graphs of the US (Oct 2011 with 6331 AS nodes) is restricted to seven hours, which is tolerable for our computers.

9. Conclusion

This paper focuses on the relation between graph perturbations and corresponding changes in the WSD and the ME1, and analyzes the physical meaning of the two spectral metrics embedded in Internet AS graphs, e.g., node classification, initial star-based structure, multi-homed transformation and core-periphery decomposition. The two metrics reflect the NLS features with eigenvalues not only toward 0 (and 2) but also restricted to 1; i.e., they can be considered as the cost function representing the NLS. Additionally, they are independent of the network size of evolving AS graphs. Therefore, our contributions are useful for understanding the Internet structure and leading future applications of the NLS in AS graphs.

Acknowledgment

We would like to thank the anonymous reviewers for their comments and suggestions, which helped improve this paper, and the CAIDA Skitter project for providing real topology datasets and information for the decomposition of the topology. This paper is supported by the National Natural Science Foundation of China (Grant nos. 61402485 and 71201169).

Appendix A. The proofs

Proof of Theorem 1. Operations (i), (ii), (iii) and (iv) cannot alter the 1-degree node number and the connection relationships of nodes in $P(G)$. Thus, p and q are unaltered. Operators (i) and (ii) are unrelated to subgraph $Inner(G)$, which means that pi , qi , ri , ii are unaltered. Operators (iii) and (iv) cannot alter nodes in QI . Thus, the alterations induced by operator (iii) are restricted to two possible scenarios: either nodes in PI_2 are transformed into RI or nodes in II are transformed into PI_2 ; i.e., $pi + ri + ii$ is unaltered. The alterations induced by operator (iv) are restricted to two possible scenarios: either nodes in PI_2 are transformed into II or nodes in RI are transformed into PI_2 ; i.e., $pi + ri + ii$ is unaltered. Therefore, $stavalue = p - q + pi + ri - qi + ii$ is constant. \square

Proof of Theorem 2. The addition of the edge induces the state that $ii := ii - 2$ and p, q, pi, qi, ri are unaltered; i.e., $stavalue = p - q + pi + ri - qi + ii$ decreases by 2. \square

Proof of Theorem 3. The addition of the edge induces the state that $ii := ii - 1$, $qi := qi + 1$ and p, q, pi, ri are unaltered; i.e., $stavalue = p - q + pi + ri - qi + ii$ decreases by 2. \square

Proof of Theorem 4. The addition of the edge induces the state that $ii := ii - 1$, $pi := pi + 1$, $qi := qi + 1$, $ri := ri + l$ and p, q are unaltered; i.e., $stavalue = p - q + pi + ri - qi + ii$ decreases by 1 if $l = 0$ and increases by $l - 1$ if $l \geq 2$. \square

Proof of Theorem 5. The addition of the edge induces the state that $pi := pi - 1$ and p, q, qi, ri, ii are unaltered; i.e., $stavalue = p - q + pi + ri - qi + ii$ decreases by 1. \square

Proof of Theorem 6. Nodes v_1 and v_2 cannot be transformed into pendants based on the operators (ii) and (iii); i.e., p and q are unaltered. Operator (i) induces the state that $ii := ii + 2$ and pi, qi, ri are unaltered; i.e., $stavalue = p - q + pi + ri - qi + ii$ increases by 2. \square

Proof of Theorem 7. The structural classification of nodes belonging to PI, RI, II, BI and QI cannot be altered by operator (i) due to $v_1, v_2 \in OI$. Nodes v_1 and v_2 cannot be transformed into pendants based on the operators (ii)–(iv); i.e., p and q are unaltered. Node v_2 cannot be transformed into II . If node v_2 is transformed into PI , the number of newly increased nodes in QI is no more than the number of newly increased nodes in PI ; i.e., $pi - qi$ does not decrease. Node v_1 must be transformed into II, PI or RI according to different $Card(Nr(v_1) \cap QI)$ values; i.e., $stavalue = p - q + pi + ri - qi + ii$ increases and $stavalue \geq stavalue + 1$. \square

Appendix B. ME1 optimization algorithm

Based on the ME1 stabilization [19], the eigenvalue power law [18] and the quasi-symmetry of the NLS about 1 [3,4], objective parameters for the optimization can be evaluated [9], e.g., the ME1 $m_{obj}(1)$ and the distribution of eigenvalues falling in intervals $\{(2(i-1)/k, 2i/k)\}_{i=1}^t$ (i.e., $\{w_{obj}^i\}_{i=1}^t$).

According to Theorems 2–7, we can design the optimization algorithm for synthetic Internet AS subgraphs generated by Inet-3.0. The algorithm is designed as follows:

Inputs: A synthetic Internet AS subgraph G , three objective values $m_{obj}(1)$, $\{w_{obj}^i\}_{i=1}^t$ and $W_{obj} = \sum_{l=1}^t (1 - \frac{2l-1}{k})^N \cdot w_{obj}^l$ (i.e., the WSD^t of the real AS subgraph) and a threshold T .

Step 1. Estimate $ME1 \approx n - 2 \cdot (1 - eps)^-$ and the WSD^t of graph G . Divide graph G into $P(G), Q(G)$ and $R(G)$, and divide subgraph $Inner(G)$ into $PI(PI_1 \cup PI_2), QI, RI, II, BI$ and $OI(OI_1 \cup OI_2)$. If $ME1 > m_{obj}(1) + 2$, go to **Step 2**. If $ME1 < m_{obj}(1) - 2$, go to **Step 3**. Otherwise, conclude the optimization.

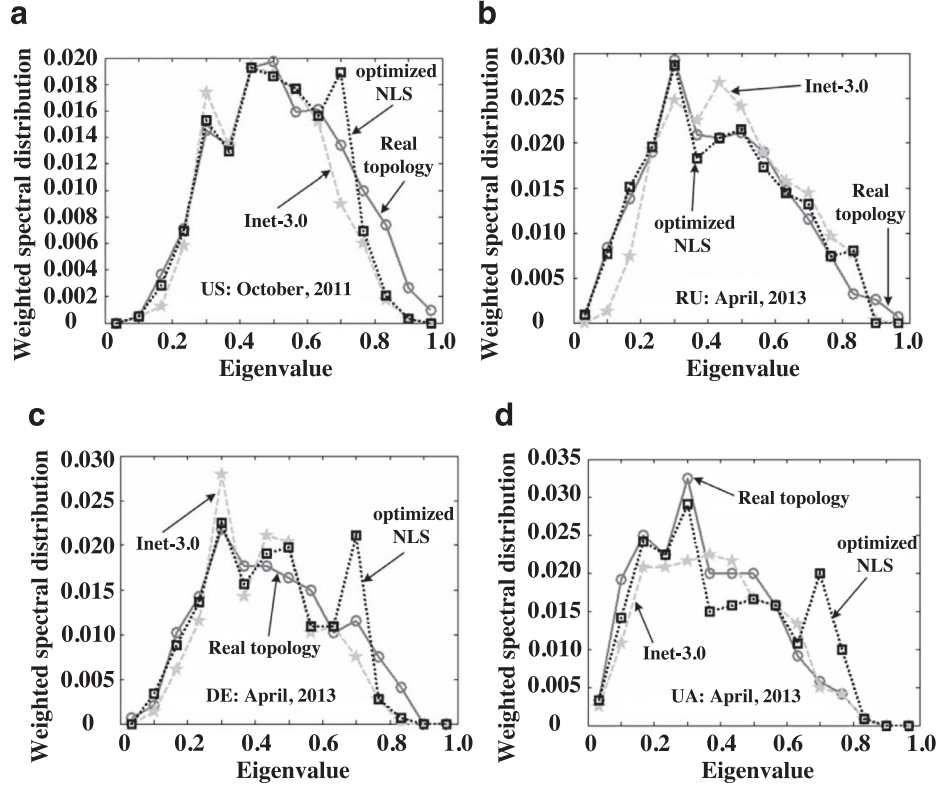


Fig. 7. Two-dimensional displays of the WSD between real topologies and corresponding graphs with optimized NLS. Based on the quasi-symmetry of the NLS about 1, only eigenvalues from 0 to 1 are shown in this figure. The x-axis represents the midpoints of equally spaced bins in [0, 1], and the y-axis represents the distribution of eigenvalues falling in the bins. With different weighting values for the eigenvalue distribution, we find that eigenvalues moving toward 0 are more important.

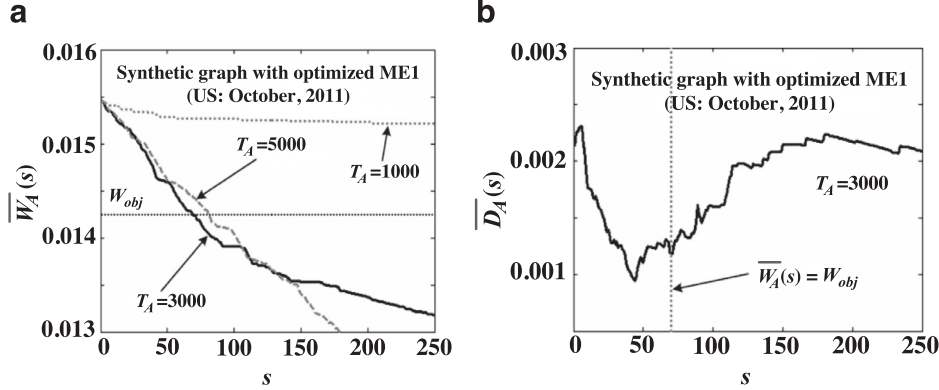


Fig. 8. $\bar{W}_A(s)$ and $\bar{D}_A(s)$ curves of a synthetic AS graph with optimized ME1. The Inet-3.0 inputs are evaluated according to the real US Internet AS subgraph in the snapshot of October 2011. W_{obj} is the WSD^t of the real AS subgraph. (a) $\bar{W}_A(s)$ curves with different T_A values. (b) $\bar{D}_A(s)$ curve with $T_A = 3000$.

Step 2. Let $AltSet = \{(v_1, v_2) | v_1 \in OI, v_2 \in Pl_2\} \cup \{(v_1, v_2) | v_1 \neq v_2, v_1 \in II \cup Pl_2 \cup Set, v_2 \in II\}$, where $Set = \{\bar{v} \in OI | Card(Nr(\bar{v}) \cap OI_1) = 0\}$. If $Card(AltSet) = 0$, conclude the optimization. If $Card(AltSet) > 0$, do the following:

Denote by $AltSet(T) = \{e_i\}_{i=1}^T$ the top T edges in $AltSet$ with decreasing order of $MC_1 = |d_u - d_v|$ if $WSD^t \geq W_{obj}$ and by $AltSet(T) = \{e_i\}_{i=1}^T$ the top T edges in $AltSet$ with increasing order of $MC_2 = \max(d_u, d_v)$ if $WSD^t < W_{obj}$.

Denote by G_i the graph induced by adding the edge e_i into graph G . Let

$$D^t(G_i, N) = \sum_{l=1}^t \left(1 - \frac{2l-1}{k}\right)^N \cdot |w_{G_i}^l - w_{obj}^l| \quad (B.1)$$

where $\{w_{G_i}^l\}_{l=1}^t$ denotes the distribution of eigenvalues falling in $\{(2(l-1)/k, 2l/k)\}_{l=1}^t$ ($k = 30$ and $t = 10$ are selected in this paper).

We calculate

$$i^* = \arg_{i=1, \dots, T} \min D^t(G_i, N) \quad (B.2)$$

and add edge e_{i^*} to graph G . Go to **Step 1**.

Step 3. Let $AltSet_1 = \{(v_1, v_2) | v_1 \in II, v_2 \in Set\}$, where $Set = \{\bar{v} \in OI | Card(Nr(\bar{v}) \cap OI_1) \geq 2\}$. If $Card(AltSet_1) > 0$, calculate i^* similar to Eq. (B.2) and add edge $e_{i^*} \in AltSet_1$ to graph G . If $Card(AltSet_1) = 0$, go to **Step 4**.

Step 4. Let $AltSet_2 = \{(v_1, v_2) | v_1, v_2 \in BI, v_2 \in Nr(v_1)\}$.

If $Card(AltSet_2) > 0$, do the following: denote by $AltSet_2(T) = \{e_i\}_{i=1}^T$ the top T edges in $AltSet_2$ with decreasing order of MC_1 if $WSD^t \leq W_{obj}$ and by $AltSet_2(T) = \{e_i\}_{i=1}^T$ the top T edges in $AltSet_2$ with increasing order of MC_2 if $WSD^t > W_{obj}$. Denote by G_i the graph induced by deleting edge e_i from graph G . We calculate

$$i^* = \arg_{i=1, \dots, T} \min D^t(G_i, N). \quad (B.3)$$

and delete edge e_{i^*} from graph G . Additionally, some edges (zero, one or two) from the operators (ii) and (iii) in [Theorem 6](#) are added to graph G (please note that a method similar to [Eq. \(B.2\)](#) is used for the addition of edges). If $Card(AltSet_2) = 0$, go to **Step 5**.

Step 5. Select a node $v_1 = \arg \min_{v \in OI} (Card(Nr(v) \setminus QI))$ and delete all edges (v_1, v_2) , where $v_2 \in Nr(v_1) \setminus QI$. Additionally, some edges from the operators (ii)–iv) in [Theorem 7](#) are added to graph G . If node v_1 cannot be found, conclude the optimization. Otherwise, go to **Step 1**. \square

The threshold T can be small for fast calculation of the algorithm (e.g., we select $T = 5$ in this paper) because the WSD optimization is another important task.

Appendix C. WSD optimization algorithm

Based on [Sections 4](#) and [5](#), the WSD optimization can keep the ME1 unaltered. Hence, this algorithm should be run after the ME1 optimization. To optimize the synthetic Internet AS graph G with optimized ME1, we should pay more attention to the $D^t(G, N)$ metric, which measures the difference between a synthetic graph G and the corresponding real AS subgraph.

Denote by $Dis^t(e_i)$ the $D^t(G, N)$ of a graph induced by adding or deleting an edge e_i for AS graph G and by $AddSet(T_A)$ the top T_A edges in $AddSet$ (see [Section 4](#)) with decreasing order of $MC_1 = |d_u - d_v|$. Let $AddSet(T_A) = \{e_i^a\}_{i=1}^{T_A}$, where $Dis^t(e_1^a) \leq Dis^t(e_2^a) \leq \dots \leq Dis^t(e_{T_A}^a)$. Simultaneously, denote by $DelSet(T_D)$ the top T_D edges in $DelSet$ (see [Section 4](#)) with decreasing order of MC_1 . Let $DelSet(T_D) = \{e_i^d\}_{i=1}^{T_D}$, where $Dis^t(e_1^d) \leq Dis^t(e_2^d) \leq \dots \leq Dis^t(e_{T_D}^d)$. Denote by $\bar{W}_A(s)$ and $\bar{D}_A(s)$, respectively, the WSD^t and $D^t(G, N)$ of a graph induced by adding a series of edges $\{e_i^a\}_{i=1}^s$ to AS graph G and by $\bar{W}_D(s)$ and $\bar{D}_D(s)$, respectively, the WSD^t and $D^t(G, N)$ of a graph induced by deleting a series of edges $\{e_i^d\}_{i=1}^s$ from AS graph G . If the deletion of edge e_s^d alters the nodes in sets $P(G)$ and QI , add the edge to AS graph G again and let $\bar{W}_D(s) = \bar{W}_D(s-1)$ and $\bar{D}_D(s) = \bar{D}_D(s-1)$.

According to the analyses of [Fig. 8](#), we can obtain that i) there are some relationships between $\bar{W}_A(s)$ and $\bar{D}_A(s)$ that are induced by different weighting values $(1-\theta)^N$ ($\theta \in (2(i-1)/k, 2i/k]$) of the eigenvalue distribution and ii) T_A (or T_D) can be smaller compared to the study of [Section 4](#). To optimize the $D^t(G, N)$ metric, the algorithm of [Section 5](#) should be improved as follows:

- (i) The thresholds T_A and T_D are set as constant and smaller values, e.g., $T_A = T_D = 3000$ for US (Oct 2011) and $T_D = 2000$ for RU (Apr 2013).
- (ii) The $W^t(e_i)$ metric is replaced by $Dis^t(e_i)$, and the $Dis^t(e_i)$ metric is used to sort the edges in $AddSet(T_A)$ and $DelSet(T_D)$.
- (iii) When s_{obj} in [Eq. \(8\)](#) or [Eq. \(9\)](#) is calculated, we calculate

$$s_{obj}^* = \arg \min_{s \in Set} \bar{D}_A(s) \quad (C.1)$$

or

$$s_{obj}^* = \arg \min_{s \in Set} \bar{D}_D(s) \quad (C.2)$$

where $Set = \{s | s = s_{obj} - i \cdot T^*; i = 1, 2, \dots; s \geq 1\}$ and T^* is 1, 2 or 3.

Then, add (or delete) the edges $\{e_i^a\}_{i=1}^{s_{obj}^*}$ (or $\{e_i^d\}_{i=1}^{s_{obj}^*}$) for AS graph G and conclude the algorithm. \square

References

- [1] M.E.J. Newman. Spectral community detection in sparse networks. 2013, <http://arxiv.org/pdf/1308.6494.pdf> (accessed Dec 2015).
- [2] D. Cvetkovic, S.K. Simic, Spectral graph theory in computer science, *IPSI BgD Trans. Adv. Res.* 8 (2) (2012) 35–42.
- [3] E.K. Cetinkaya, M.J.F. Alenazi, J.P. Rohrer, et al., Topology connectivity analysis of Internet infrastructure using graph spectra, in: *Proceedings of 4th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, IEEE, 2012, pp. 752–758.

- [4] E.K. Cetinkaya, M.J.F. Alenazi, A.M. Peck, et al., Multilevel resilience analysis of transportation and communication networks, *Telecommun. Syst.* (2015), doi:10.1007/s11235-015-9991-y.
- [5] G. Zhang, B. Qoitoin, S. Zhou, Phase changes in the evolution of the IPv4 and IPv6 AS-Level Internet topologies, *Comput. Commun.* 34 (5) (2011) 649–657.
- [6] S. Zhou, Understanding the evolution dynamics of Internet topology, *Phys. Rev. E* 74 (1) (2006) 016124.
- [7] R.G. Clegg, C. Di Cairano-Gilfedder, S. Zhou, A critical look at power law modelling of the Internet, *Comput. Commun.* 33 (3) (2010) 259–268.
- [8] S. Zhou, R.J. Mondragon, Accurately modeling the Internet topology, *Phys. Rev. E* 70 (6) (2004) 066108.
- [9] B. Jiao, Y. Zhou, J. Du, et al., Study on the stability of the topology interactive growth mechanism using graph spectra, *IET Commun.* (2014), doi:10.1049/iet-com.2014.0183.
- [10] C. Gkantsidis, M. Mihail, E. Zegura, Spectral analysis of Internet topologies, *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies 1* (2003) 364–374.
- [11] C. Francesc, G. Silvia, A star-based model for the eigenvalue power law of Internet graphs, *Phys. A: Stat. Mech. Appl.* 351 (2–4) (2005) 680–686.
- [12] J. Wu, M. Barahona, Y.J. Tan, et al., Robustness of random graphs based on graph spectra, *Chaos* 22 (2012) 043101 <http://dx.doi.org/10.1063/1.4754875>.
- [13] W. Liu, H. Sirisena, K. Pawlikowski, Weighted algebraic connectivity metric for non-uniform traffic in reliable network design, in: *Proceedings of the International Conference on Ultra Modern Telecommunications*, St. Petersburg, Russia, 2009, pp. 1–6.
- [14] D. Fay, H. Haddadi, A. Thomason, et al., Weighted spectral distribution for Internet topology analysis: theory and applications, *IEEE/ACM Trans. Netw.* 18 (1) (2010) 164–176.
- [15] D. Fay, H. Haddadi, S. Uhlig, et al., Discriminating graphs through spectral projections, *Comput. Netw.* 55 (15) (2011) 3458–3468.
- [16] B. Jiao, Y.P. Nie, J.M. Shi, et al., Accurately and quickly calculating the weighted spectral distribution, *Telecommun. Syst.* (2015), doi:10.1007/s11235-015-0077-7.
- [17] S.K. Butler, *Eigenvalues and Structures of Graphs*, University of California, San Diego, 2008 Doctor Thesis.
- [18] L. Trajkovic, Analysis of Internet topologies, *IEEE Circuits Syst. Mag.* 10 (2010) 48–54.
- [19] Vukadinovic D., Huang P., Erlebach T. A Spectral Analysis of the Internet Topology. ETH TIK-NR, 2001, pp. 1–11.
- [20] D. Vukadinovic, P. Huang, T. Erlebach, On the spectrum and structure of Internet topology graphs, in: *Innovative Internet Computing Systems*, Springer Berlin Heidelberg, 2002, pp. 83–95.
- [21] Jin C., Chen Q., Jamin S. Inet: Internet Topology Generator. Technical report CSE-TR-433-00, University of Michigan EECS Dept., 2000.
- [22] H. Haddadi, D. Fay, S. Uhlig, et al., Tuning topology generators using spectral distributions, in: *Proceedings of SPEC International Performance Evaluation Workshop, Germany, Lecture Notes in Computer Science*, 2008, pp. 154–173.
- [23] B.M. Waxman, Routing of multipoint connections, *IEEE J. Sel. Areas Commun.* 6 (9) (1988) 1617–1622.
- [24] A.L. Barabasi, R. Albert, Emergence of scaling in random networks, *Science* 286 (5439) (1999) 509–512.
- [25] T. Bu, D. Towsley, On distinguishing between Internet power law topology generators, in: *Proceedings of IEEE Infocom 2002*, New York, NY, 2002.
- [26] Winick J., Jamin S. Inet-3.0: Internet Topology Generator. University of Michigan, Technical Report CSE-TR-456-02, 2002.
- [27] H. Haddadi, D. Fay, A. Jamakovic, et al., On the importance of local connectivity for Internet topology models, in: *Proceedings of 21st International Teletraffic Congress, ITC 21*, Paris, France, September, 2009, pp. 1–8.
- [28] S.Y. Shafi, M. Arcak, L.E. Ghaoui, Graph weight allocation to meet Laplacian spectral constraints, *IEEE Trans. Autom. Control* 57 (7) (2012) 1872–1877.
- [29] X. Ying, X. Wu, Randomizing social networks: a spectrum preserving approach, in: *Proceedings of SIAM International Conference on Data Mining*, Atlanta, Georgia, USA, 2008, pp. 739–750.
- [30] A. Sydney, C. Scoglio, D. Gruenbacher, Optimizing algebraic connectivity by edge rewiring, *Appl. Math. Comput.* 219 (2013) 5465–5479.
- [31] Archipelago Measurement Infrastructure, <http://www.caida.org/projects/ark/> (accessed Dec 2015).
- [32] Iffinder Alias Resolution Tool, <http://www.caida.org/tools/measurement/> (accessed Dec 2015).
- [33] S. Zhou, G.Q. Zhang, Chinese Internet AS-level topology, *IET Commun.* 1 (2) (2007) 209–214.
- [34] A. Lakhina, J.W. Byers, M. Crovella, et al., On the geographic location of Internet resources, *IEEE J. Sel. Areas Commun.* 21 (6) (2003) 934–948.
- [35] H. Haddadi, M. Rio, A. Moore, Network topologies: inference, modeling, and generation, *IEEE Commun. Surv. Tutor.* 2nd Quart. 10 (2) (2008) 48–69.
- [36] M. Faloutsos, P. Faloutsos, C. Faloutsos, On power-law relationships of the Internet topology, *ACM SIGCOMM Comput. Commun. Rev.* 29 (4) (1999) 251–262.
- [37] Diot C. Sprint tier 1 IP Backbone: Architecture, Traffic Characteristics, and Some Measurement Results, 2001.
- [38] E. Gregori, A. Improta, L. Lenzini, et al., Discovering the geographic properties of the Internet AS-level topology, *Netw. Sci.* (2013) 1–9.
- [39] Y. Tian, R. Dey, Y. Liu, et al., Topology Mapping and Geolocating for China's Internet, *IEEE Trans. Parallel Distrib. Syst.* 24 (9) (2013) 1908–1917.
- [40] Meenakshi S.P., Raghavan S.V. Forecasting and Event Detection in Internet Resource Dynamics using Time Series Models. 2013, <http://arxiv.org/pdf/1306.6413.pdf>, (accessed Dec 2015).