Beyond technical solutions: Understanding the role of governance structures in Internet routing security

Brenden Kuerbis, School of Public Policy, Georgia Institute of Technology
Milton Mueller, School of Public Policy, Georgia Institute of Technology
Rio Maulana, School of Information Studies, Syracuse University

Internet routing security anomalies often make headlines when they occur (Cowie, 2010, RIPE 2008). But not all networks on the Internet (known technically as Autonomous Systems) experience anomalies, and some experience more than others. While efforts to standardize and deploy secure routing technologies continue and commercial services to help mitigate them are now commonplace, a broader empirically-based contextual understanding of these anomalies does not exist.

Informed by theories of institutional economics and networked governance (Jones et al, 1997; Mueller et al, 2013), and using existing data from large-scale monitoring projects operated by computer scientists, this research begins to shed light on why network operators experience different levels of routing security anomalies. Our research method for the entire project uses quantitative measures of routing anomalies over time (the dependent variable) and a set of independent variables that reflect variations in the macro, meso and micro level governance structures among Autonomous Systems. This paper, representing an early stage of the research, focuses mainly on the independent variables at the meso level; i.e., “structural embeddedness” (SE). SE is defined as the extent to which Autonomous Systems are embedded within interconnections of other Autonomous Systems.

Previous work in physics suggests that structural embeddedness (SE) among Autonomous Systems should make them more resilient against attacks (Pastor-Satorras, Vázquez & Vespignani, 2001). This paper explores this argument as it relates to routing security, quantifying the SE of Autonomous Systems within the routing topology and then testing to see how SE is related to routing security anomalies. The paper also quantifies the route advertisement activity of each Autonomous System for which it has data, and correlates that to the number of anomalies. In a later phase of the research, we will use these findings to select specific Autonomous Systems for in-depth qualitative interviews about the decisions and actions they employ to secure their routing operations in order to correlate anomalies with micro-level and macro-level governance structures.

Data and Method

Our research method requires quantitative measures of routing anomalies over time (the dependent variable) and a set of independent variables that reflect variations in the governance structures among Autonomous Systems. This section describes how we defined these variables and gathered data for them.

Dependent variable: a longitudinal view of routing anomalies

Numerous systems and heuristics have been proposed by computer scientists and commercial companies to monitor Internet routing and identify anomalies. At a high level, monitoring systems work by identifying differences between observed interdomain routing announcements sent using the Border Gateway Protocol (BGP) and intended or expected route announcements. A routing announcement includes a prefix (i.e., an IP address block) and the AS number through which that prefix can be reached. Routers compile announcements into tables they reference to send traffic to the correct destination. Route announcement data are collected by numerous systems (e.g., BGPmon, RouteViews, RIPE-RIS, etc.) that maintain monitoring points on the Internet. Anomalies can be identified by observing routing announcement data inconsistent with 1) the BGP, or 2) routing policy and resource use data contained within various repositories (such as those maintained by Internet Routing Repositories (IRRs) and the Regional Internet Registries (RIRs)), or 3) expected business relationships between operators[1]. Monitoring systems continue to evolve to address known shortcomings, such as their limited view(s) of the Internet (Zhang et al. 2007), over- or under-estimation, and identification of new types of routing anomalies.

For example, the Cyclops system operated by UCLA's Internet Research Lab generates anomalies data on a daily basis, including transient announcement of prefixes, unexpected removal of prefix announcements, and announcement of bogus or incorrectly configured prefixes. More recently, researchers at Tsinghua University and Tsinghua National Laboratory for Information Science and Technology have developed the Argus system, which collects information about observed resource origin, AS adjacency and routing policy anomalies. Argus addresses weaknesses observed in previous systems, including distinguishing between types of prefix hijacks and eliminating false negatives. Other systems are currently being developed to improve the monitoring and identify specific cases of routing anomalies like traffic interception (Gill and Dainotti, 2014). Given the absence of unified criteria for quantifying the occurrence of anomalies, and of any comprehensive study of the prevalence and impact of routing anomalies over time, we intend to collect and archive routing anomaly data from the Argus system with the AS as the unit of analysis. Our project will be designed to allow us to monitor ongoing research in this area and adapt or update our measures of anomalies accordingly.

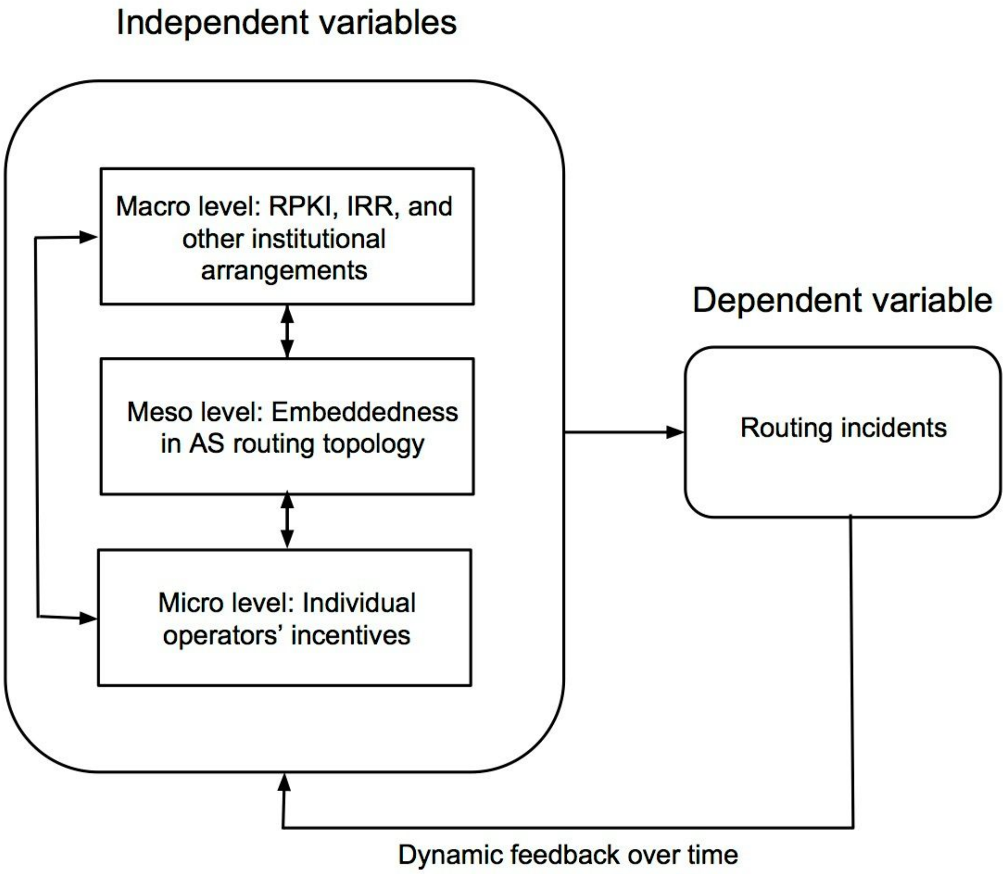
Independent variables: macro, meso and micro

As the diagram of our analytical model (Figure 1, below) shows, data will be collected and analyzed at the macro, meso and micro levels. The first step is to detect variations in the organizational structures among Autonomous Systems at the meso level. We expect to find that distinct configurations at the meso level will vary in their susceptibility to routing anomalies. Using these observations as a guide for further inquiry, we will identify institutional

[1] An example of this is the "valley-free" assumption, where one does not expect to see BGP announcements that would result in traffic routing from a customer to provider to another customer of that provider before routing to the broader internet.

arrangements at the macro level and use interviews to explore the different factors shaping the incentives and decisions of operators at the micro level.

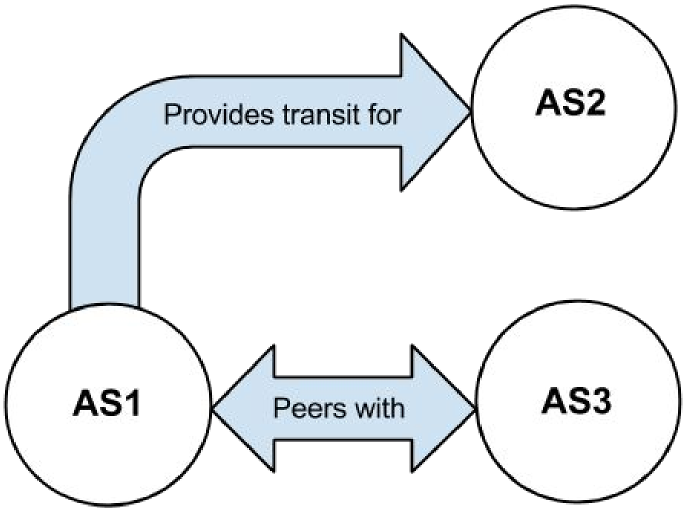
Figure 1: Analytical model


Independent variables
Macro level: RPKI, IRR, and other institutional arrangements
Meso level: Embeddedness in AS routing topology
Micro level: Individual operators' incentives
Dependent variable
Routing incidents
Dynamic feedback over time

Meso Level

Meso-level research will involve analysis of the Internet's routing topology using graph data generated by computer science researchers. Data generated by the Center for Applied Internet Data Analysis (CAIDA) at the University of San Diego provides longitudinal snapshots of AS routing relationships, including whether AS pairs have peer, provider or customer relationships. Figure 2 (below) outlines how we model AS relationships as reported by CAIDA. For peering, the organizational relationship between Autonomous Systems is bidirectional. In Figure 2, AS1 and AS3 would have one indegree and and one outdegree each as a result. For transit, the relationship direction is from the transit provider to the customer. In Figure 2, AS1 provides transit to AS2, and therefore has one outdegree, while AS2 has one indegree. Our modeling is based on organizational relationships, not actual packet flows which would be bidirectional in all cases.

Figure 2: AS relationship model



	AS1	AS2	AS3
indegree	1	1	1
outdegree	2	0	1

Using this representation of the Internet's topology we measure the structural embeddedness of an AS. To derive SE measures, we utilize k-Nearest-Neighbor measurements derived previously by Pastor-Satorras, Vázquez & Vespignani (2001) and implemented in the Sci2 tool.[2] First, we convert the CAIDA raw relationship data to a directed edge network file as explained above in Figure 2. For each AS we then calculate indegree and outdegree. We then use those measurements to generate four k-Nearest-Neighbor scores for each AS, which are a measure of the correlation between the degree of a AS and that of its neighboring Autonomous Systems. Scores include:

1. *kInIn*, which compares the indegree of an AS with the average indegree of their incoming neighbors (i.e., the neighboring Autonomous Systems adjacent to incoming edges);
2. *kInOut*, which compares the indegree of an AS with the average outdegree of their incoming neighbors (i.e., the neighboring Autonomous Systems adjacent to incoming edges);
3. *kOutIn*, which compares the outdegree of an AS with the average indegree of their outgoing neighbors (i.e., the neighboring Autonomous Systems adjacent to outgoing edges);
4. *kOutOut*, which compares the outdegree of an AS with the average outdegree of their outgoing neighbors (i.e., the neighboring Autonomous Systems adjacent to outgoing edges).

[2] Sci2 Team. (2009). Science of Science (Sci2) Tool. Indiana University and SciTech Strategies, https://sci2.cns.iu.edu.

The role and importance of structural embeddedness within social, economic and governance networks is well known. In other policy domains, theory suggests that SE is a precondition for effective network governance, with higher levels of SE helping to ensure the spread of information and permitting actors to scrutinize one another effectively, thereby invoking trust and minimizing uncertainty. Our method will allow us to explore how embeddedness of specific Autonomous Systems within the graph may evolve over time, highlighting different types of networked governance arrangements and different organizational policies and incentives.

Macro Level

This aspect of the research will track the evolution of RPKI and IRR implementation, the diffusion of RPKI use by network operators, and IRR usage. RPKI is a security technology that implements a hierarchy of digital certificates that can be used to authenticate the use of address blocks by ASes. . Internet Routing Registries (IRR) is an older technology that allows network operators to register and validate routing policies of ASes. While RPKI and IRRs are often viewed solely technical enhancements to the Internet's infrastructure, they both have governance implications. The deployment of RPKI requires ISPs to trade-off greater security for less routing autonomy. (Kuerbis & Mueller, 2013) If RPKI trusted authorities (e.g., RIRs or large ISPs) are faulty, misconfigured, compromised, or compelled to misbehave, they can dramatically impact routing. (Cooper et al. 2013) Similar governance issues arise with IRRs, with one RIR recently consulting with network operators to assess demand and criteria in establishing an IRR to be used in route validation.[3] The study will eventually (1) monitor the deployment of RPKI by the Regional Internet Registries and (2) understand the usage of IRRs by ISPs, including analyzing the organizational, contractual or policy elements of both systems and the differences between implementations. This research will allow us to find out whether macro-level institutional structures such as RPKI or IRR use is associated with a reduction in anomalies among the operators who use them. Furthermore, it will also allow us to determine whether particular institutional arrangements governing RPKI or IRRs facilitate or impede their widespread adoption, and adoption of related technologies like BGPSEC.

Micro Level

Micro-level research will involve both qualitative interviews with operators and the collection of qualitative data about their incentives to adopt and use routing security technologies (e.g., IRRs and RPKI) and practices (e.g., ingress and egress route filtering) to protect routing security. These interviews will also elicit information related to how operator's contracts with peering and transit partners might impact routing security. They will also provide a reality check on explaining some of the patterns observed in meso-level analysis.

Research Questions

As noted before, this paper concentrates on meso-level analyses; i.e. we correlate levels of structural embeddedness in the Internet's routing topology with the number of anomalies. We also explore the relationship between anomalies and the number of prefix advertisements made

[3] https://www.arin.net/participate/acsp/community_consult/03-17-2015_irr.html

by Autonomous Systems. All three types of data (SE, number of prefixes advertised and anomalies) are available longitudinally, allowing us to assess trends. In the current paper, we test two propositions in quantitative terms:

Proposition 1. Higher levels of structural embeddedness among Autonomous Systems are negatively correlated with the number of routing anomalies.

We may also be able to answer another question:

Proposition 2: Are any other independent variables related to an Autonomous System's routing practices significantly correlated with variation in the number of routing anomalies?

Results

Combining the Routeviews, CAIDA and Argus data, we observed activity for 51,436 Autonomous Systems over 41 months between June 2011 and October 2014. We aggregated these observations on a monthly basis producing 1,431,649 records. Observations of Autonomous Systems ranged from a minimum of 1 month (.86%) to a maximum of 35 months (60.51%), reflecting intermittent use of some AS numbers by network operators. Each record shows the number of routing anomalies the AS experienced in that time frame (which could be 0), the average number of prefixes it advertised during that period, and measures of its structural embeddedness during that month.

From those 1.43 million cases, we make the following initial observations:

- The absolute number of routing security anomalies is small when viewed in comparison to overall Internet routing activity. Even after supplementing publicly available data and using the most liberal interpretation of that data, 3.09% (1587) of Autonomous Systems observed have experienced an incident. Nonetheless, 2,455 anomalies of varying severity did occur over that 41 month period, and some Autonomous Systems experienced multiple anomalies in a month.

- There are still clear shortcomings in existing routing incident monitoring and reporting system(s), and a need for improved systems that capture, categorize and publicly report routing security anomalies. Seemingly major anomalies that were reported in the technical press over the course of the study did not always show up in the Argus data.

- The data indicate a positive (.108**) statistically significant correlation between the number of routing anomalies experienced by an AS and the number of prefixes it advertises (Table 2). If the test group is limited to AS's that have experienced one or more anomalies, the positive correlation strengthens to .286** (Table 4).

- The data indicate a relatively strong (0.269**) positive correlation between the number of routing anomalies experienced by an AS and the number of peering and transit relationships (out degrees) it maintains (Table 2). If the test group is limited to Autonomous Systems that have experienced one or more anomalies, the positive correlation strengthens to 0.469** (Table 4).

- Contrary to the expectations established by Pastor-Satorras, R., Vázquez, A., & Vespignani, A. (2001) SE has at best a very weak role in explaining the prevalence or absence of anomalies. If the number of anomalies is correlated with k measures across all Autonomous Systems, SE accounts for -0.001 to 0.012** of variation (See Table 2, row 1). Put bluntly, the effect of SE across all ASs is negligible. When the correlation test is limited to Autonomous Systems that have experienced one or more anomalies, however, higher levels of SE are indeed associated with lower rates of anomalies, but the correlation is quite weak. The range of negative correlation goes from -.041* (for kInOut) to -.054* (for kOutIn), and the statistical significance (.05) is weaker (Table 4).

With respect to Proposition 1, these results do not support our original hypothesis. The answer to Proposition 2, however, is clearly Yes. The strongest factors influencing susceptibility to routing anomalies appear to be the number of peering and transit relationships an AS maintains with other Autonomous Systems, followed by the number of prefix advertisements it announces.

In some ways, the findings appear to be paradoxical: an AS's greater exposure to and engagement with other Autonomous Systems (through more peering and transit relationships and more prefix announcements) increases their exposure to routing anomalies, whereas broader measures of Structural Embeddedness (even after controlling for number of prefix announcements) indicate that SE has no effect upon, or slightly decreases, their susceptibility to routing anomalies. These results might make more sense, however, if the correlations are broken down by different types of anomalies. For instance, origin anomalies (when a prefix is advertised by an unauthorized or unexpected AS) can be announced from anywhere in the Internet's topology and have nothing to do, per se, with how embedded an AS is. More granular analysis based on incident type may reveal that SE does actually influence AS exposure to specific incident types, such as path anomalies or adjacency anomalies.

These findings do reinforce our study's focus upon governance structures. The initial findings essentially rule out pure topological properties (i.e., SE measures) as a major factor explaining routing security anomalies, making it more likely that specific operational and organizational practices (e.g., filtering, number of transit relationships) and institutional factors (e.g., legal and contractual relationships) will prove to play a significant role. Our continuing research hopes to identify and test these factors.

Data Tables

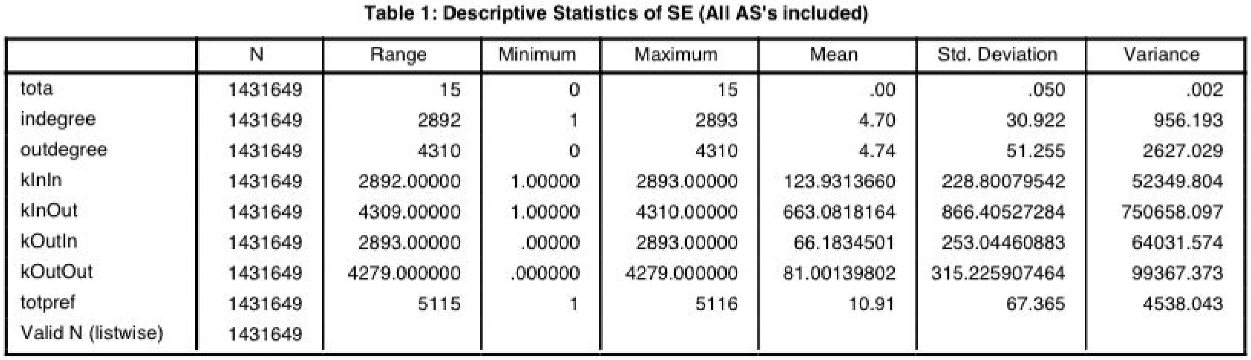


Table 1: Descriptive Statistics of SE (All AS's included)

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
tota	1431649	15	0	15	.00	.050	.002
indegree	1431649	2892	1	2893	4.70	30.922	956.193
outdegree	1431649	4310	0	4310	4.74	51.255	2627.029
kInIn	1431649	2892.00000	1.00000	2893.00000	123.9313660	228.80079542	52349.804
kInOut	1431649	4309.00000	1.00000	4310.00000	663.0818164	866.40527284	750658.097
kOutIn	1431649	2893.00000	.00000	2893.00000	66.1834501	253.04460883	64031.574
kOutOut	1431649	4279.000000	.000000	4279.000000	81.00139802	315.225907464	99367.373
totpref	1431649	5115	1	5116	10.91	67.365	4538.043
Valid N (listwise)	1431649						

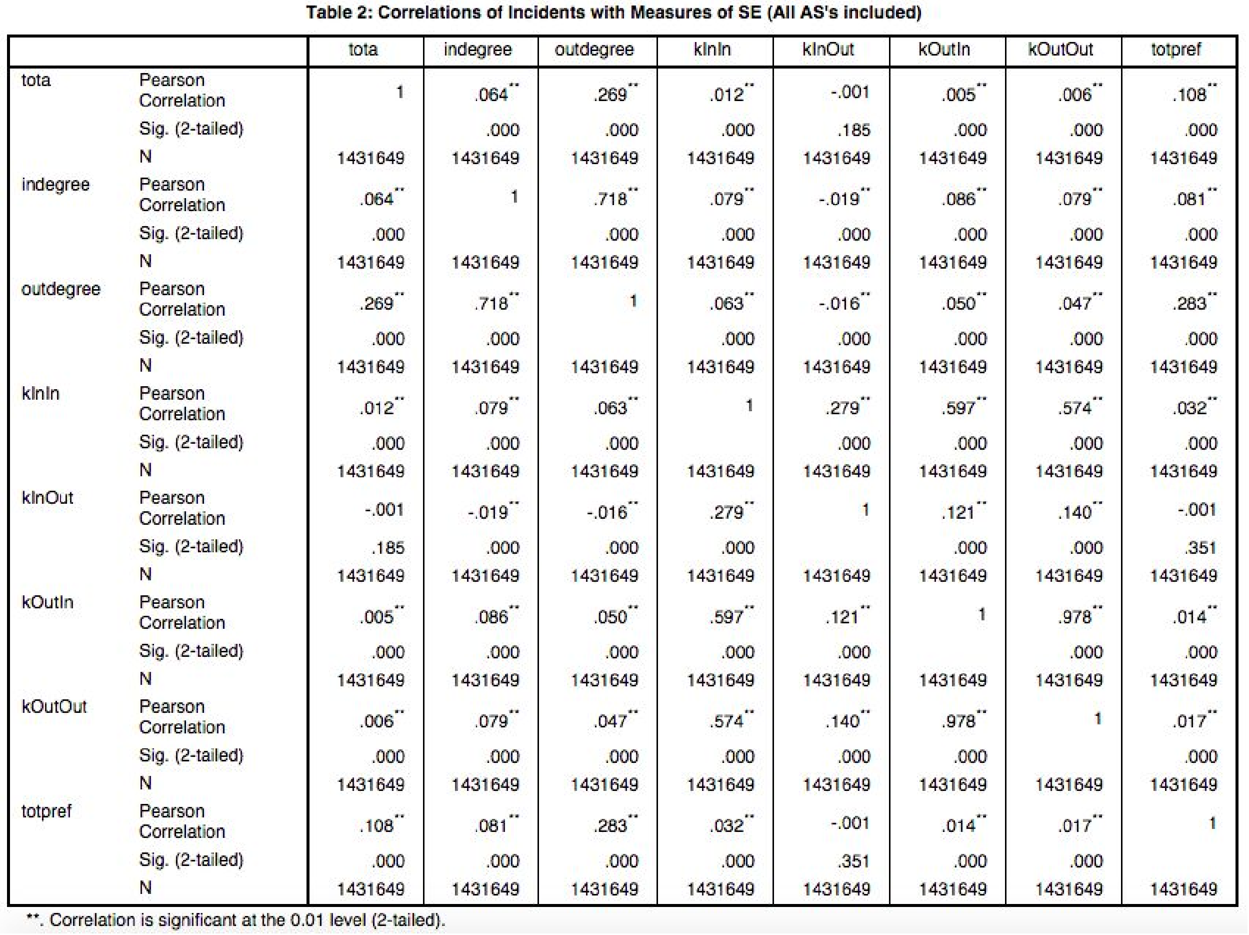
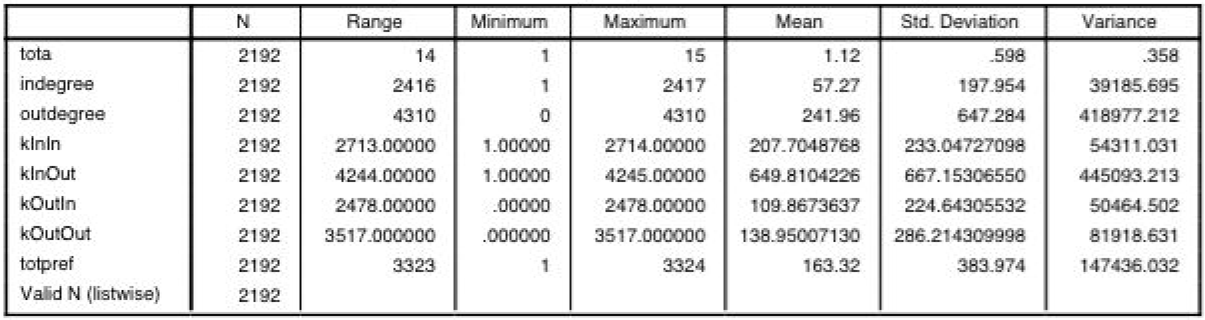


Table 2: Correlations of Incidents with Measures of SE (All AS's included)

		tota	indegree	outdegree	kInIn	kInOut	kOutIn	kOutOut	totpref
tota	Pearson Correlation	1	.064**	.269**	.012**	-.001	.005**	.006**	.108**
	Sig. (2-tailed)		.000	.000	.000	.185	.000	.000	.000
	N	1431649	1431649	1431649	1431649	1431649	1431649	1431649	1431649
indegree	Pearson Correlation	.064**	1	.718**	.079**	-.019**	.086**	.079**	.081**
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000	.000
	N	1431649	1431649	1431649	1431649	1431649	1431649	1431649	1431649
outdegree	Pearson Correlation	.269**	.718**	1	.063**	-.016**	.050**	.047**	.283**
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000	.000
	N	1431649	1431649	1431649	1431649	1431649	1431649	1431649	1431649
kInIn	Pearson Correlation	.012**	.079**	.063**	1	.279**	.597**	.574**	.032**
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000	.000
	N	1431649	1431649	1431649	1431649	1431649	1431649	1431649	1431649
kInOut	Pearson Correlation	-.001	-.019**	-.016**	.279**	1	.121**	.140**	-.001
	Sig. (2-tailed)	.185	.000	.000	.000		.000	.000	.351
	N	1431649	1431649	1431649	1431649	1431649	1431649	1431649	1431649
kOutIn	Pearson Correlation	.005**	.086**	.050**	.597**	.121**	1	.978**	.014**
	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000	.000
	N	1431649	1431649	1431649	1431649	1431649	1431649	1431649	1431649
kOutOut	Pearson Correlation	.006**	.079**	.047**	.574**	.140**	.978**	1	.017**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000		.000
	N	1431649	1431649	1431649	1431649	1431649	1431649	1431649	1431649
totpref	Pearson Correlation	.108**	.081**	.283**	.032**	-.001	.014**	.017**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.351	.000	.000	
	N	1431649	1431649	1431649	1431649	1431649	1431649	1431649	1431649

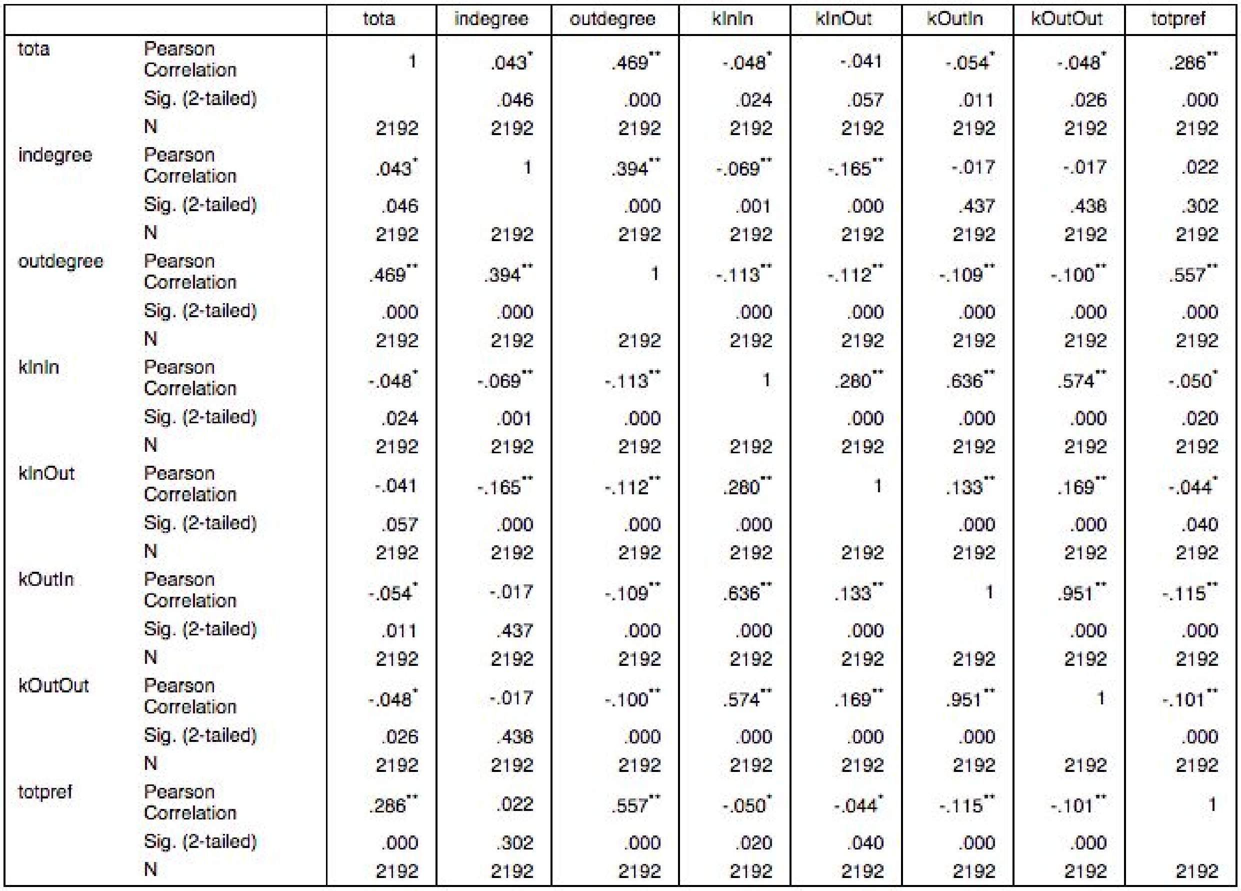
**. Correlation is significant at the 0.01 level (2-tailed).

Table 3: Descriptive Statistics of SE (Only AS's with Incidents >0 included)



	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
tota	2192	14	1	15	1.12	.598	.358
indegree	2192	2416	1	2417	57.27	197.954	39185.695
outdegree	2192	4310	0	4310	241.96	647.284	418977.212
kInIn	2192	2713.00000	1.00000	2714.00000	207.7048768	233.04727098	54311.031
kInOut	2192	4244.00000	1.00000	4245.00000	649.8104226	667.15306550	445093.213
kOutIn	2192	2478.00000	.00000	2478.00000	109.8673637	224.64305532	50464.502
kOutOut	2192	3517.000000	.000000	3517.000000	138.95007130	286.214309998	81918.631
totpref	2192	3323	1	3324	163.32	383.974	147436.032
Valid N (listwise)	2192						

Table 4: Correlations of Incidents with SE (Only AS's with Incidents >0 included)



		tota	indegree	outdegree	kInIn	kInOut	kOutIn	kOutOut	totpref
tota	Pearson Correlation	1	.043*	.469**	-.048*	-.041	-.054*	-.048*	.286**
	Sig. (2-tailed)		.046	.000	.024	.057	.011	.026	.000
	N	2192	2192	2192	2192	2192	2192	2192	2192
indegree	Pearson Correlation	.043*	1	.394**	-.069**	-.165**	-.017	-.017	.022
	Sig. (2-tailed)	.046		.000	.001	.000	.437	.438	.302
	N	2192	2192	2192	2192	2192	2192	2192	2192
outdegree	Pearson Correlation	.469**	.394**	1	-.113**	-.112**	-.109**	-.100**	.557**
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000	.000
	N	2192	2192	2192	2192	2192	2192	2192	2192
kInIn	Pearson Correlation	-.048*	-.069**	-.113**	1	.280**	.636**	.574**	-.050*
	Sig. (2-tailed)	.024	.001	.000		.000	.000	.000	.020
	N	2192	2192	2192	2192	2192	2192	2192	2192
kInOut	Pearson Correlation	-.041	-.165**	-.112**	.280**	1	.133**	.169**	-.044*
	Sig. (2-tailed)	.057	.000	.000	.000		.000	.000	.040
	N	2192	2192	2192	2192	2192	2192	2192	2192
kOutIn	Pearson Correlation	-.054*	-.017	-.109**	.636**	.133**	1	.951**	-.115**
	Sig. (2-tailed)	.011	.437	.000	.000	.000		.000	.000
	N	2192	2192	2192	2192	2192	2192	2192	2192
kOutOut	Pearson Correlation	-.048*	-.017	-.100**	.574**	.169**	.951**	1	-.101**
	Sig. (2-tailed)	.026	.438	.000	.000	.000	.000		.000
	N	2192	2192	2192	2192	2192	2192	2192	2192
totpref	Pearson Correlation	.286**	.022	.557**	-.050*	-.044*	-.115**	-.101**	1
	Sig. (2-tailed)	.000	.302	.000	.020	.040	.000	.000	
	N	2192	2192	2192	2192	2192	2192	2192	2192

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

References

Cowie, J. (2010). China's 18-minute mystery. *Renesys blog.* http://www.renesys.com/blog/2010/11/chinas-18-minute-mystery.shtml.

Jones, C., W.S. Hesterly, and S.P. Borgatti. (1997). A general theory of network governance: Exchange conditions and social mechanisms. The Academy of Management Review, 22(4):911–945.

Mueller, M., A. Schmidt and B. Kuerbis, (2013). Internet Security and Networked Governance in International Relations. *International Studies Review* 15(1), 86-104, March.

Pastor-Satorras, R., Vázquez, A., & Vespignani, A. (2001). Dynamical and Correlation Properties of the Internet. *Physical Review Letters*, *87*(25), 258701. doi:10.1103/PhysRevLett.87.258701

RIPE. (2008) Youtube hijacking: A RIPE-NCC ris case study. http://www.ripe.net/internet-coordination/news/industry-developments/youtube-hijacking-a-ripe-ncc-ris-case-study,