LA-UR- 01-5784

Approved for public release; distribution is unlimited.

Title: Empirical Characterization of Infrastructure Networks

Author(s): Brian W. Bush / D-4
Christina R. Files / NIS-7
David R. Thompson / D-5

Submitted to: worldwide web

Los Alamos
NATIONAL LABORATORY

Empirical Characterization of Infrastructure Networks

B. W. Bush, C. R. Files, and D. R. Thompson
Decision Applications Division
Los Alamos National Laboratory
22 October 2001

Abstract

Critical infrastructure protection is a recognized problem of national importance. Infrastructure networks such as electric power, natural gas, communications, and transportation systems have an inherent graph-theoretic structure. Quantitatively characterizing the essential properties of infrastructure networks for various domains lays a valuable foundation for studying the universal features (especially criticality, robustness, etc.) and specific characteristics of such networks. We construct an extensive reference data set of infrastructure network graphs: 44 graphs of 13 types with nearly one million vertices and over one million edges. After regularizing these graphs, we compute more than fifty metrics related to connectivity, distance scale, cyclicity, cliquishness, and redundancy. We contrast these metrics for different types of infrastructures, study their interrelationship, and use them to cluster and classify systems. We consider both intact networks and networks that have been degraded by the removal of some vertices or edges either at random or systematically—this provides insight as to the robustness of the network if it were subject to a natural disaster or an attack.

Table of Contents

I. Introduction

A. Motivation

Presidential Decision Directive No. 63 has recognized critical infrastructure protection as a problem of national importance. Moreover, the 1999-2004 LANL Strategic Plan identifies the need to develop techniques that enable the nation to accurately assess the vulnerability of critical infrastructures. Our work on generic graph-based analysis of interconnected infrastructures supports this mission by providing insight into the general behavior of these networks and by furnishing methods for their analysis. Issues of criticality and causality comprise questions of special importance in these infrastructures. Criticality analysis involves the identification of components and events for which the infrastructure dynamics have particular sensitivity. Causality analysis identifies potential interactions between events in time on the network whereby one event contributes to the occurrence of another.

Graph- and network-based analysis techniques have proven valuable methods for capturing important aspects of criticality and causality. For example, one can use graph analyses to find the minimum number of cuts in an electric power system that will result in a blackout for a particular region of interest, identify key single or multiple points of failure for a natural gas system, or find the fastest path between points in a transportation system. These analyses have the significant advantage of avoiding the time-consuming, resource-intensive, and expensive process of developing detailed simulation models. We consider numerous additional aspects of critical infrastructures from a graph-theoretic analysis perspective to better characterize issues of criticality and causality.

Quantitatively characterizing the essential properties of infrastructure networks for various domains will lay a valuable foundation for studying the universal features (especially criticality, robustness, etc.) of such networks. For instance, electric power networks tend to have a hierarchically linked ring structure with some radial edges, whereas natural gas networks possess a more tree-like layout; additionally, little is known about the typical pattern of inter-domain infrastructure dependencies.

B. Approach

Critical infrastructure networks such as electric power, natural gas, communications, and transportation systems have an inherent graph-theoretic structure. In order to understand the universal and specific features of such networks, we first construct an extensive database of infrastructure networks. In fact, we have collected and formatted as graphs the descriptions of dozens of diverse networks, many of them with national coverage. This provides a uniform starting point for the rest of the analysis and a reference data set for future work. We next regularize the graphs by cleaning them to remove small components and artifacts of their representation; this makes them more readily comparable to one another.

Preliminary to analyzing the graphs, we construct a set of metrics for measuring properties that reflect critical features of the graphs. These properties characterize the distance scales, connectivity, clustering, etc. of the graphs, their vertices, and their edges. We have implemented and thoroughly tested algorithms that efficiently calculate the metrics.

Our analysis of the metrics for intact infrastructure networks focuses on understanding the differences between the various infrastructures and the relationships between the metrics. We look for evidence of scaling behavior in the distributions of measurements. Clustering and classification techniques are applied to quantitatively differentiate networks. We also briefly discuss generative models of infrastructure networks and methods for identifying corrupt or missing elements in the source data used to construct their graphs.

In addition to studying intact networks, we also investigate the consequences of degrading networks by removing vertices or edges from their graph. We consider removing elements randomly and also removing elements most likely to be critical. This provides insight as to the robustness of the network if it were subject to a natural disaster or an attack.

C. Notation

Throughout this report we generally use the notation of G. Chartrand and L. Lesniak [CL 96] for graph-theoretic quantities. To supplement this, we have used the notation of

Watts [Wa 99] for the "small world" measures in graphs. Table 1 summarizes our notation.

Table 1. Notation used throughout this report in equations and data files.

Field Name	***Symbol***	***Description***	***Data File***									
			Summary	Graph	Component	Vertex	Edge	Pair	Degree	Distance	Cycle	Cut
`db`		graph data set used: 0 = "raw," 1 = "clean0", 2 = "clean1"	√									
`run`		unique compute run name	√									
`style`		style of vertex or edge deletion: 0 = no deletions, 1 = vertices randomly deleted, 2 = vertices deleted in order of highest degree first, 3 = articulation points deleted first then other vertices (all in order of highest degree), -1 = edges randomly deleted, -2 = edges deleted in order of highest total degree of their end vertices first, -3 = bridges deleted first then other edges (all in order of highest total degree)	√									
`f`	f	fraction of vertices or edges deleted	√									
`graph`		numerical id of the graph	√	√	√	√	√	√	√	√	√	√
`type`		numerical type of the graph	√	√	√							
`subtype`		numerical subtype of the graph	√	√	√							
`name`		name of the graph	√	√	√	√	√	√	√	√	√	√
`j`		numerical id of the component			√	√	√					
`id`		numerical id of the vertex				√						
`id1`		numerical id of the first vertex					√	√				
`id2`		numerical id of the second vertex					√	√				
`V`	V	order	√	√	√							
`E`	E	size	√	√	√							
`n`	n	number of components	√	√								
`delta`	δ	minimum degree of a vertex	√	√	√							
`Ddelta`	Δ	maximum degree of a vertex	√	√	√							
`r`	r	radius	√	√	√							
`rHat`	$\hat{r}$	normalized radius	√	√	√							
`d`	d	diameter	√	√	√							
`dHat`	$\hat{d}$	normalized diameter	√	√	√							
`g`	g	girth	√	√	√							
`kappa`	κ	vertex connectivity, if computed	√	√	√							
`kappa1`	κ_1	edge connectivity, if computed	√	√	√							

Table 1 (continued). Notation used throughout this report in equations and data files.

Field Name	Symbol	Description	Data File									
			Summary	Graph	Component	Vertex	Edge	Pair	Degree	Distance	Cycle	Cut
`chi`	χ	vertex chromatic number (upper bound)	√	√	√							
`gamma`	γ	vertex domination number (upper bound), if computed	√	√								
`gammaHat`	$\hat{\gamma}$	normalized vertex domination number (upper bound), if computed	√	√								
`beta`	β	vertex independence number (lower bound)	√	√	√							
`betaHat`	$\hat{\beta}$	normalized vertex independence number (lower bound)	√	√	√							
`I`	i	lower vertex independence number (upper bound)	√	√	√							
`iHat`	$\hat{i}$	normalized lower vertex independence number (upper bound)	√	√	√							
`alpha`	α	vertex covering number (upper bound)	√	√	√							
`alphaHat`	$\hat{\alpha}$	normalized vertex covering number (upper bond)	√	√	√							
`aBar`	$\bar{a}$	fraction of vertices that are articulation points	√	√	√							
`bBar`	$\bar{b}$	fraction of edges that are bridges	√	√	√							
`cBar`	$\bar{c}$	mean length of the shortest cycle at a vertex	√	√	√							
`qBar`	$\bar{q}$	mean size of the minimum cut between a pair of vertices	√	√								
`kBar`	$\bar{k}$	mean degree of a vertex	√	√	√							
`Lbar`	$\bar{L}$	characteristic path length	√	√	√							
`LbarHat`	$\hat{\bar{L}}$	normalized characteristic path length	√	√	√							
`gammaBar`	$\bar{\gamma}$	clustering coefficient	√	√	√	√						
`phiBar`	$\bar{\phi}$	fraction of edges that are shortcuts	√	√	√							
`psiBar`	$\bar{\psi}$	fraction of vertex pairs that are contractions	√	√	√							
`psiStar`	ψ^*	fraction of vertex pairs having overlapping neighborhoods that are contractions	√	√	√							
`elTilde`	$\tilde{\ell}$	normalized entropy of the length distribution of shortest paths	√	√								
`cTilde`	$\tilde{c}$	normalized entropy of the length distribution of shortest cycles	√	√								
`qTilde`	$\tilde{q}$	normalized entropy of the size distribution of minimum cuts	√	√								
`kTilde`	$\tilde{k}$	normalized entropy of the degree distribution	√	√								

Table 1 (continued). Notation used throughout this report in equations and data files.

Field Name	*Symbol*	*Description*	***Data File***									
			Summary	Graph	Component	Vertex	Edge	Pair	Degree	Distance	Cycle	Cut
`k`	k	degree				√			√			
`e`	e	eccentricity				√						
`eHat`	$\hat{e}$	normalized eccentricity				√						
`a`	a	whether the vertex is an articulation point				√						
`L`	L	mean shortest path length				√						
`Lhat`	$\hat{L}$	normalized mean shortest path length				√						
`psi`	ψ	number of contractions				√		√				
`c`	c	shortest cycle length				√					√	
`x`	x	abscissa				√						
`y`	y	ordinate				√						
`b`	b	whether the edge is a bridge					√					
`phi`	ϕ	whether the edge is a shortcut					√					
`s`	s	whether the vertices are connected						√				
`q`	q	size of the minimum cut						√				√
`el`	ℓ	shortest path length								√		
`observations`		number of observations (for a value in a histogram)							√	√	√	√

II. Infrastructure Network Graphs

We have constructed 44 network graphs for a variety of infrastructure types. Table 2 lists the names we have assigned to these graphs along with some measurements related to their size. The data files for the graphs all use the standard XGMML format (see the Appendix).

Table 2. Order and size of infrastructure network graphs.

type	*subtype*	*name*	*order*	*size*	*components*
transportation (1)	airport (1)	air-sample	23	234	1
		air-tickets	655	48624	13
	highways (2)	highways-canada	5232	4718	537
		highways-mexico	518	762	1
		highways-usa	90415	124782	62
	railways (3)	railways-canada	691	843	8
		railways-mexico	179	207	2
		railways-usa	133752	160168	304
	roads (4)	roads-dallas	9863	14750	4
		roads-portland	100852	126161	23
	waterways (5)	waterways-usa	6263	6870	6
	intermodal transport (6)	nts	101588	163932	476
communications (2)	utility communications (7)	communications-bge	36	41	1
		communications-gpc	401	99	303
		communications-jepco	25	25	3
		communications-pepco	38	44	7
		communications-pge	475	322	207
		communications-pnm	50	49	2
		communications-sceg	28	26	2
		communications-sdge	51	45	11
		communications-sepco	7	5	2
		communications-sngc	156	111	49
		communications-tpc	67	49	18
		communications-vepco	195	93	106
	microwave backbones (8)	microwave-fcc	42313	10644	32456
	telephone (9)	telephone-lerg	33854	106253	1434
	internet (10)	internet-bell	122622	162133	605
		internet-nlanr	6474	12572	1
energy (3)	electric power transmission (11)	electric-ecar	4196	5889	2
		electric-ercot	4724	5669	7
		electric-frcc	2082	2605	4
		electric-maac	2836	3701	1
		electric-main	3443	4261	9
		electric-mapp	4478	5449	12
		electric-nepool	2180	2674	6
		electric-nypp	4117	5245	1
		electric-serc	5531	7201	4
		electric-spp	3730	4483	6
		electric-wscc	12226	14696	24
	petroleum transmission (12)	petroleum-mapsearch	160859	170818	3180
	natural gas transmission (12)	gas-florida	32	31	1
		gas-utah	160	159	1
logistics (4)	military logistics (13)	fde	43	310	4
interdependence (5)	interdependence (14)	interdependence	475341	792196	3605

A. Transportation

1. Airports

Our first U.S. Airports graph ("air-sample") comes from a database of airplane flights constructed by Roberto Tamassia using the EasySABRE database of 1992 [Ta 92]. It is a directed graph with the attributes listed in Table 3. Its layout is shown in Figure 1. As might be expected for air travel, this graph is near to being complete—i.e., almost every vertex is directly connected to every other. The major deficiency in this graph is its small size (only major U.S. airports are included). Other than that, it does not exhibit any anomalous data.

Table 3. Attributes of the "air-sample" graph.

attribute	*description*
abbreviation	Airport abbreviation
timezone	Time Zone (offset from GMT)
x	X coordinate on map
y	Y coordinate on map
name	Name of City/Airport
airline	Airline
flight	Flight
departure	Departure Time
arrival	Arrival Time
meals	Meals [S = snack | L = lunch | D = dinner | B = breakfast | # = depends on class]
stops	Stops during flight
aircraft	Aircraft type
booking	Booking Classes

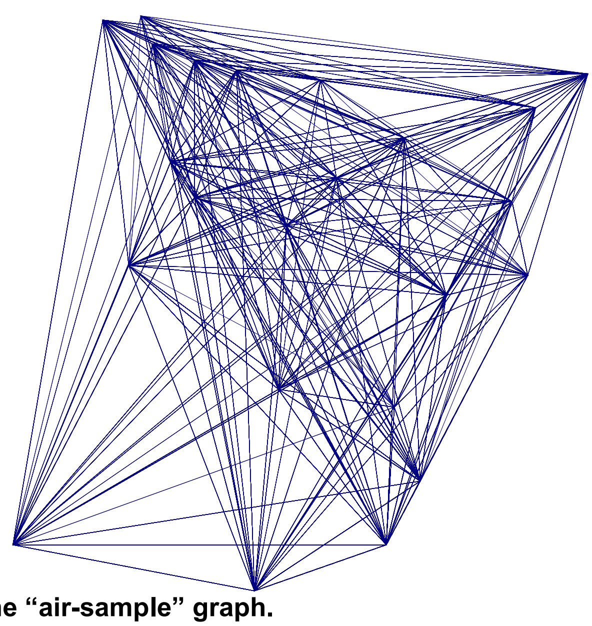

Figure 1. Layout of the "air-sample" graph.

Our second, more extensive U.S. airports graph ("air-tickets") comes from the airline-ticket origin-destination database maintained by the Office of Airline Information in the Bureau of Transportation Statistics [OAI 00]. It is an undirected graph with the attributes listed in Table 4. Figure 2 shows the layout of a representative portion of the graph.

Once again, this graph is fairly complete. We have noticed some internal inconsistencies within the source data used to construct the graph, however, so some edges may be missing in our representation.

Table 4. Attributes of the "air-tickets" graph.

attribute	*description*
code	airport code
area	airport area number
name	airport name
latitude	airport latitude (degrees)
longitude	airport longitude (degrees)
passOutQuart	quarterly outgoing passengers
passInQuart	quarterly incoming passengers
passTotQuart	quarterly total passengers
passOutYear	yearly outgoing passengers
passInYear	yearly incoming passengers
passTotYear	yearly total passengers
domesticQuart	quarterly passengers on international route
domesticYear	yearly passengers on international route
sourceGenerated	yearly passengers generated at source
targetGenerated	yearly passengers generated at target
mileage	mileage between source and target

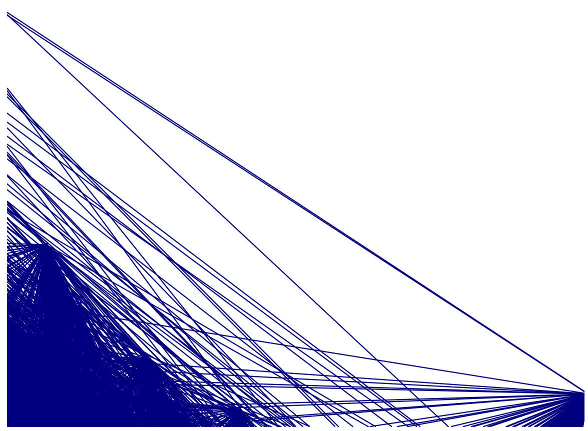

Figure 2. Layout of a representative portion of the "air-tickets" graph.

2. Highways

Our highway system graphs come from the U.S. Bureau of Transportation Statistics NORTAD database [BTS 98]. We have one graph each for the United States, Canada, and Mexico. These are undirected graphs.

The U.S. Highways graph ("highways-usa") has the attributes listed in Table 5. Figure 3 shows the layout of a representative portion of the graph. The vertices in this graph generally have low degree. Most of the components represent geographically disconnected regions such as Hawaii, Alaska, or Puerto Rico. There seems to be very little anomalous data in this graph, although there are some very small components (order less than ten).

Table 5. Attributes of the "highways-usa" graph.

attribute	***description***
LON	The longitude of the node
LAT	The latitude of the node
DESCRIPT	Name or identification for the node/link
SIGN1	Primary Sign Route
SIGN2	Alternate Sign Route
SIGN3	Alternate Sign Route
MILES	An accurate measurement in miles for the link chain
FACTYPE	Describes the permissible flow of traffic over the link [1 = One way flow | 2 = bi-directional flow]
TOLL	Identifies links which have one or more toll features associated with travel along the link [0 = Not A Toll Road | 1 = Toll Road]
LANES	Number of lanes in both directions
ACONTROL	Describes the degree of access [0 = Unknown | 1 – Full Access Control | 2 - Partial Access Control | 3 – No Access Control]
MEDIAN	Describes the type of median [0 - Unknown | 1 – Divided Highway | 2 - Undivided Highway]
SURFACE	Identifies the predominant surface [0 = Unknown | 1 – Paved | 2 - Unpaved | 3 - Ferry]
FCLASS	Identifies the assigned functional class of each link [01 = Rural Interstate | 02 = Rural Principal Arterial | 06 = Rural Minor Arterial | 07 = Rural Major Collector | 08 = Rural Minor Collector | 09 = Rural Local | 11 = Urban Interstate | 12 = Urban Freeway or Expressway | 14 = Urban Principal Arterial | 16 = Urban Minor Arterial | 17 = Urban Collector | 19 = Urban Local]
RU_CODE	Rural/Urban classification [1 = Rural | 2 = Small Urban (1990 pop 5,000-49,999) | 3 = Large Urban (1990 > 50,000)]
STATUS	Describes availability of the link [0 = Proposed/Under construction | 1 = Open to traffic]

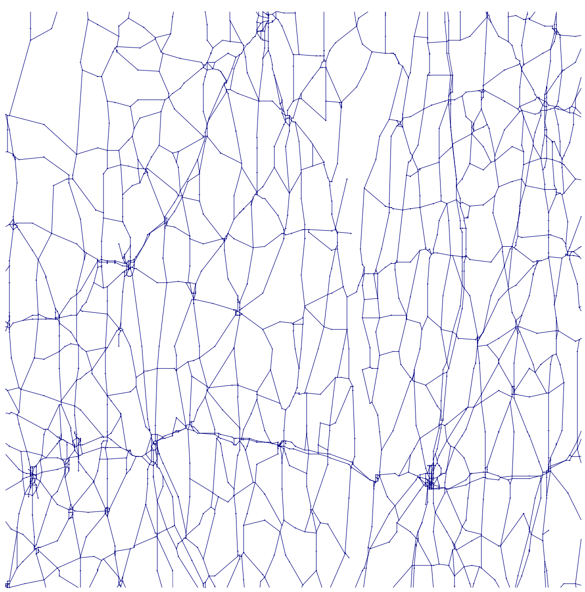

Figure 3. Layout of a representative portion of the "highways-usa" graph.

The Canada Highways graph ("highways-canada") has the attributes listed in Table 6. Figure 4 shows the layout of a portion of the graph. Once again, the vertices in this graph have low degree. There seems to be a lot of contamination in the data, however: the graph has many disconnected components, seems very non-planar, and has edges that possess an exceeding long geographic extent. Because of these problems, we will exclude this graph from many of the analyses we perform for the other graphs.

Table 6. Attributes of the "highways-canada" graph.

attribute	*description*
LON	The longitude of the node
LAT	The latitude of the node
DESCRIPT	Name or identification for the node/link
LANES	Number of lanes
SPD_LIM	Speed limit in kilometers

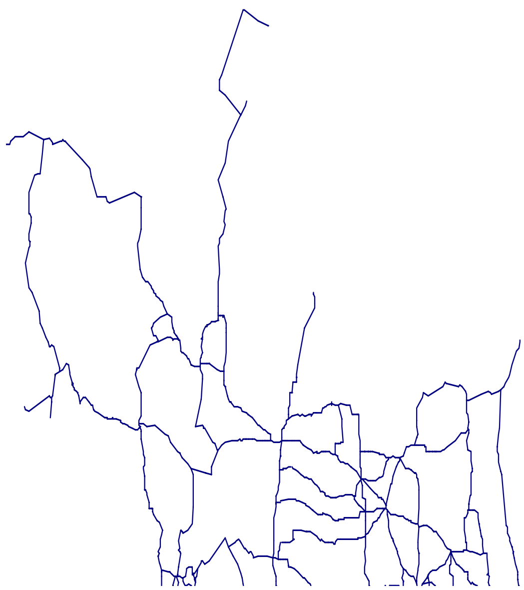

Figure 4. Layout of a portion of the "highways-canada" graph.

The Mexico Highways graph ("highways-mexico") has the attributes listed in Table 7. Figure 5 shows the layout of a representative portion of the graph. This is very similar to the U.S. Highways graph. It only has a single component and seems free of anomalies.

Table 7. Attributes of the "highways-mexico" graph.

attribute	*description*
LON	The longitude of the node
LAT	The latitude of the node
DESCRIPT	Name or identification for the node/link
NAME	Highway Name
MODE	Highway Mode [AUTOPISTA | CARRETERO | TRANSBORDADOR]
CLASS	Highway Classification
LANES	Number of Lanes
TERRAIN	Type of topography [LOMERIO | PLANO | MONTANOSO]
RUR_URB	Rural / Urban categorization [RURAL | SUBURBANO | URBANO |]

Figure 5. Layout of a representative portion of the "highways-mexico" graph.

3. Railways

Our railway system graphs come from the U.S. Bureau of Transportation Statistics NORTAD database [BTS 98]. We have one graph each for the United States, Canada, and Mexico. These graphs are undirected.

The U.S Railways graph ("railways-usa") has the attributes listed in Table 8. Figure 6 shows the layout of a representative portion of it. Most of the vertices in this graph are of low degree, including an unusually large number of degree-two vertices. The degree-two vertices may be an artifact of the geographic representation of the graph—i.e., these may be geographic "shape" points rather than locations where a station exists or where the properties of the rail line change. Aside from there being quite a few components of small order, the graph has no major anomalies.

Table 8. Attributes of the "railways-usa" graph.

attribute	*description*
LON	The longitude of the node
LAT	The latitude of the node
DESCRIPT	Name or identification for the node/link
RROWNER	Railroad Owner Name Abbreviation

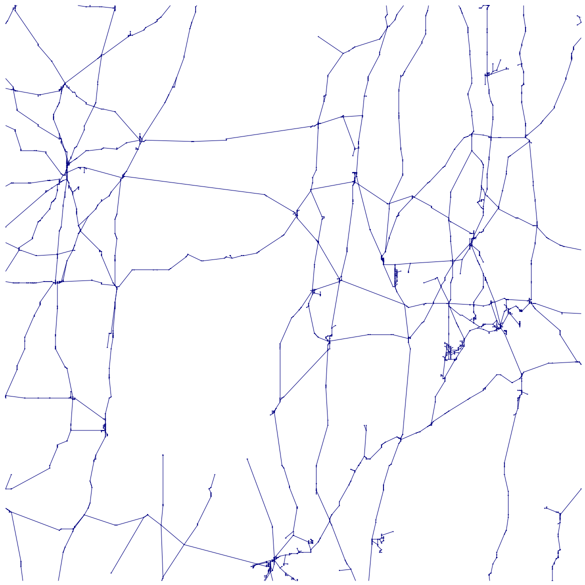

Figure 6. Layout of a representative portion of the "railways-usa" graph.

The Canada Railways graph ("railways-canada") has the attributes listed in Table 9. Figure 7 shows the layout of a representative portion of it. Although it lacks the problems noted in the Canada Highways graph, its geographic layout is somewhat suspicious because it seems very non-planar. There are only a few small components and vertices having small degree, however.

Table 9. Attributes of the "railways-canada" graph.

attribute	***description***
LON	The longitude of the node
LAT	The latitude of the node
DESCRIPT	Name or identification for the node/link
MILES	Length of link, in miles

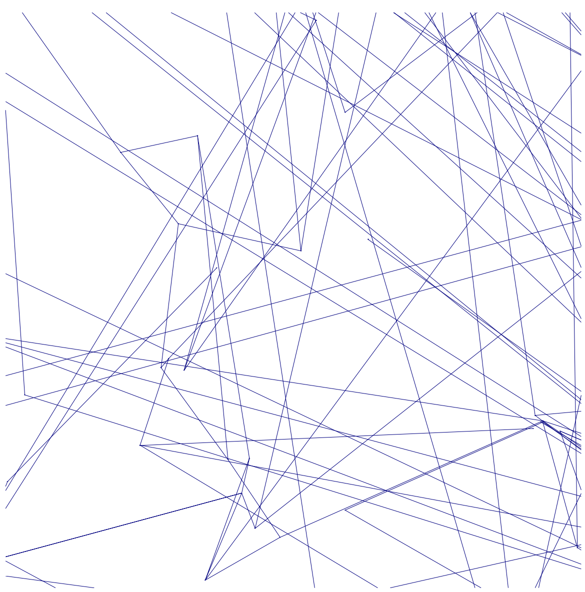

Figure 7. Layout of a representative portion of the "railways-canada" graph.

The Mexico Railways graph ("railways-mexico") has the attributes listed in Table 10. Figure 8 shows the layout of a representative portion of it. The vertices have small degree and there are only a couple of small components—nearly all of the vertices are in the largest component. Although the layout shows a lot of overlapping edges, there seem to be few anomalies in the graph.

Table 10. Attributes of the "railways-mexico" graph.

attribute	*description*
LON	The longitude of the node
LAT	The latitude of the node
DESCRIPT	Name or identification for the node/link
MODE	Rail mode [F.C. DIESEL]
LINE	Railroad Line
TYPE	Rail type [VIA DOBLE | VIA SENCILLA]

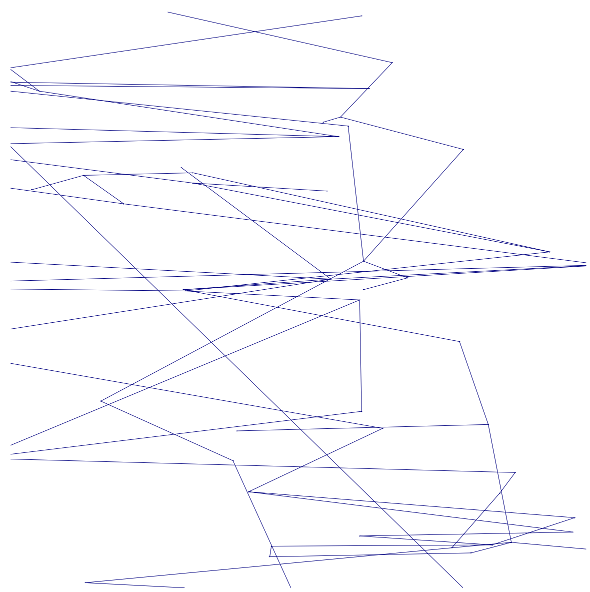

Figure 8. Layout of a representative portion of the "railways-mexico" graph.

4. Roads

Our roadway graphs come from databases compiled for the TRANSIMS project at Los Alamos National Laboratory [TRA 01]. These directed graphs represent the detailed layout of roads within a city.

The Dallas Galleria Roads graph ("roads-dallas") has the attributes listed in Table 11. Figure 9 shows the layout of a representative portion of it. This network contains all of the major roads in the Dallas-Ft. Worth, U.S.A., area and all of the local streets in a three-mile by three-mile area centered around the Galleria shopping center, as of 1994. Most of the anomalies in this graph were previously corrected while preparing it for use in a traffic microsimulation.

Table 11. Attributes of the "roads-dallas" graph.

attribute	*description*
EASTING	UTM easting for the node (meters)
NORTHING	UTM northing for the node (meters)
TAPA	Transportation analysis zone
LANES	Number of lanes
LENGTH	Length of link (meters)
SPEEDLMT	Speed limit on link (meters/second)
FREESPD	Free-flow speed on link (meters/second)
FUNCTCLASS	Functional class on link [COLLECTOR | FREEWAY | FUNCTCLASS | LIGHTRAIL | LOCAL | OTHER | PRIARTER | RAMP | SECARTER | WALKWAY | XPRESSWAY]

Figure 9. Layout of a representative portion of the "roads-dallas" graph.

The Portland Roads graph ("roads-portland") has the attributes listed in Table 12. Figure 10 shows the layout of a representative portion of it. This network contains all of the roads in the greater Portland, Oregon, U.S.A., area, as of 1996. Once again, most of the anomalies in this graph were previously corrected while preparing it for use in a traffic microsimulation.

Table 12. Attributes of the "roads-portland" graph.

attribute	*description*
EASTING	UTM easting for the node (meters)
NORTHING	UTM northing for the node (meters)
ELEVATION	Elevation of node (meters)
TAPA	Transportation analysis zone
NAME	Name of link
LANES	Number of lanes
LENGTH	Length of link (meters)
SPEEDLMT	Speed limit on link (meters/second)
FREESPD	Free-flow speed on link (meters/second)
FUNCTCLASS	Functional class on link [COLLECTOR | FREEWAY | FUNCTCLASS | LIGHTRAIL | LOCAL | OTHER | PRIARTER | RAMP | SECARTER | WALKWAY | XPRESSWAY]
VEHICLE	Vehicles allowed on link [AUTO | BUS | LIGHTRAIL | WALK]

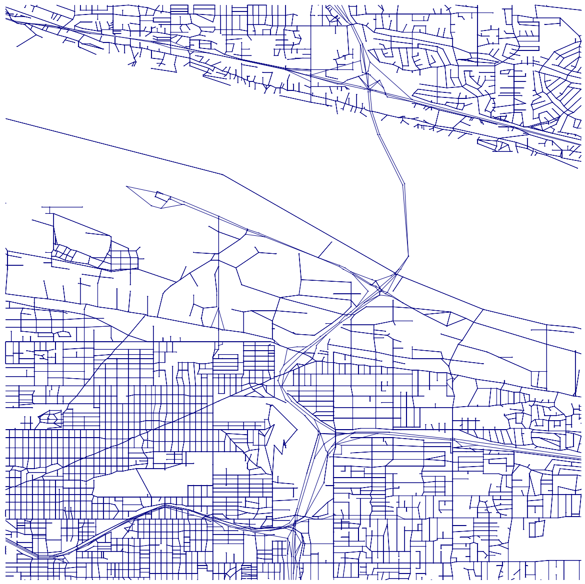

Figure 10. Layout of a representative portion of the "roads-portland" graph.

5. Waterways

Our U.S. Waterway graph ("waterways-usa") comes from the U.S. Bureau of Transportation Statistics NTAD database [BTS 00]. This graph is undirected and has the attributes listed in Table 13. Figure 11 shows the layout of a representative portion of it. Nearly all of the vertices belong to the largest component, but there are a couple of very small components. There are several vertices of high degree, as might be expected for major ports. There seem to be no major anomalies in this graph.

Table 13. Attributes of the "waterways-usa" graph.

attribute	***description***
LON	The longitude of the node
LAT	The latitude of the node
DESCRIPT	Name or identification for the node/link
LLEN	Line length, in miles
GEO	Geographic Class [G = Great Lakes | O = Ocean / Offshore | I = Inland]
FUNC	Functional Class [N = no traffic, non-navigable | S = shallow draft (i.e., no deep draft ocean vessels) | D = deep draft | B = both | U = special vessels only (fishing, pleasure craft, etc; normally no freight traffic)]
WTYPE	Waterway Type [1 = Harbor (including harbor channels), Bay | 2 = Intracoastal Waterway | 3 = Sea-lane | 4 = Sea-lane with separation zone | 5 = Open water | 6 = River, creek, thoroughfare, lake | 7 = Estuary | 8 = Channel (not including harbor channels) | 9 = Canal | 10 = Great Lakes direct link (major ports) | 11 = Great Lakes indirect link | 12 = Corps of Engineers Lock]

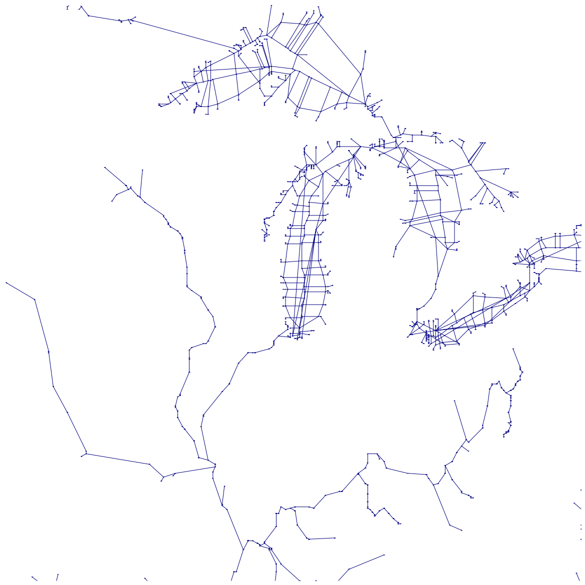

Figure 11. Layout of a representative portion of the "waterways-usa" graph.

6. Intermodal Transportation

Our National Transportation System graph ("nts") comes from Oak Ridge National Laboratory [Oa 97]. It has the attributes listed in Table 14. Figure 12 shows the layout of a representative portion of it. This graph is directed, but because some inconsistencies exist in the data, so it should be used as an undirected graph. NTS has many components, some of which are explained because they are islands. Others were deemed errors in the data. There are also vertices of unreasonably high degree, and vertices of degree zero.

Table 14. Attributes of the "nts" graph.

attribute	***description***
DESCRIPTION	Description of node
MODE	Mode of node/link
LAT	Latitude of node
LON	Longitude of node
IMPEDANCE	Impedance of link

Figure 12. Layout of a representative portion of the "nts" graph.

B. Communications

1. Utility Communications

Our utility communications graphs come from data collected for the IAAP project at Los Alamos National Laboratory [IAA 01]. These have the attributes listed in Table 15. They were constructed from the FCC Master Frequency Database [Pe 95], using publications by the utilities (journal articles, conference proceedings, etc.), and by directly contacting the utilities. Because complete information about the communication system was often unavailable, these graphs typically lack certain types of components: copper wire connections are often not listed, but microwave links always are. These are directed graphs. Note that the electric power and natural gas utilities have graphs of different character.

Table 15. Attributes of the utility communications graphs.

attribute	***description***
name	Name of the node
latitude	Latitude of the node
longitude	Longitude of the node
medium	Type of transmission medium [FO = optical fiber | MW-D = digital microwave | MW-A = analog microwave | MW-H = microwave hybrid | TEL = telephone | CATV = cable television]

The layout of the Baltimore Gas & Electric Communications graph ("communications-bge") is show in Figure 13. The microwave and optical fiber connections are completely represented in this graph.

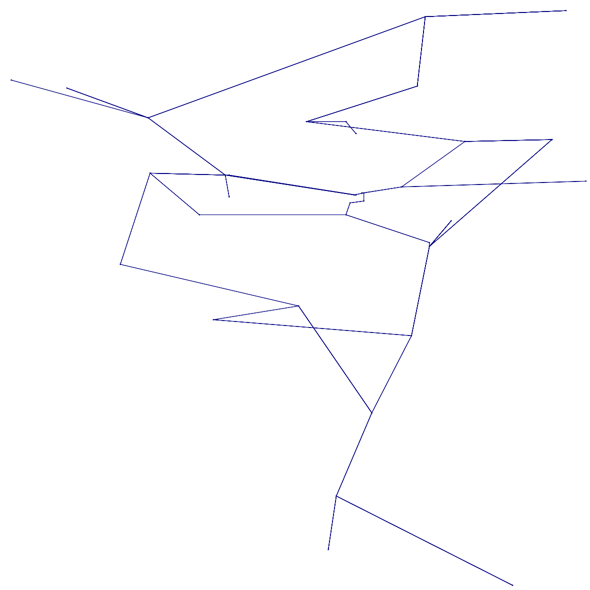

Figure 13. Layout of the "communications-bge" graph.

The layout of the Georgia Power Corporation Communications graph ("communications-gpc") is shown in Figure 14. The microwave links are completely represented, but the copper wire connections are absent; this accounts for the presence of isolated vertices and disconnected components.

Figure 14. Layout of the "communications-gpc" graph.

The layout of the Jacksonville Electric & Power Corporation Communications graph ("communications-jepco") is shown in Figure 15. This is a complete representation of the microwave network.

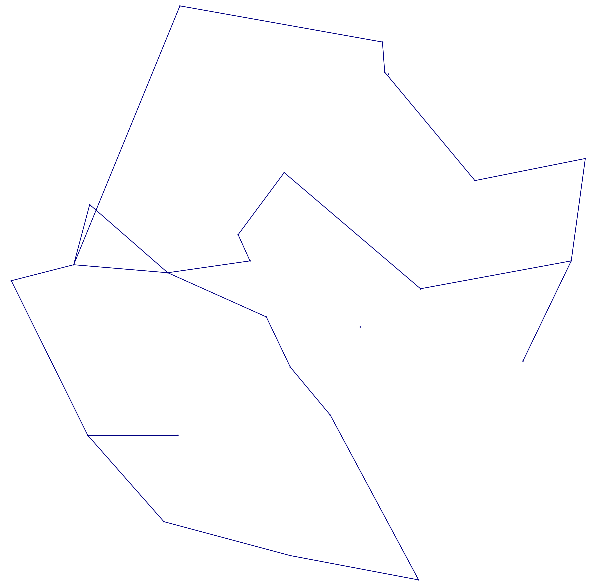

Figure 15. Layout of the "communications-jepco" graph.

The layout of the Potomac Electric & Power Corporation Communications graph ("communications-pepco") is shown in Figure 16. This is a complete representation of the microwave and optical fiber components.

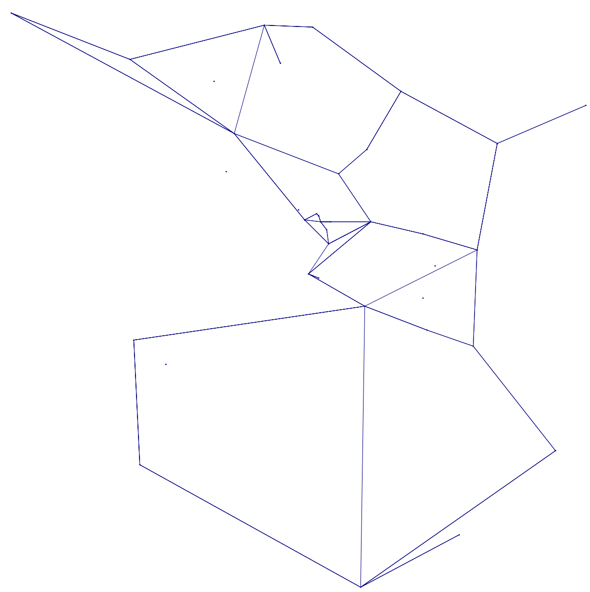

Figure 16. Layout of the "communications-pepco" graph.

The layout of the Pacific Gas & Electric Communications graph ("communications-pge") is shown in Figure 17. This is a complete representation of the microwave links, but no copper wire connections are present.

Figure 17. Layout of the "communications-pge" graph.

The layout of the Public Service Company of New Mexico Communications graph ("communications-pnm") is shown in Figure 18. This is a complete representation of the microwave and optical fiber networks.

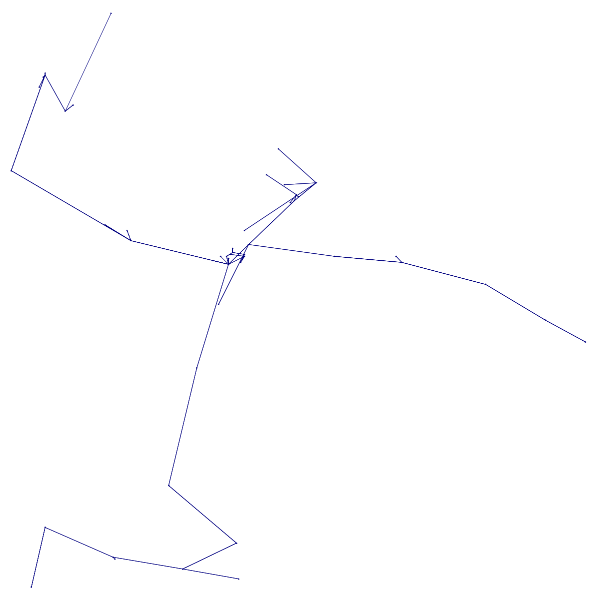

Figure 18. Layout of the "communications-pnm" graph.

The layout of the South Carolina Electric & Gas Communications graph ("communications-sceg") is shown in Figure 19. This is a complete representation of the microwave network.

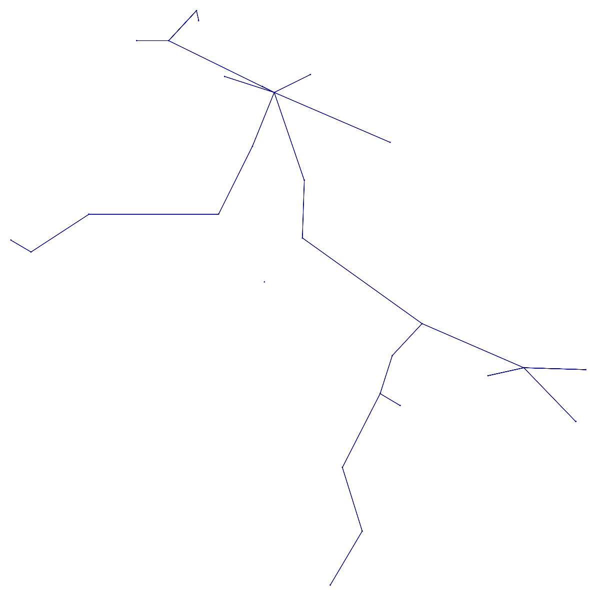

Figure 19. Layout of the "communications-sceg" graph.

The layout of the San Diego Gas & Electric Communications graph ("communications-sdge") is shown in Figure 20. This is a complete representation of the microwave, optical fiber, and copper wire networks.

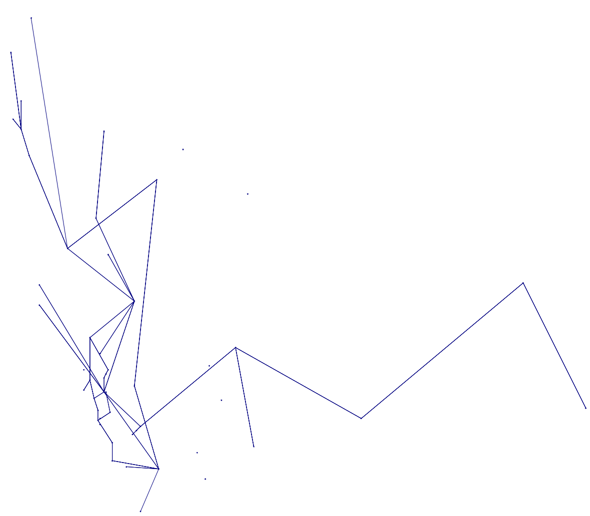

Figure 20. Layout of the "communications-sdge" graph.

The layout of the Savannah Electric & Power Company Communications graph ("communications-sepco") is shown in Figure 21. This is a complete representation of the microwave network.

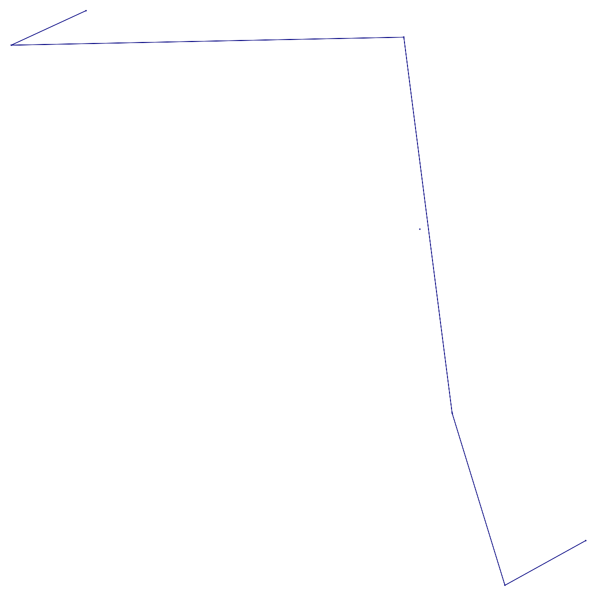

Figure 21. Layout of the "communications-sepco" graph.

The layout of the Southern Natural Gas Company Communications graph ("communications-sngc") is shown in Figure 22. This is a complete representation of the microwave network.

Figure 22. Layout of the "communications-sngc" graph.

The layout of the Transcontinental Corporation Communications graph ("communications-tpc") is shown in Figure 23. This is a complete representation of the microwave network.

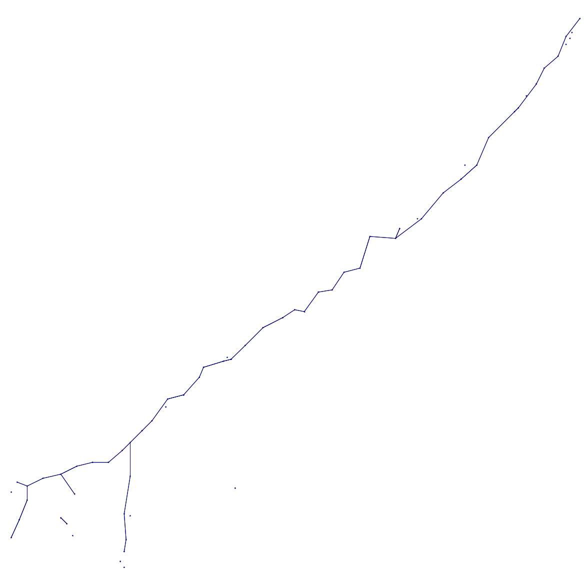

Figure 23. Layout of the "communications-tpc" graph.

The layout of the Virginia Electric & Power Company Communications graph ("communications-vepco") is shown in Figure 24. This is a complete representation of the microwave network, but lacks the copper wire connections.

Figure 24. Layout of the "communications-vepco" graph.

2. Microwave Backbones

Our source for non-governmental microwave communication networks in the United State is the U.S. Federal Communication Commission's Master Frequency Database [Pe 95]. This is a directed graph with the attributes listed in Table 16. Figure 25 shows the layout of a representative portion of it. Because systems owned by different organizations are typically not connected, this graph contains many disconnected components. There are also many isolated vertices. The data contain relatively few other anomalies, however.

Table 16. Attributes of the "microwave-fcc" graph.

attribute	*description*
id	Numeric ID of the node/link
name	Name of the node
latitude	Latitude of the node
longitude	Longitude of the node
medium	Type of transmission medium [MW = microwave | UHF-H = high-band UHF]

Figure 25. Layout of a representative portion of the "microwave-fcc" graph.

3. Telephone

Our source for the telephone network graph ("telephone-lerg") is the Local Exchange Routing from Bellcore [Be 98]. This is an undirected graph. Anomalies are almost completely absent from this graph because it was constructed from a working database used for routing communications between local telephone exchange switches. Many of the nodes have high degree. Note that this graph does not include information on long-distance (inter-LATA) service, as most of such data is proprietary.

Table 17. Attributes of the "telephone-lerg" graph.

attribute	***description***
lata	Local access and transport area
switch	Switch ID
type	Equipment type
atc	Whether the central office is an access tandem
eoc	Whether the central office is and end office
pmc	Whether the central office is a public mobile carrier
rcc	Whether the central office is a radio common carrier
owner	Owner name
street	Street address
city	City
state	State abbreviation
zip	Zip code
longitude	Longitude
latitude	Latitude
homing	Homing information bits: [0 = originating feature group B tandem | 1 = originating feature group C tandem | 2 = originating feature group D tandem | 3 = originating operator services tandem | 4 = originating FG B intermediate tandem | 5 = originating FG C intermediate tandem | 6 = originating FG D tandem | 7 = originating local tandem | 8 = originating intraLATA tandem | 9 = originating circuit switched data tandem | 10 = terminating feature group B tandem | 11 = terminating feature group C tandem | 12 = terminating feature group D tandem | 13 = terminating operator services tandem | 14 = terminating FG B intermediate tandem | 15 = terminating FG C intermediate tandem | 16 = terminating FG D tandem | 17 = terminating local tandem | 18 = terminating intraLATA tandem | 19 = terminating circuit switched data tandem | 20 = host | 21 = STP1 | 22 = STP2 | 23 = originating 500 SSP | 24 = originating 800 SSP | 25 = ISDN foreign served office | 26 = actual switch ID | 27 = LNP SCP1 | 28 = LNP SCP2 | 29 = AIN ISCP/SCP1 | 30 = AIN ISCP/SCP2]

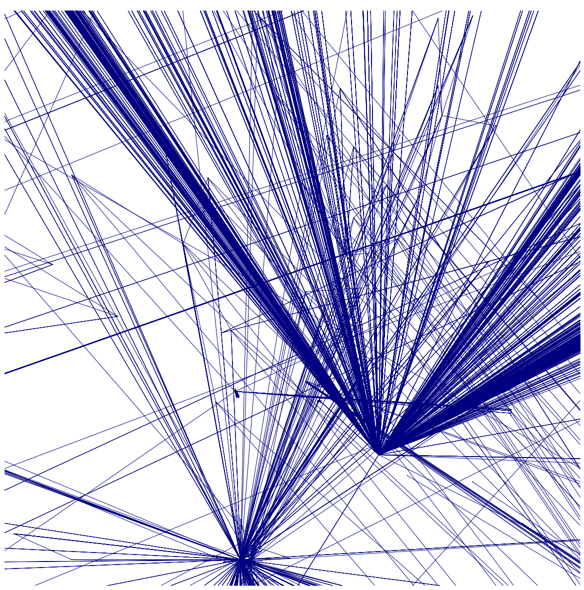

Figure 26. Layout of a representative portion of the "telephone-lerg" graph.

4. Internet

Our first source for an Internet graph ("internet-nlanr") comes from the Internet Autonomous Systems database maintained by the National Laboratory for Applied

Network Research [Na 00]. An autonomous system is a “collection of gateways (routers) that fall under one administrative entity and cooperate using a common Interior Gateway Protocol (IGP).” This provides complete high-level coverage of the Internet, but does not contain information about routing within autonomous systems. There are no attributes for this graph. Figure 27 shows the layout of a representative portion of the graph.

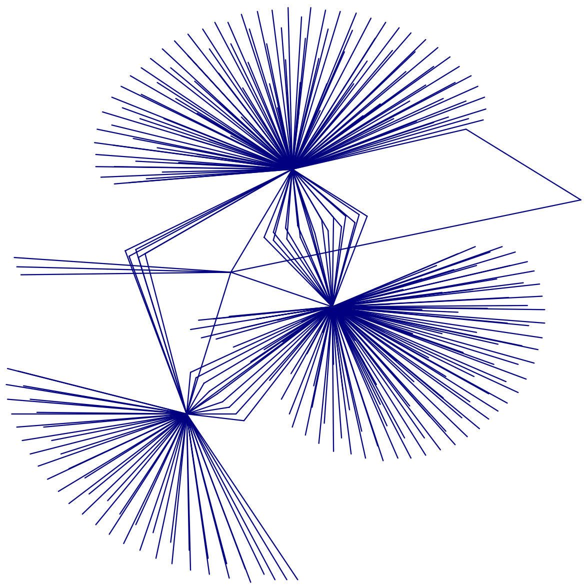

Figure 27. Layout of a representative portion of the “internet-nlanr” graph.

Another source for an Internet graph (“internet-bell”) is the Internet Mapping Project at Bell Labs [CB 00]. This graph was constructed by tracing the routes of UDP probes to randomly selected Internet hosts. Because of this construction method, this graph is likely to be biased in its coverage. Table 18 lists the attributes of this undirected graph. Figure 28 shows a layout of a representative portion of the graph.

Table 18. Attributes of the “internet-bell” graph.

attribute	*description*
IP1	First byte of IP address
IP2	Second byte of IP address
IP3	Third byte of IP address
IP4	Fourth byte of IP address

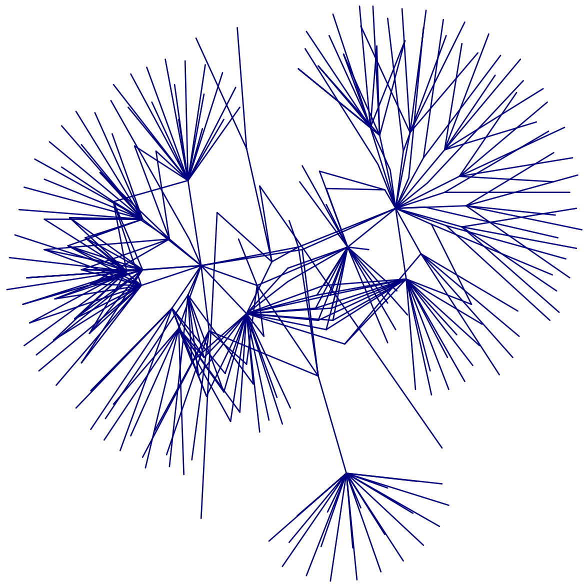

Figure 28. Layout of a representative portion of the "internet-bell" graph.

C. Energy

1. Electric Power Transmission

Our electric power system graphs come from data collected for the IAAP project at Los Alamos National Laboratory [IAA 01]. These undirected graphs have the attributes listed in Table 19. They were constructed from the Federal Energy Regulatory Commission [Fe 99] and Energy Information Administration [EIA 99] filings. The latitudes and longitudes are not known, but approximated by the IAAP project. Complete information about the transmission system is often unavailable, so some vertices remain equivalenced. Distribution networks are not included. Due to filing errors, a few disconnected sub-components exist in each of the electric power networks. In general, the vertices are of low degree, with a few high degree vertices consistent with major hubs in cities. Figure 29 shows the geographic location of the various electric power system graphs described below.

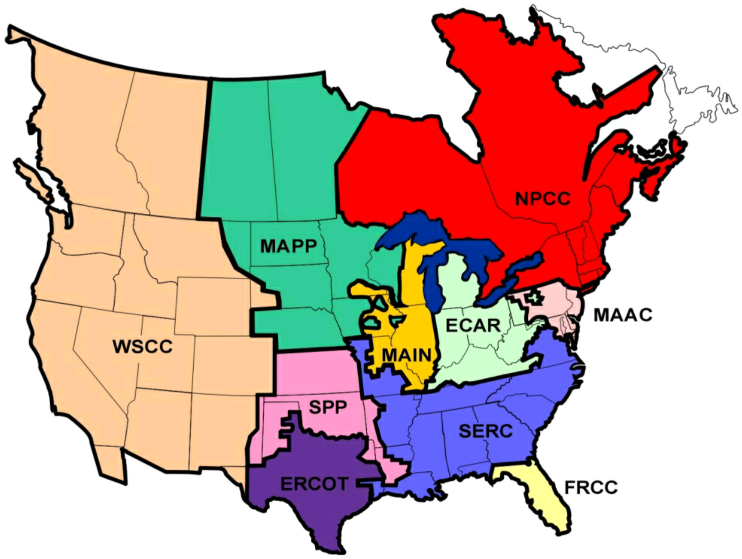

Figure 29. Geographic extent of NERC regions.

Table 19. Attributes of the electric power system graphs.

attribute	*description*
id	ID string for the bus/line
name	Name of the bus
latitude	Latitude of the bus
longitude	Longitude of the bus
intertie	Whether the bus is an intertie [1 = yes | 0 = no]
generator	Whether the bus has a generator [1 = yes | - = no]
voltage	Voltage at bus (kV)
load mw	Real load at bus (MW)
load mvar	Reactive load at bus (MVAR)
mw limit	Maximum real power at bus (MW)
mvar limit	Maximum reactive power at bus (MVAR)
max capac mvar	Maximum capacitive compensation at bus (MVAR)
max react mvar	Maximum reactive compensation at bus (MVAR)
capac banks	Number of capacitor banks at bus
react banks	Number of reactor banks at bus
transformer count	Number of transformers at bus
length	Length of line (miles)
maximum mva	Maximum power for line (MVA)
design voltage	Design voltage of line (kV)
circuits	Number of circuits
lines	Number of lines
inductive reactance	Inductive reactance of line
capacitive reactance	Capacitive reactance of line
resistance	resistance of line
line charging	line charging capacitance
ac	Whether line is A.C [1 = A.C | 0 = D.C.]
overhead	Whether line is overhead [1 = overhead | 0 = underground]
in service	Whether line is normally in service [1 = yes | 0 = no]

The layout of a representative portion of the ECAR Power System graph ("electric-ecar") is shown in Figure 30. ECAR is located in the north central United States. It comprises of five states from Tennessee to Michigan. The data for ECAR from FERC included data from the neighboring NERC regions. Thus, some errors might have been introduced extracting ECAR. Regardless, very few errors were found in the data.

Figure 30. Layout of a representative portion of the "electric-ecar" graph.

The layout of a representative portion of the ERCOT Power System graph ("electric-ercot") is shown in Figure 31. ERCOT comprises of a region covering most of Texas. The ERCOT data is consistently of high quality.

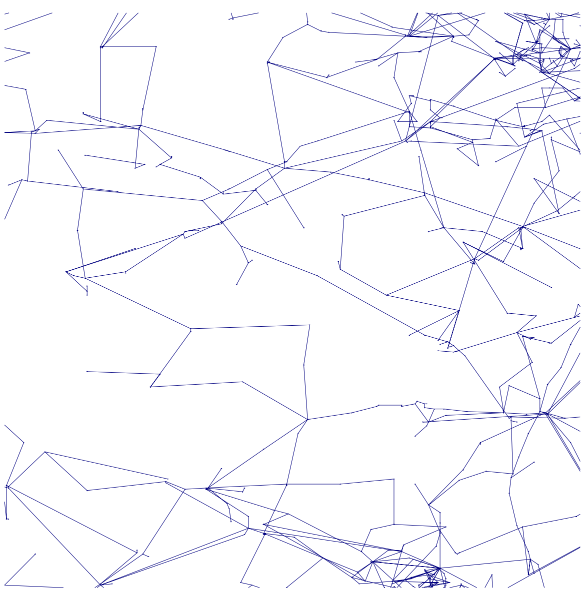

Figure 31. Layout of a representative portion of the "electric-ercot" graph.

The layout of a representative portion of the FRCC Power System graph ("electric-frcc') is shown in Figure 32. FRCC covers most of Florida, from the panhandle to the keys. The data is in good shape, although the FRCC data had to be extracted from its bordering

NERC region of SERC. FRCC filed the fewest number of vertices (substations, busses, taps, etc) of all the regions.

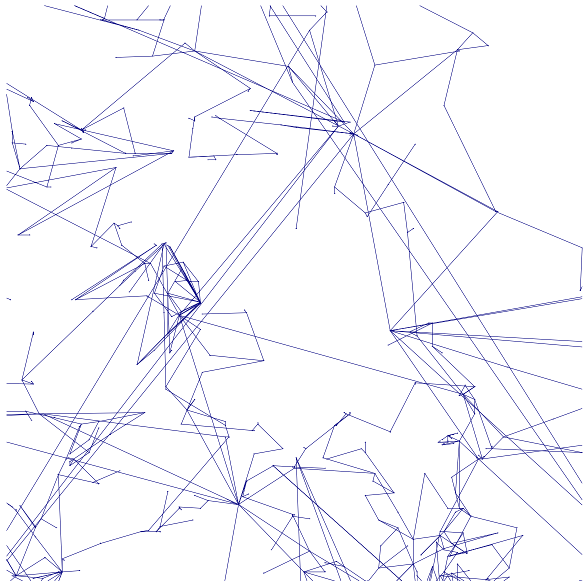

Figure 32. Layout of a representative portion of the "electric-frcc" graph.

The layout of a representative portion of the MAAC Power System graph ("electric-maac") is shown in Figure 33. The majority of MAAC is in Pennsylvania, although it covers several bordering states as well. Geographically it is one of the smallest regions, but it does not have the fewest number of electrical components. MAAC is one of only two regions to have only one graph component, an indication that the data for MAAC is in great shape.

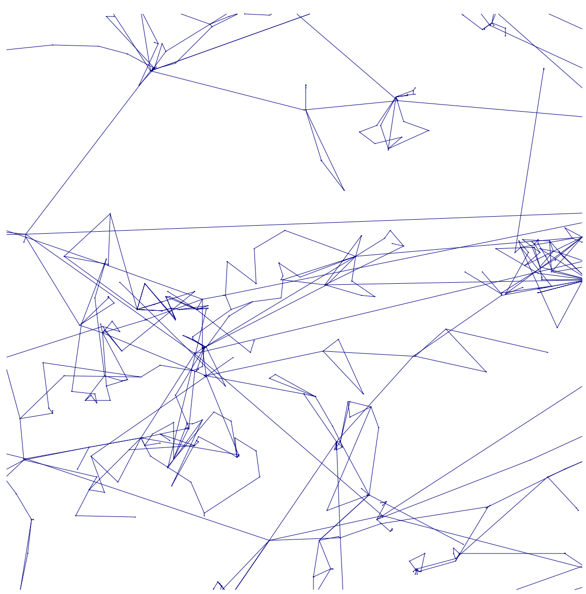

Figure 33. Layout of a representative portion of the "electric-maac" graph.

The layout of a representative portion of the MAIN Power System graph ("electric-main") is shown in Figure 34. MAIN is located in the north central United States. It covers Wisconsin, and much of the bordering area. MAIN's data had to be extracted from the neighboring areas of MAPP, ECAR and SERC. A few data errors might have been introduced in this process, although the data looked very good overall.

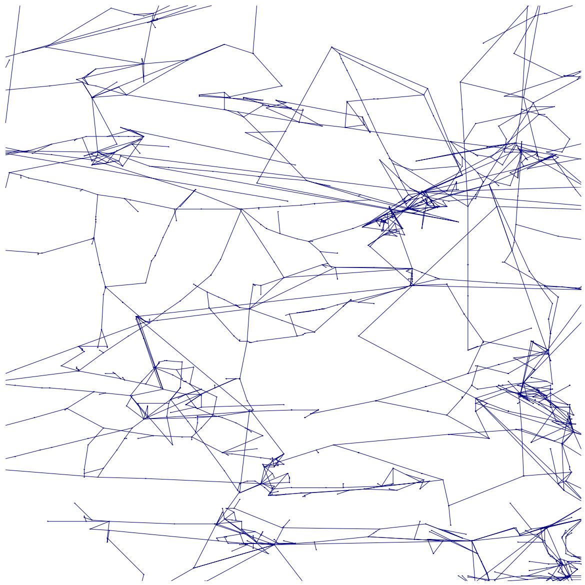

Figure 34. Layout of a representative portion of the "electric-main" graph.

The layout of a representative portion of the MAPP Power System graph ("electric-mapp") is shown in Figure 35. MAPP covers a very large geographic area in the north central United States and into Canada. It includes the states of Nebraska, South Dakota, North Dakota, Minnesota, Iowa, and the provinces of Saskatchewan and Manitoba. The data is pretty good, although there were several isolated vertices. Confidence in the data is restricted to the data in the United States—because this data is collected to satisfy an American mandate, the effort used to file the Canadian data is questionable.

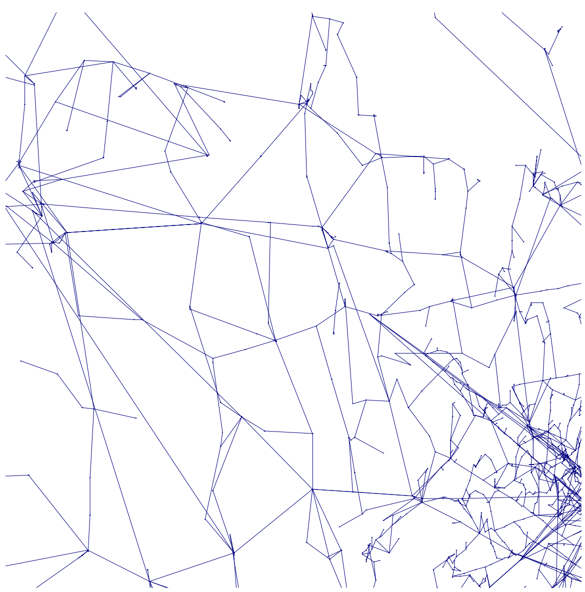

Figure 35. Layout of a representative portion of the "electric-mapp" graph.

The layout of a representative portion of the NEPOOL Power System graph ("electric-nepool") is shown in Figure 36. NEPOOL is located in northeastern United States, and into Canada. It includes all of the states east of New York, and Quebec. It also is one of the smallest NERC regions, in terms of number of electrical components.

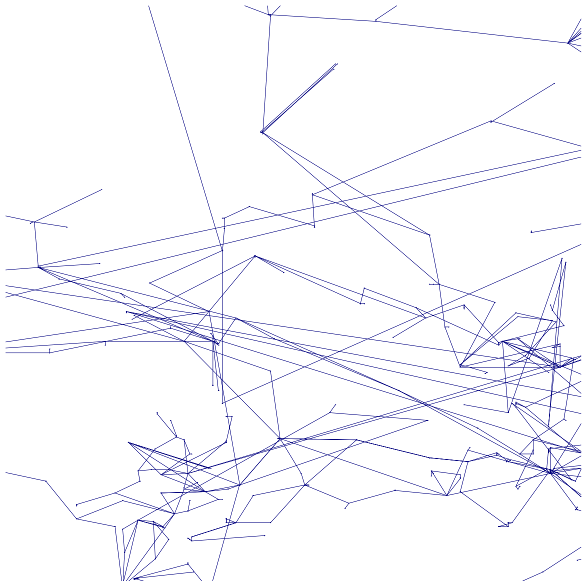

Figure 36. Layout of a representative portion of the "electric-nepool" graph.

The layout of a representative portion of the NYPP Power System graph ("electric-nepool") is shown in Figure 37. NYPP consists of New York and Ontario. It is one of only two NERC regions to have no isolated components in the data. This is an indication of high quality data in NYPP.

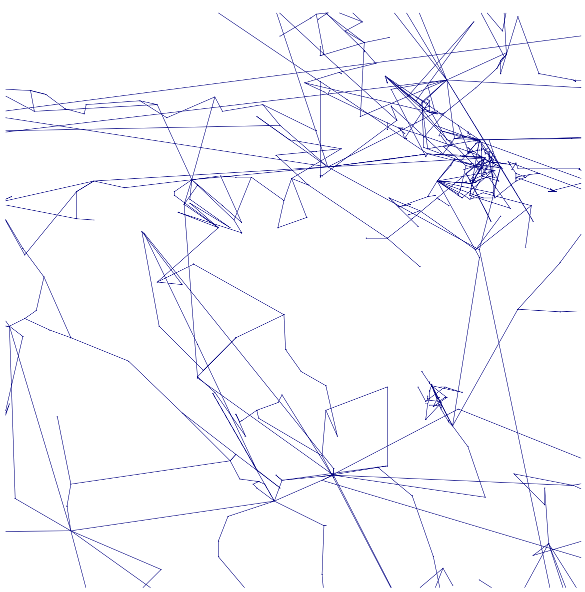

Figure 37. Layout of a representative portion of the "electric-nypp" graph.

The layout of a representative portion of the SERC Power System graph ("electric-serc") is shown in Figure 38. SERC is the second largest NERC region, both geographically and in terms of number of electrical components. It includes all of the southeast United

States, excluding Florida. In all, there are nine states either wholly or partially in SERC. Given its size, it is a surprise that there were so few data errors. For instance, there are only three isolated components in the entire region.

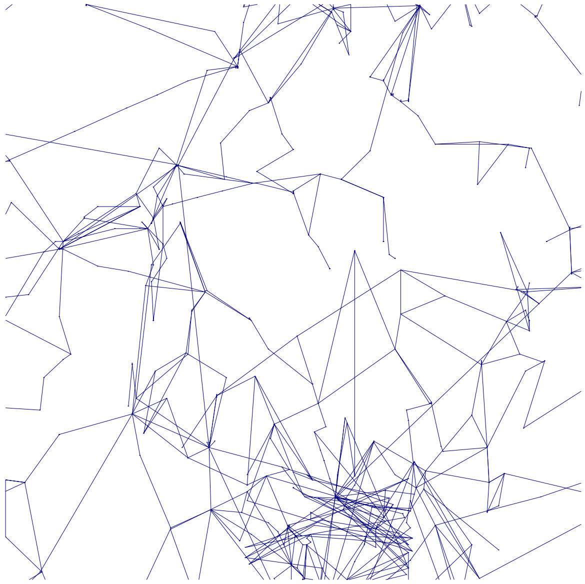

Figure 38. Layout of a representative portion of the "electric-serc" graph.

The layout of a representative portion of the SPP Power System graph ("electric-spp") is shown in Figure 39. SPP is located just north of Texas. It contains Kansas, Oklahoma, part of Louisiana and northern Texas. The data quality seems to be typical of the NERC regions.

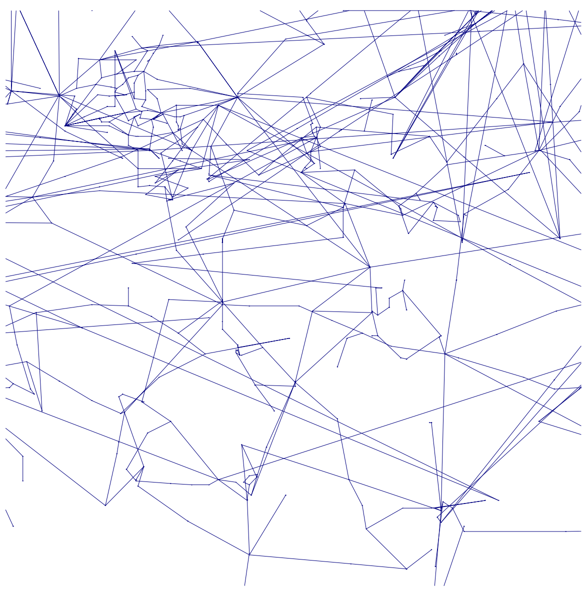

Figure 39. Layout of a representative portion of the "electric-spp" graph.

The layout of a representative portion of the WSCC Power System graph ("electric-wscc") is shown in Figure 40. WSCC consists of all of the western United States and parts of Mexico, Alberta and British Columbia. It is connected to the rest of the US power grid by a few AC-DC-AC connections. WSCC filings changed dramatically in the past few years when the number of vertices filed went from eight thousand to over twelve thousand. It is the largest NERC region, in terms of both electrical components and geographic size. It also had the largest number of isolated vertices.

Figure 40. Layout of a representative portion of the "electric-wscc" graph.

2. Petroleum Transmission

Our petroleum transmission system graph comes from the Pipeline, Facility & Interconnect Inventory developed by Penwell Mapsearch [Pe 98]. It is an undirected graph with the attributes listed in Table 20. Figure 41 shows the layout of a representative portion of the graph. We performed extensive processing of this data set in order to put in into the form of a graph—the source data did not explicitly provide information on how pipeline segments connect with one another. The graph is relatively clean for a data set of this size, but it does contain disconnected components due to the variety of companies and products represented.

Table 20. Attributes of the "petroleum-mapsearch" graph.

attribute	*description*
LAT	The latitude of the node
LON	The longitude of the node
PIP_ID	The ID of the pipeline
SEG_NO	The segment number of the pipeline
OWN_ID	The company that owns the pipeline
OPR_ID	The company who operates the pipeline
CMDTY	The commodity type [C = Crude oil | L = LPG/NGL | N = Natural gas | U = Unknown | R = Refined products | P = Petrochemical | M = Miscellaneous]
LTYPE	The classification of pipe [GT = Gathering system field line | TL = Gathering system main line | TR = Transmission/trunk line | LD = Local distribution]
FLOW	Unknown
DIAMETER	The diameter of the pipe (inches)
CMDTY_2	The commodity subtype (see comment)
STATUS	The status of the pipeline [AC = active | AB = abandoned | ID = idle/inactive | PR = proposed]
INT_ID	The ID of the interconnect

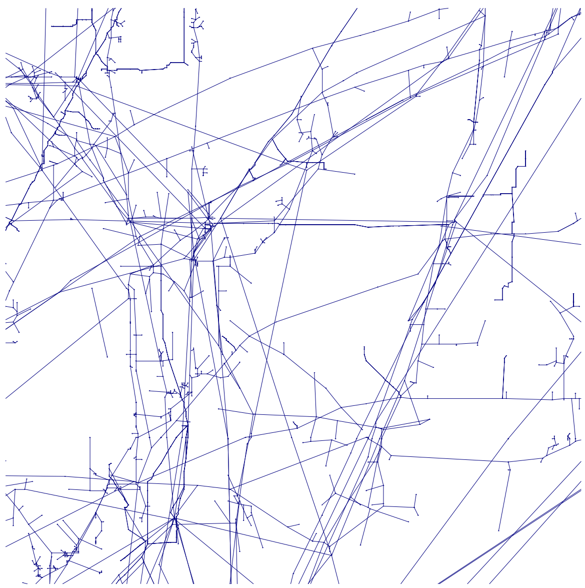

Figure 41. Layout of a representative portion of the "petroleum-mapsearch" graph.

3. Natural Gas Transmission

Our natural gas transmission system data for the states of Florida and Utah were constructed by Cetin Unal [Un 01] from databases provided by Argonne National Laboratory. These are undirected graphs with the attributes listed in Table 21. Figure 42 and Figure 43 illustrate the layout of these graphs. These carefully constructed representations of the natural gas networks in these states form a fairly complete description of the transmission systems there.

Table 21. Attributes of the "gas-florida" and "gas-utah" graphs.

attribute	*description*
name	Name of node/link
pressure	Pipeline pressure (psia)
id	ID of node/link
maximumPressure	Maximum pipeline pressure (psia)
slack	Whether the node is slack [true | false]
inService	Whether the node is in service [true | false]
latitude	Latitude of the node
longitude	Longitude of the node
altitude	Altitude of the node
outgoingFlow	Outgoing pipeline flow (CFD)
length	Length of pipeline (miles)
maximumFlow	Maximum pipeline flow (CFD)
efficiency	Pipeline efficiency
incomingFlow	Incoming pipeline flow (CFD)
diameter	Pipeline diameter (inches)

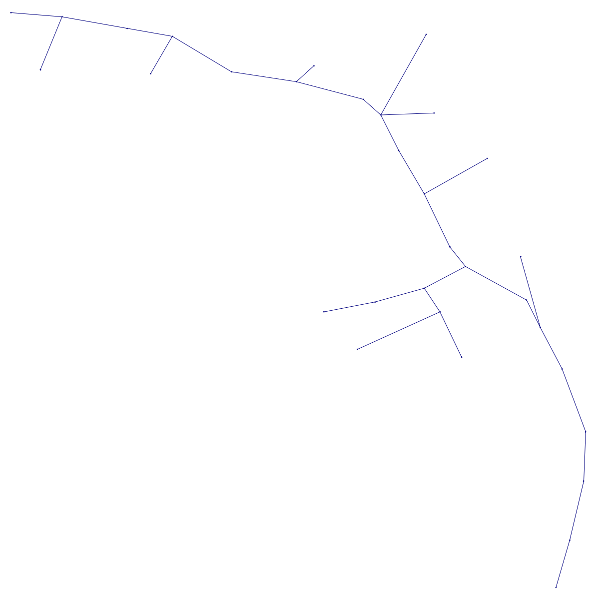

Figure 42. Layout of the "gas-florida" graph.

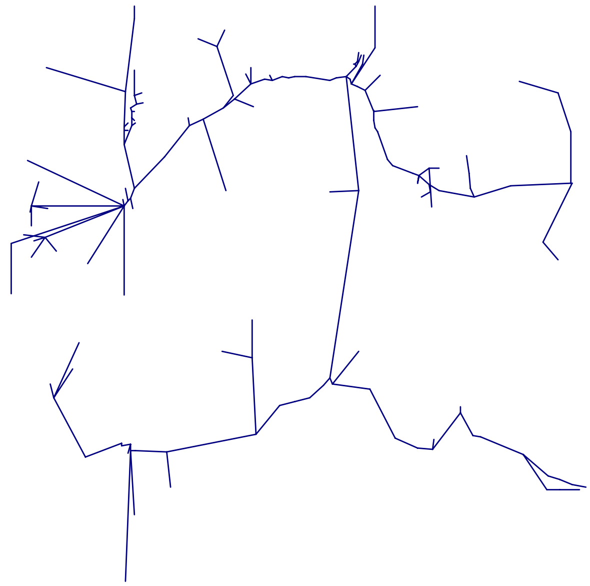

Figure 43. Layout of the "gas-utah" graph.

D. Logistics

Our example logistics graph is the Department of Defense simplified multimodal logistics network [De 92] used in LANL's Force Deployment Estimator analysis tool. It is a directed graph with the attributes listed in Table 22. Figure 44 shows the layout of the graph. This is nearly a complete graph.

Table 22. Attributes of the "fde" graph.

attribute	*description*
LON	The longitude of the node
LAT	The latitude of the node
FLD1	Method of transportation
FLD2	Distance
FLD3	Number of days for round trip
FLD4	Hours not moving during transport

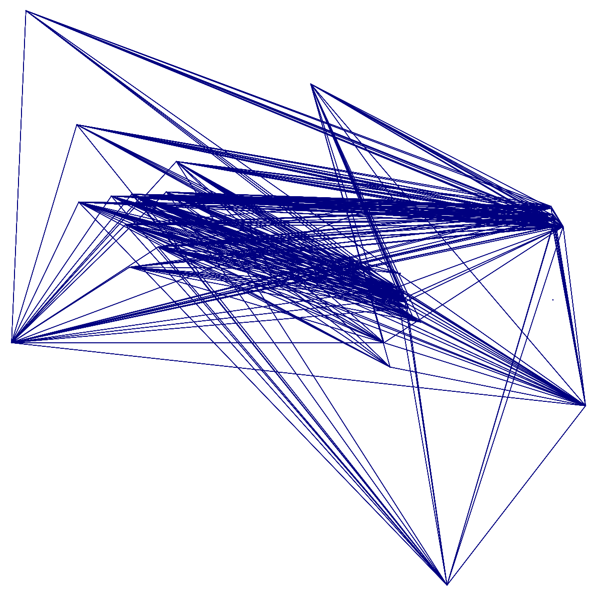

Figure 44. Layout of the "fde" graph.

E. Interdependence

We define an infrastructure interdependency as a physical, logical, or functional connection from one infrastructure to another, the loss or severing of which would affect the operation of the dependent infrastructure. In order to construct a graph representing the interdependencies between various infrastructures, we start with seven layers representing individual infrastructures (see Table 23)—we chose these 17 graphs because they had sufficient quality, completeness, and geographic extent. Next, to estimate how these infrastructures interconnect, we rely on a rule set developed by Jacque Grenier [Gr 00] that assesses the degree of one infrastructure's dependence upon another (see Table 24 for the subset of this matrix relevant to the seven infrastructure layers). If we map this matrix onto our seven layers, we find the dependencies listed in Table 25. To construct the interdependence edges in the combined graph, we identified the vertices in different layers that are within one-half nautical mile of each other and are listed as a high dependency in Table 25—these edges make up 14% of the total number of edges in the graph. Table 26 lists what connections between layers are potentially present for these infrastructures. The resulting graph has the attributes listed in Table 27. Figure 45 illustrates the layout of the graph.

Table 23. Infrastructure graphs used to construct the layers of the "interdependence" graph.

layer	*graph*
electric (1)	electric-ecar
	electric-ercot
	electric-frcc
	electric-maac
	electric-main
	electric-mapp
	electric-nepool
	electric-nypp
	electric-serc
	electric-spp
	electric-wscc
petroleum (2)	petroleum-mapsearch
air (3)	air-tickets
railway (4)	railways-usa
highway (5)	highways-usa
waterway (6)	waterways-usa
telephone (7)	telephone-lerg

Table 24. Matrix of interdependencies for several key infrastructures (extracted from [Gr 00]). The horizontal rows show which infrastructures depend on the infrastructures in the vertical columns; the abbreviations H, M, and L refer to high, medium, and low degrees of dependency.

Industry	*Element*	*Electrical Power*	*Natural Gas*	*Oil Industry*	*Aviation*	*Air Carriers*	*Airport Services*	*Air Navigation Services*	*Rail*	*Surface*	*Trucking*	*Roads Infrastructure*	*Marine*	*Port Services*	*Marine Navigation*	*Telecommunications*
Energy & Utilities	Electrical Power			M					L		L	M				H
	Natural Gas	L		L							L	L	L			M
	Oil Industry	H							L		H	M	M			M
Transportation	Aviation															
	Air Carriers	M	L	H			H	H			M	M				M
	Airport Services	M	M	H		M		H			M	L				M
	Air Navigation Services	L	M	M		M	H				L	L			L	H
	Rail	M	L	H		L					H	L	M	M		M
	Surface															
	Trucking	M	L	H		M	L		M			H	M	M		H
	Roads Infrastructure	M	L	H							M					M
	Marine	L	L	H					M		M	M		H	H	M
	Port Services	H	L	M					H		H	M	H		M	M
	Marine Navigation	M	L	M							L	L	M	L		M
Communications	Telecommunications	H	L	M								M				

Table 25. Dependencies between the seven infrastructure layers.

	dependee						
dependent	electric	petroleum	air	railway	highway	waterway	telephone
electric		M		L	M		**H**
petroleum	**H**			L	**H**	M	M
air	M	**H**			M		**H**
railway	M	**H**	L		**H**	M	M
highway	M	**H**	M	M		M	**H**
waterway	**H**	**H**		**H**	**H**		M
telephone	**H**	M			M		

Table 26. Potential number of connections between infrastructure layers: the numbers in bold represent the number of connections in the "interdependence" graph we have constructed.

	dependee						
dependent	electric	petroleum	air	railway	highway	waterway	telephone
electric	**62405**	3680	71	7855	7898	657	**3493**
petroleum	**4933**	**170818**	57	11038	**9751**	777	3879
air	35	**33**	**48625**	66	102	8	**50**
railway	17799	**19674**	347	**160168**	**44467**	4078	26349
highway	8029	**8099**	252	20715	**124782**	1405	**15155**
waterway	**419**	**368**	6	**965**	**762**	**6870**	292
telephone	**4106**	3519	166	12479	13701	594	**106253**

Table 27. Attributes of the "interdependence" graph.

attribute	***description***
layer	name of the node's layer
graphid	id of the original graph
localid	id of the node within the original graph
x	longitude
y	latitude
z	layer offset
sourcelayer	name of the layer supplying service
targetlayer	name of the layer receiving service

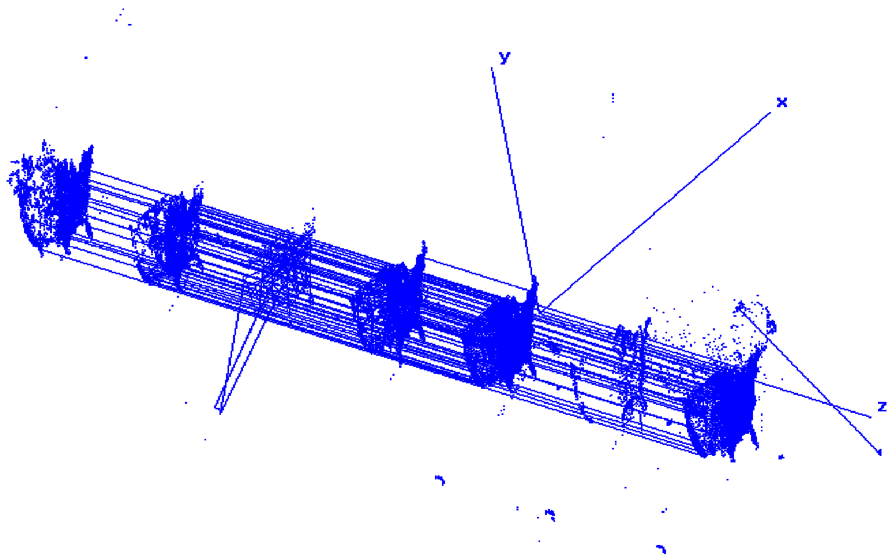

Figure 45. Layout of the "interdependence" graph, with many edges deleted for clarity of display.

III. Methods of Characterization

A. Preprocessing

1. Cleanup

Because the infrastructure graphs come from widely different data sources, there is a lot of heterogeneity in the representation of components. It is also possible for different databases to represent the same infrastructure system with different graphs. Some databases, for example, create vertices so that a geographic layout of the graph follows the geographic extent of the infrastructure: in some sense these are "extra" vertices because no infrastructure component exists at this point—there is just a bend in a pipe or railroad track there. In other cases, databases contain numerous disconnected components that are too small for meaningful analysis: the "microwave-fcc" database contains many small microwave transmission systems, for example. In order to make our analyses more robust, we have preprocessed the infrastructure graphs to "clean up" some of the artifacts present in the source databases.

Our "raw" data set is a simple translation of the source data for the infrastructure into a graph. We have not cleaned up this data at all—it has merely been reformatted.

Our "clean0" data set has the small components in the "raw" graphs removed. This eliminates subgraphs that are too small to be interesting in the analysis. Table 28 shows the cutoff for component size during preprocessing.

Table 28. Size of smallest components retained in graph preprocessing.

graph	*minimum component size*
air-sample	23
air-tickets	643
communications-bge	36
communications-gpc	13
communications-jepco	23
communications-pepco	32
communications-pge	268
communications-pnm	47
communications-sceg	27
communications-sdge	41
communications-sepco	6
communications-sngc	107
communications-tpc	49
communications-vepco	81
electric-ecar	4194
electric-ercot	4718
electric-frcc	2079
electric-maac	2836
electric-main	3424
electric-mapp	4457
electric-nepool	2175
electric-nypp	4117
electric-serc	5526
electric-spp	3711
electric-wscc	12202
fde	40
gas-florida	32
gas-utah	160
highways-canada	953
highways-mexico	518
highways-usa	11
interdependence	357339
internet-bell	121665
internet-nlanr	6474
microwave-fcc	5
nts	11
petroleum-mapsearch	100
railways-canada	654
railways-mexico	177
railways-usa	132432
roads-dallas	9857
roads-portland	100733
telephone-lerg	31768
waterways-usa	6251

Our “clean1” data set has the small components of the “raw” graphs removed along with trivial degree-two vertices. By “trivial,” we mean degree-two vertices that are not part of a cycle of length three. This eliminates vertices that do not contribute topological information to the graph and that tend to lengthen the distance between dependent vertices. Table 29 shows how preprocessing affects the number of vertices, edges, and components in the various infrastructure graphs. Later we will discuss how preprocessing affects the measurements used in our analyses.

Table 29. Effect of preprocessing on the infrastructure graphs.

	raw			*clean0*			*clean1*		
name	***V***	***E***	***n***	***V***	***E***	***n***	***V***	***E***	***n***
air-sample	23	234	1	23	234	1	23	234	1
air-tickets	655	48624	13	643	48624	1	643	48624	1
communications-bge	36	41	1	36	41	1	26	31	1
communications-gpc	401	99	303	65	62	3	42	39	3
communications-jepco	25	25	3	23	25	1	7	9	1
communications-pepco	38	44	7	32	44	1	22	34	1
communications-pge	475	322	207	268	321	1	205	258	1
communications-pnm	50	49	2	47	47	1	34	34	1
communications-sceg	28	26	2	27	26	1	17	16	1
communications-sdge	51	45	11	41	45	1	26	30	1
communications-sepco	7	5	2	6	5	1	2	1	1
communications-sngc	156	111	49	107	110	1	47	50	1
communications-tpc	67	49	18	49	48	1	12	11	1
communications-vepco	195	93	106	81	84	1	60	63	1
electric-ecar	4196	5889	2	4194	5888	1	2658	4352	1
electric-ercot	4724	5669	7	4718	5669	1	2910	3861	1
electric-frcc	2082	2605	4	2079	2605	1	1480	2006	1
electric-maac	2836	3701	1	2836	3701	1	2060	2925	1
electric-main	3443	4261	9	3424	4250	1	2484	3310	1
electric-mapp	4478	5449	12	4457	5439	1	3027	4009	1
electric-nepool	2180	2674	6	2175	2674	1	1760	2259	1
electric-nypp	4117	5245	1	4117	5245	1	2608	3736	1
electric-serc	5531	7201	4	5526	7199	1	3182	4855	1
electric-spp	3730	4483	6	3711	4468	1	2143	2900	1
electric-wscc	12226	14696	24	12202	14695	1	8381	10874	1
fde	43	310	4	40	310	1	38	308	1
gas-florida	32	31	1	32	31	1	21	20	1
gas-utah	160	159	1	160	159	1	107	106	1
highways-canada	5232	4718	537	953	966	1	148	161	1
highways-mexico	518	762	1	518	762	1	445	689	1
highways-usa	90415	124782	62	90246	124659	12	53850	88263	12
interdependence	458755	771265	1	458755	771265	1	357339	669849	1
internet-bell	122622	162133	605	121665	161780	1	83071	123186	1
internet-nlanr	6474	12572	1	6474	12572	1	5486	11584	1
microwave-fcc	42313	10644	32456	9539	10233	91	6394	7088	91
nts	101588	163932	476	100985	163780	11	82361	145156	11
petroleum-mapsearch	160859	170818	3180	142166	154791	64	111436	124061	64
railways-canada	691	843	8	654	813	1	579	738	1
railways-mexico	179	207	2	177	206	1	109	138	1
railways-usa	133752	160168	304	132432	159082	1	106316	132966	1
roads-dallas	9863	14750	4	9857	14747	1	8144	13034	1
roads-portland	100852	126161	23	100733	126059	1	88379	113705	1
telephone-lerg	33854	106253	1434	31768	105301	1	31225	104758	1
waterways-usa	6263	6870	6	6251	6863	1	4709	5321	1

2. Directedness & Weighting

Because there are no vertex or edge weights that would be uniformly appropriate to graphs of such different types of infrastructures, we limit our present analyses to unweighted and undirected graphs. This avoids the problem of making an arbitrary or divergent choice of weights in these graphs. Later work will examine weights on these graphs, however.

B. Fundamental Measurements

We have measured a wide variety of graph-theoretic properties for the vertices, edges, pairs of vertices, and components of the infrastructure graphs. Many of these are properties of interest for any graph, but others focus on attributes that relate to the criticality and robustness of an infrastructure.

Note that when averages below are calculated over quantities that might be undefined for some objects, the undefined values are omitted from the average.

1. Vertices

Degree. The degree of a vertex, $k(v)$, is the number of edges connected to it [CL 96]. This is a simple measure of connectivity in a graph.

Eccentricity. The eccentricity of a vertex is the length of the longest shortest path from the vertex to another vertex [CL 96]: $e(v) = \max_{u \in V(G)} \{\ell(u,v)\}$. We use a breadth-first search to measure this. The eccentricity is a measure of the size and connectedness of a graph. Vertices with low eccentricity are near all the vertices in their component.

Articulation Point. A vertex is an articulation point if and only if removing it from the graph increases the number of components in the graph [CL 96]. We write $a(v) = 1$ or $a(v) = 0$, depending on whether a vertex is an articulation point or not. This is a measure of the fragility of the graph at the vertex. We use a depth-first-search labeling algorithm to find articulation points.

Mean Distance. The mean distance from a vertex is the average length of shortest paths from the vertex to other vertices [Wa 99]: $L(v) = \operatorname*{ave}_{u \in V(G)} \{\ell(u,v)\}$. We use a breadth-first search to measure this. The mean distance is another measure of the extent of a graph. Vertices with low mean distance are typically near the vertices in their component.

Clustering Coefficient. The clustering coefficient for a vertex is the ratio of the number of edges between the vertex's neighbors to the number of edges there would be between the neighbors if they formed a complete subgraph [Wa 99]: $\bar{\gamma}(v) = \left|\{u \mid \ell(u,v)=1\}\right| \Big/ \binom{k(v)}{2}$. This is a measure of how interconnected a vertex's neighbors are. Vertices with low clustering coefficient have few connections between their neighbors; the converse is also true.

Number of Contractions. A pair of vertices that are not connected by an edge and have only one common neighbor is a contraction [Wa 99]. We denote the number of contractions involving a vertex by $\psi(v)$. This is a measure of the number of unique length-two paths between vertices. Vertices with many contractions have many unique connections to nearby vertices.

Shortest Cycle. The length of the shortest cycle at a vertex is written $c(v)$. We measure this using a breadth-first search. This length provides information about the length of alternative paths at a vertex.

2. Edges

Bridge. An edge is a bridge if and only if removing it from the graph increases the number of components in the graph [CL 96]. We write $b(e)=1$ or $b(e)=0$, depending on whether an edge is a bridge or not. This is a measure of the fragility of the graph at the edge. We use a depth-first labeling algorithm to find bridges.

Shortcut. An edge is a shortcut if it does not have a length-two path between its vertices [Wa 99]. We write $\phi(e)=1$ or $\phi(e)=0$, depending on whether the edge is a shortcut or not. This measures the availability of short alternative paths in a graph. A shortcut is not a part of a triangle of edges.

3. Vertex Pairs

Note that we only perform measurements between pairs of vertices on a random sample of pairs for larger graphs due to the amount of computation that would be needed for a complete sample.

Connectivity. We write $s(u,v)=1$ or $s(u,v)=0$, depending on whether two vertices are in the same component of a graph [CL 96]. This is a measurement of disjointedness in the graph.

Minimum Cut. The minimum cut between two vertices, $q(u,v)$, is the minimum number of edges that would have to be deleted from the graph to put the vertices in separate components [Gi 85]. This measures the robustness of the graph. Small minimum cuts are an indication that the graph can be easily separated into additional components.

Contraction. A pair of vertices that are not connected by an edge and have only one common neighbor is a contraction [Wa 99]. We write $\psi(u,v)=1$ or $\psi(u,v)=0$, depending on whether a pair is a contraction. This is a measure of the number of unique length-two paths between vertices. Vertices with many contractions have many unique connections to nearby vertices.

Distance. The distance between two vertices, $\ell(u,v)$, is the number of edges in the shortest path between them [Gi 85]. We use a breadth-first search algorithm for measuring this. The distance is an indication of the extent of a graph.

4. Graphs and Components

Order. The order of a graph, $|V(G)|$, is the number of vertices it contains [CL 96]. This is the simplest measure of how large a graph is. We use this to scale some of the other extensive properties of the graph.

Size. The size of a graph, $|E(G)|$, is the number of edges it contains [CL 96]. This is another measure of how large a graph is. It could also be used to scale other measured quantities.

Components. The number of components of a graph, $n(G)$, is the number of disconnected subgraphs in the graph [CL 96]. A depth-first search algorithm is used to measure this. This helps characterize the disconnectedness in the graph. Graphs with a large number of components are very disconnected.

Minimum Degree. The minimum degree of a graph is written $\delta(G) = \min_{v \in V(G)} \{k(v)\}$ [CL 96]. This helps characterize the degree distribution of vertices in a graph by quantifying how connected the least connected of a graph's vertices are.

Maximum Degree. The maximum degree of a graph is written $\Delta(G) = \max_{v \in V(G)} \{k(v)\}$ [CL 96]. This helps characterize the degree distribution of vertices in a graph by quantifying how connected the most connected of a graph's vertices are.

Radius. The radius of a graph is the minimum eccentricity of the vertices in the graph [CL 96]: $r(G) = \min_{v \in V(G)} \{e(v)\}$. This helps characterize the eccentricity distribution of vertices in a graph by quantifying how eccentric the least eccentric vertices in a graph are. Graphs with small radius have some vertices very close to all other vertices; conversely, graphs with large radius have no vertices very close to all other vertices.

Diameter. The diameter of a graph is the maximum eccentricity of the vertices in the graph [CL 96]: $d(G) = \max_{v \in V(G)} \{e(v)\}$. This helps characterize the eccentricity distribution of vertices in a graph by quantifying how eccentric the most eccentric vertices in a graph are. Graphs with small diameter have all vertices very close to all other vertices; conversely, graphs with large diameter have all vertices very far from some other vertices.

Girth. The girth of a graph is the minimum cycle length in the graph [CL 96]: $g(G) = \min_{v \in V(G)} \{c(v)\}$. This is not defined for acyclic graphs. The girth is a measure of the length of alternative paths between vertices. Graphs with no girth are forests; graphs with large girth have at most one short path between any pair of vertices.

Vertex Connectivity. The vertex connectivity, $\kappa(G)$, is the minimum number of vertices that would have to be removed from the graph to increase the number of components in the graph [CL 96]. We use an articulation-point-finding algorithm to compute this, but we do not compute it for all graphs. Vertex connectivity gives us an overall assessment of the connectedness of the graph. Graphs with low vertex connectivity are easily split apart by deleting vertices.

Edge Connectivity. The edge connectivity, $\kappa_1(G)$, is the minimum number of edges that would have to be removed from the graph to increase the number of components in the graph [CL 96]. We use a bridge-finding algorithm to compute this, but we do not compute it for all graphs. Edge connectivity gives us an overall assessment of the connectedness of the graph. Graphs with low edge connectivity are easily split apart by deleting edges.

Vertex Chromatic Number. The chromatic number of a graph, $\chi(G)$, is the minimum integer k for which the graph is k-colorable [CL 96]. We use a greedy heuristic to estimate an upper bound for this. The chromatic number measures the mutual connectedness of vertices in a graph. Graphs with low chromatic number do not have complete subgraphs of high order. This is related to the size of cliques in the graph.

Vertex Domination Number. The dominating number of a graph, $\gamma(G)$, is the minimum cardinality among the dominating sets of the graph [CL 96]. (A dominating set is a set for which every vertex in the graph is adjacent to or in the set [CL 96].) We use a greedy heuristic to estimate an upper bound for this. The domination number measures the overall adjacency of vertices in a graph. Graphs with low domination number have a few vertices that are adjacent to the rest of the vertices; the converse is also true.

Vertex Independence Number. The independence number of a graph, $\beta(G)$, is the maximum cardinality among the independent sets of vertices of the graph [CL 96]. (An independent set is a set where no two vertices are adjacent [CL 96].) We use a greedy heuristic to estimate a lower bound for this. The independence number also measures the overall adjacency of vertices in a graph. Graphs with low independence number have many vertices that are adjacent to the rest of the vertices; the converse is also true.

Lower Vertex Independence Number. The lower independence number of a graph, $i(G)$, is the minimum cardinality of a maximal independent set of vertices in the graph [CL 96]. We use a greedy heuristic to estimate an upper bound for this. The lower independence number also measures the overall adjacency of vertices in a graph. Graphs with low lower independence number have a few vertices that are adjacent to the rest of the vertices; the converse is also true.

Vertex Covering Number. The covering number of a graph, $\alpha(G)$, is the minimum cardinality of a vertex cover of the graph [CL 96]. (A vertex cover is a set of vertices that is connected to every edge in the graph [CL 96].) We use a greedy heuristic to estimate an upper bound for this. The covering number also measures the overall adjacency of vertices in a graph. Graphs with low covering number have a few vertices that are adjacent to the rest of the vertices; the converse is also true.

Fraction of Articulation Points. The fraction of the vertices that are articulation points is written $\bar{a}(G) = \underset{v \in V(G)}{\text{ave}} \{a(v)\}$. This indicates how common articulation points are. Graphs with a high fraction have many vertices which, if any is deleted, will break the graph into additional disconnected subgraphs.

Fraction of Bridges. The fraction of the edges that are bridges is written $\bar{b}(G) = \underset{e \in E(G)}{\text{ave}} \{b(e)\}$. This indicates how common bridges are. Graphs with a high fraction have many edges which, if any is deleted, will break the graph into additional disconnected subgraphs.

Average Shortest Cycle. The average length of the shortest cycle at vertices in a graph is written $\bar{c}(G) = \underset{v \in V(G)}{\text{ave}} \{c(v)\}$. Note that $c(v)$ is not defined if the vertex does not have a cycle, so these do not contribute to the average. This average measures the typical size of alternative paths between vertices. Graphs where this average is small generally have short alternate paths between vertices if an alternate path exists.

Average Minimum Cut. The average size of the minimum cut between vertices in a graph is written $\bar{q}(G) = \underset{v, w \in V(G)}{\text{ave}} \{q(v, w)\}$. This measures the typical size of minimum cuts. Graphs where this average is small typically have vertices that can be separated by deleting a few edges.

Average Degree. The average degree of a graph is written $\bar{k}(G) = \underset{v \in V(G)}{\text{ave}} \{k(v)\}$ [Wa 99]. This measures the typical connectedness of vertices. Graphs where this is small typically have vertices connected only to a few others.

Characteristic Path Length. The characteristic path length in a graph is the median of the mean path lengths from the vertices in the graph [Wa 99]: $\bar{L}(G) = \underset{v \in V(G)}{\text{med}} \{L(v)\}$. Note that since $L(v)$ is not defined for vertices of degree zero, these do not contribute to the median. This median measures a typical distance between vertices in a graph. Graphs where this is low have vertices that are generally close to one another; the converse is also true.

Clustering Coefficient. The clustering coefficient for a graph is the average of the clustering coefficients of the vertices [Wa 99]: $\bar{\gamma}(G) = \underset{v \in V(G)}{\text{ave}} \{\bar{\gamma}(v)\}$. Note that since $\bar{\gamma}(v)$ is not defined for vertices with degree less than two, these do not contribute to the average. This average measures the typical interconnectedness between a vertex and its neighbors. Graphs where this average is low have vertices that generally do not have connections between their neighbors; the converse is also true.

Shortcut Fraction. The shortcut fraction for a graph is the fraction of edges in the graph that are shortcuts [Wa 99]: $\bar{\phi}(G) = \underset{e \in E(G)}{\text{ave}} \{\phi(e)\}$. This is another measurement of the typical redundancy of paths in a graph. Graphs with low shortcut fraction typically have alternate paths of length two between the vertices connected to an edge; the converse is also true.

Contraction Fraction. The contraction fraction for a graph is the fraction of vertex pairs that are contractions [Wa 99]: $\overline{\psi}(G) = \underset{v,w \in V(G)}{ave}\{\psi(v,w)\}$. This measures likelihood of two unconnected vertices having a unique common neighbor. Graphs with low contraction fraction have few unconnected vertex pairs with unique common neighbors; the converse is also true.

Modified Contraction Fraction. Our calculation of $\overline{\psi}$ strictly follows the definition in Reference [Wa 99]. It appears that the calculations in Reference [Wa 99] differ from their stated definition, however. In order to compare with those results, we have also defined a modified contraction fraction for a graph as the fraction of vertex pairs having overlapping neighborhoods that are contractions: $\psi^*(G) = \underset{\ell(v,w) \le 2}{ave}\{\psi(v,w)\}$. This modification eliminates the consideration of distant pairs in the average.

C. Derived Measurements

1. Normalization

In order to make the extensive quantities measured in graphs more useful for analysis, we have developed a normalization scheme to convert these to intensive quantities. We use a simple scaling in terms of the order of the graph or its square root. We use the order for cardinality-related properties and its square root for distance-related quantities. These simple approaches appear to adequately normalize both types of quantities.

Radius. We write the normalized radius as $\hat{r} = r / \sqrt{V}$. Figure 46 shows that this normalization process removes the order dependence from the measurement.

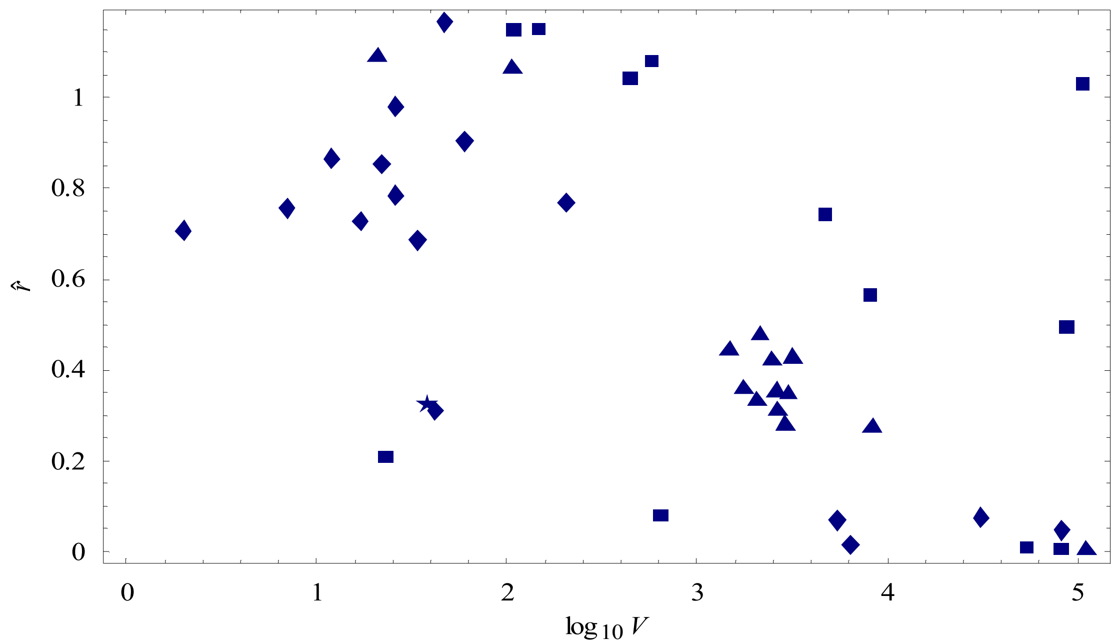

Figure 46. Dependence of normalized radius on the order of the graph for the "clean1" data set. The boxes represent transportation networks, the diamonds communications networks, the triangles energy networks, and the stars logistics networks.

Diameter. We write the normalized diameter as $\hat{d} = d / \sqrt{V}$. Figure 47 shows that this normalization process removes the order dependence from the measurement.

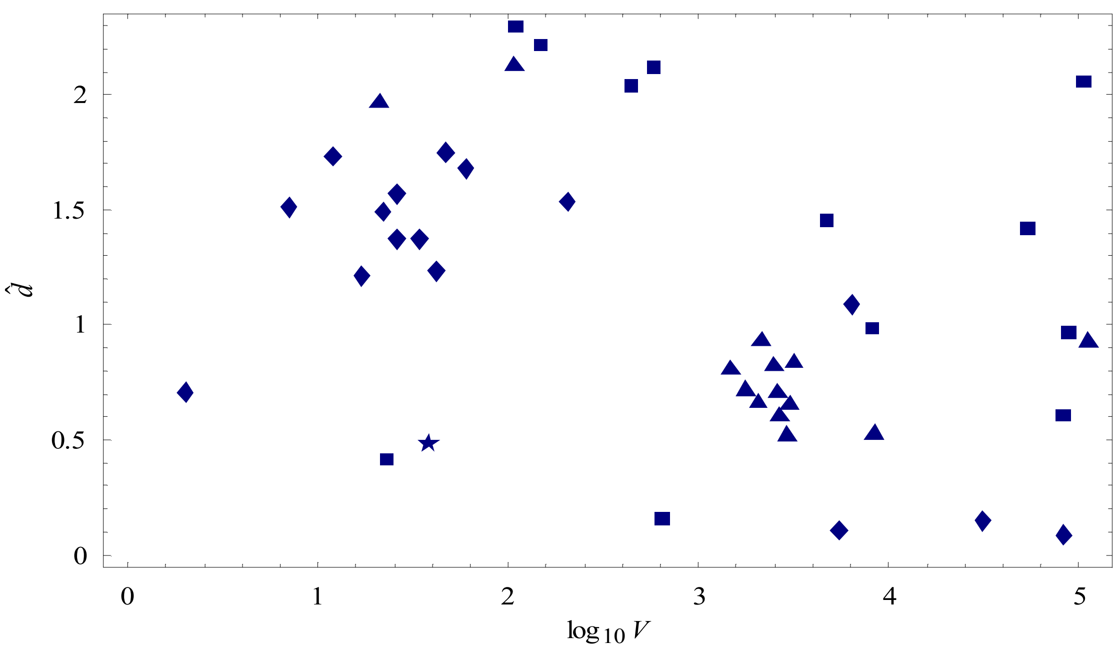

Figure 47. Dependence of normalized diameter on the order of the graph for the "clean1" data set. The boxes represent transportation networks, the diamonds communications networks, the triangles energy networks, and the stars logistics networks.

Vertex Domination Number. We write the normalized domination number as $\hat{\gamma} = \gamma / V$. This normalized quantity always falls between zero and one. Figure 48 shows that this normalization process removes the order dependence from the measurement.

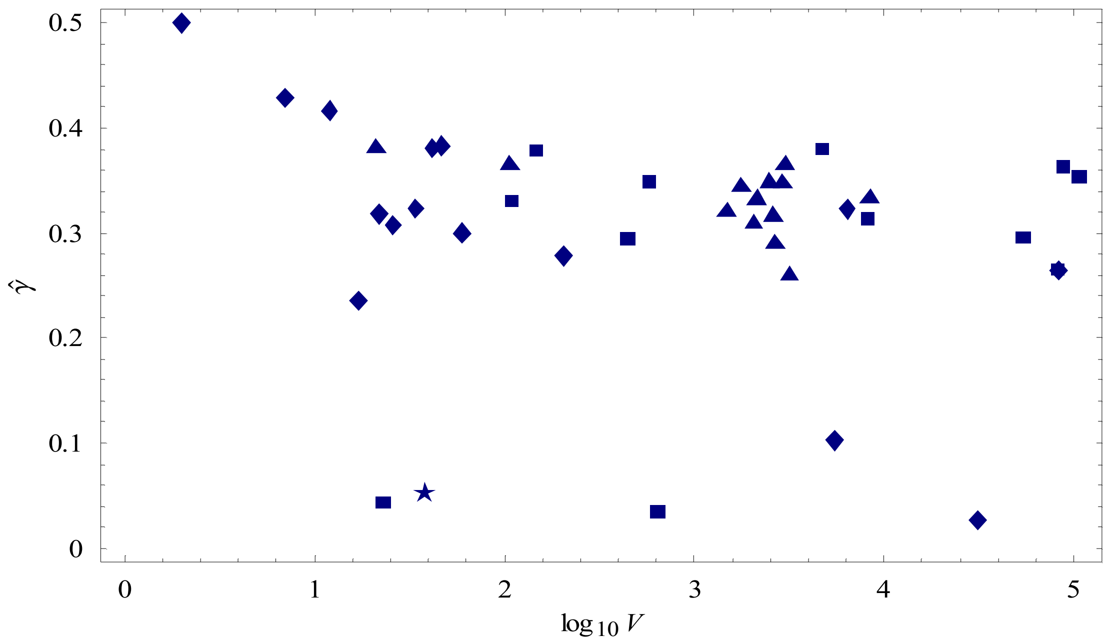

Figure 48. Dependence of normalized vertex domination number on the order of the graph for the "clean1" data set. The boxes represent transportation networks, the diamonds communications networks, the triangles energy networks, and the stars logistics networks.

Vertex Independence Number. We write the normalized independence number as $\hat{\beta} = \beta / V$. This normalized quantity always falls between zero and one. Figure 49 shows that this normalization process removes the order dependence from the measurement.

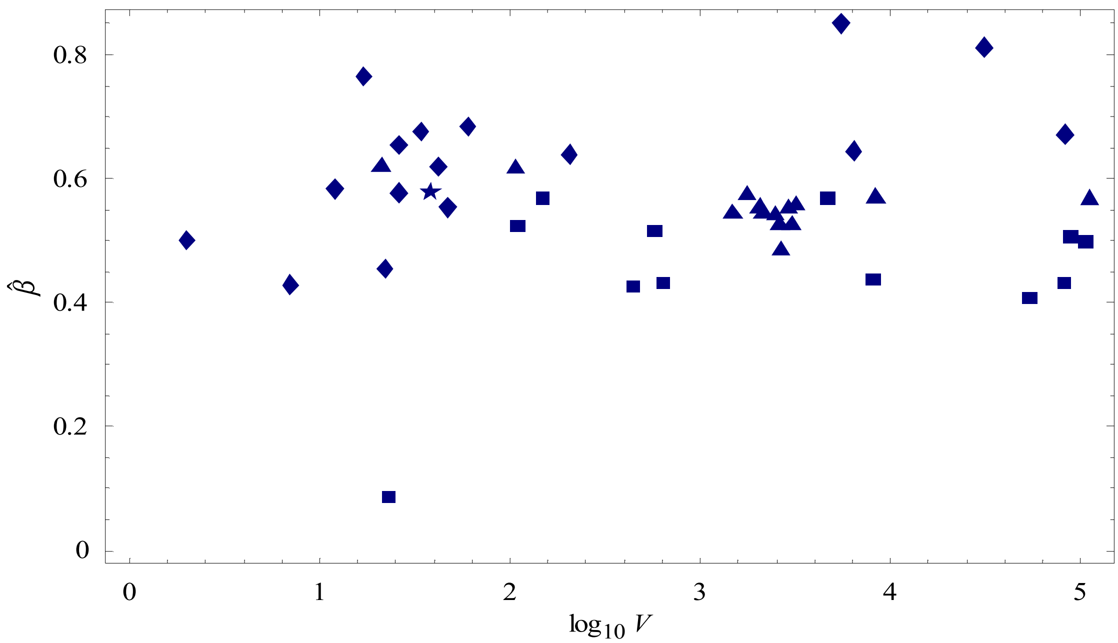

Figure 49. Dependence of normalized vertex independence number on the order of the graph for the "clean1" data set. The boxes represent transportation networks, the diamonds communications networks, the triangles energy networks, and the stars logistics networks.

Lower Vertex Independence Number. We write the normalized lower independence number as $\hat{i} = i / V$. This normalized quantity always falls between zero and one. Figure 50 shows that this normalization process removes the order dependence from the measurement.

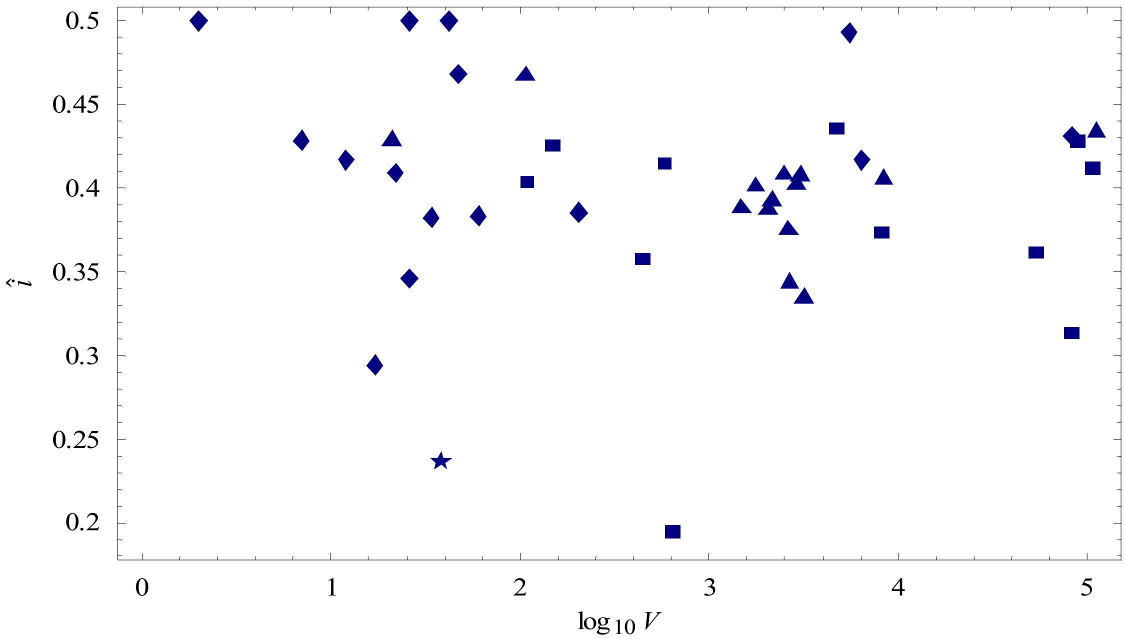

Figure 50. Dependence of normalized lower vertex independence number on the order of the graph for the "clean1" data set. The boxes represent transportation networks, the diamonds communications networks, the triangles energy networks, and the stars logistics networks.

Vertex Covering Number. We write the normalized covering number as $\hat{\alpha} = \alpha / V$. This normalized quantity always falls between zero and one. Figure 51 shows that this normalization process removes the order dependence from the measurement.

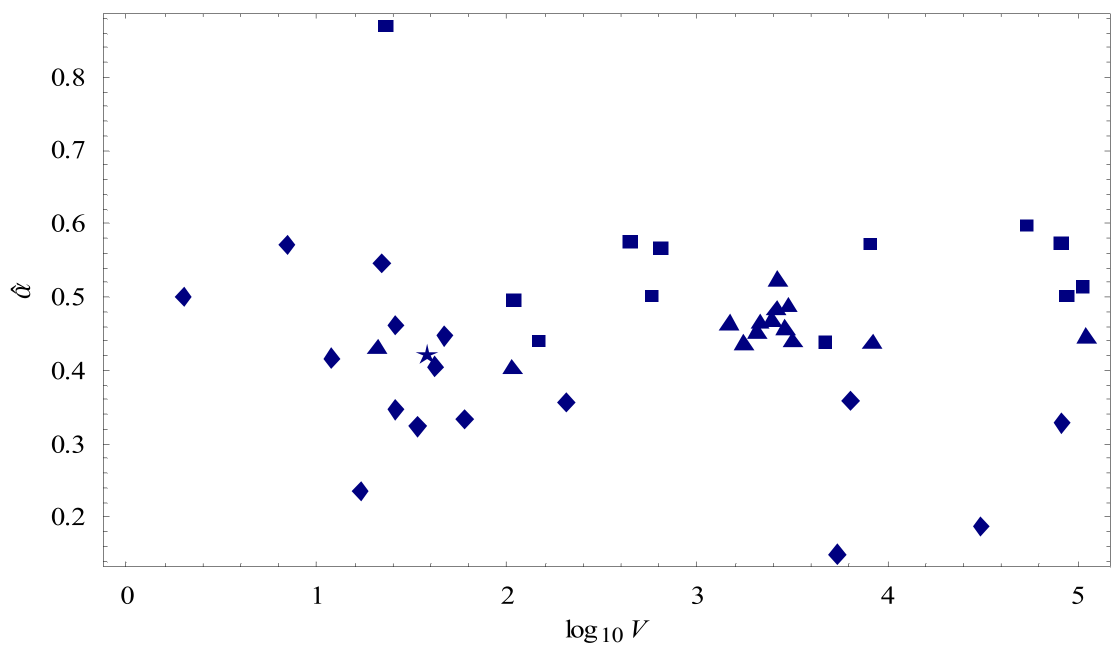

Figure 51. Dependence or normalized vertex covering number on the order of the graph for the "clean1" data set. The boxes represent transportation networks, the diamonds communications networks, the triangles energy networks, and the stars logistics networks.

Characteristic Path Length. We write the normalized characteristic path length as $\hat{\bar{L}} = \bar{L} / \sqrt{V}$. Figure 52 shows that this normalization process removes the order dependence from the measurement.

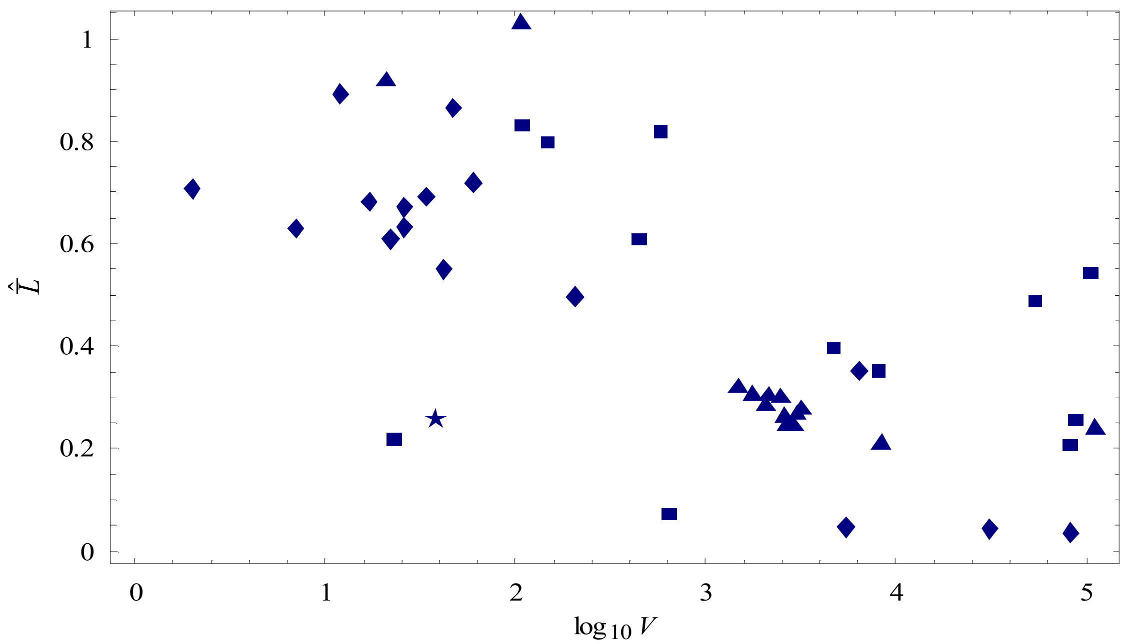

Figure 52. Dependence of normalized characteristic path length on the order of the graph for the "clean1" data set. The boxes represent transportation networks, the diamonds communications networks, the triangles energy networks, and the stars logistics networks.

Eccentricity. We write the normalized eccentricity as $\hat{e} = e / \sqrt{V}$. This has the same normalization characteristics as radius and diameter.

Mean Distance. We write the normalized mean distance as $\hat{L} = L / \sqrt{V}$. This has the same normalization characteristics as characteristic path length.

2. Entropies

We calculate normalized entropies from the empirical distributions of distance (ℓ), shortest cycle length (c), minimum cut (q), and degree (k) in the graphs. We denote their normalized entropies by $\tilde{\ell}$, $\tilde{c}$, $\tilde{q}$, and $\tilde{k}$, respectively. The unnormalized entropies are calculated as $S = -\sum_i f_i \log f_i$, where f_i is the fraction of the observations falling in bin i. We normalize this as $\tilde{S} = S / \log N$, where N is the total number of observations. This normalized quantity always falls between zero and one.

3. Inequalities

We have also investigated using graph-theoretic bounds for various to estimate quantities that are not easily measured. It turns out that in most cases these bounds are too broad to be interesting or useful. Below we collect the inequalities we investigated, using the notation of the original sources.

$\operatorname{rad} G \le \operatorname{diam} G \le 2 \operatorname{rad} G$ [CL 96]

$a(G) \le 1 + \left\lfloor \frac{\max \delta(G')}{2} \right\rfloor$, where G' is an induced subgraph of G [CL 96]

$a(G) \le 1 + \left\lfloor \frac{\Delta(G)}{2} \right\rfloor$ [CL 96]

$a'(G) \ge \left\lceil \frac{1+\max \delta(G')}{2} \right\rceil$, where G' is an induced subgraph of G [CL 96]

$a(G) \le a'(G)$ [CL 96]

$a'(G) = \max \left\lceil \frac{|E(H)|}{|V(H)|-1} \right\rceil$, where H is an nontrivial induced subgraph of G [CL 96]

$\kappa(G) \le \kappa'(G) \le \delta(G)$ [CL 96]

$\frac{\kappa(G)}{\beta(G)} \le t(G) \le \frac{\kappa(G)}{2}$ [CL 96]

$I(G) \ge \min_{1 \le t \le n} \{\max\{t, d_t + 1\}\}$, where $d_1 \ge d_2 \ge \ldots \ge d_n$ is the degree sequence [CL 96]

$I(G) \ge \left\lceil \frac{n-\kappa(G)}{\beta(G)} \right\rceil + \kappa(G)$ [CL 96]

$I(G) \ge \left\lceil 2\sqrt{n\, t(G)} - t(G) \right\rceil$, if $G \ne K_n$ [CL 96]

$I(G) \ge 1 + \delta(G)$ [CL 96]

$\operatorname{gen}(G) \le \operatorname{gen}_M(G)$ [CL 96]

$\operatorname{gen}_M(G) \le \left\lfloor \frac{m-n+1}{2} \right\rfloor$ [CL 96]

$\chi(G) \le \Delta(G)$, if G is neither an odd cycle nor a complete graph [CL 96]

$\chi(G) \le 1 + \Delta(G)$ [CL 96]

$\chi(G) \le \max_{1 \le i \le n} \{\min\{i, \deg v_i + 1\}\}$ [CL 96]

$\chi(G) \le 1 + \ell(G)$, where $\ell(G)$ denotes the length of the longest path in G [CL 96]

$\chi(G) \le 1 + \max \delta(H)$, where H is an induced subgraph of H [CL 96]

$2\sqrt{n} \le \chi(G) + \chi(\overline{G}) \le n + 1$ [CL 96]

$n \le \chi(G)\,\chi(\overline{G}) \le \sqrt{\frac{n+1}{2}}$ [CL 96]

$\chi'(G) = \chi(L(G))$ [CL 96]

$\chi'(G) \le 1 + \Delta(G)$ [CL 96]

$\chi'(G) = \Delta(G)$ if the vertices of maximum degree are independent [CL 96]

$\alpha(G) + \beta(G) = n$ if there are no isolated vertices [CL 96]

$\alpha'(G) + \beta'(G) = n$ if there are no isolated vertices [CL 96]

$\left\lceil \frac{n}{1+\Delta(G)} \right\rceil \le \beta'(G) \le \left\lfloor \frac{n}{2} \right\rfloor$ if there are no isolated vertices [CL 96]

$\left\lceil \frac{n}{2} \right\rceil \le \alpha'(G) \le \left\lfloor \frac{n\,\Delta(G)}{1+\Delta(G)} \right\rfloor$ if there are no isolated vertices [CL 96]

$i'(G) \le \beta'(G) \le 2\,i'(G)$ [CL 96]

$\mathrm{ban}(G) \ge \left\lceil \frac{\Delta(G)}{2} \right\rceil$ [CL 96]

$\mathrm{ban}(G) \ge \chi(G) - 1$ [CL 96]

$\mathrm{ban}(G) \ge \kappa(G)$ [CL 96]

$\left\lceil \frac{n}{\beta(G)} \right\rceil - 1 \le \mathrm{ban}(G) \le n - \left\lfloor \frac{\beta(G)}{2} \right\rfloor - 1$ [CL 96]

$\mathrm{ban}(G) \le n - \mathrm{diam}\, G$ [CL 96]

$\gamma(G) \le \frac{n}{2}$ [CL 96]

$\gamma(G) \le \frac{n(1+\ln(\delta(G)+1))}{\delta(G)}$ if $\delta(G) \ge 2$ [CL 96]

$\left\lceil \frac{n}{1+\Delta(G)} \right\rceil \le \gamma(G) \le n - \Delta(G)$ [CL 96]

$\gamma(G) \le n - \kappa(G)$ [CL 96]

$\gamma(G) \le \min\{\alpha(G),\ \alpha'(G),\ \beta(G),\ \beta'(G)\}$ if there are no isolated vertices [CL 96]

$n - m \le \gamma(G) \le n + 1 - \sqrt{1 + 2m}$ [CL 96]

$3 \leq \gamma(G) + \gamma(\overline{G}) \leq N + 1$ [CL 96]

$2 \leq \gamma(G)\,\gamma(\overline{G}) \leq n$ for $n \geq 2$ [CL 96]

$\gamma(G) + \gamma(\overline{G}) \leq \frac{n+4}{2}$ for $n \geq 2$ if there are no isolated vertices [CL 96]

$\gamma(L(G)) = i(L(G))$ [CL 96]

$\gamma(G) + i(G) \leq n$ if there are no isolated vertices [CL 96]

$i(G) \leq n + 2 - 2\sqrt{n}$ for $n \geq 2$ if G is connected [CL 96]

$\mathrm{ir}(G) \leq \gamma(G) \leq i(G)$ [CL 96]

$\mathrm{ir}(G) \geq \frac{2n}{3\,\Delta(G)}$ if $\Delta(G) \geq 2$ [CL 96]

$\gamma(G) \leq 2\,\mathrm{ir}(G) - 1$ [CL 96]

$\mathrm{ir}(G) \leq \gamma(G) \leq i(G) \leq \beta(G) \leq \Gamma(G) \leq \mathrm{IR}(G)$ [CL 96]

$\theta'(G) \leq \left\lfloor \frac{\sqrt{2m}}{3} + \frac{3}{2} \right\rfloor$ [Ci 95]

$\beta(G) \geq \sum_i \frac{1}{1+d_i}$ [Ha 98]

$\theta'(G) \geq \left\lceil \frac{m}{3n-6} \right\rceil$ for $n \geq 2$ [MOS 98]

$\theta'(G) \leq \left\lfloor \frac{n+7}{6} \right\rfloor$ for $n > 10$ [MOS 98]

$\theta'(G) \leq \left\lceil \frac{\Delta(G)}{2} \right\rceil$ for $n > 10$ [MOS 98]

$I(G) = \min\{N(G),\ 1 + \min_v I(G-v)\}$ for nontrivial graphs [BBG 92]

$I(G) \leq \alpha(G) + 1$ [BBG 92]

$I(G) \geq \chi(G)$ [BBG 92]

$I(G) + I(\overline{G}) \geq n + 1$ [BBG 92]

$I(G)\,I(\overline{G}) \geq n$ [BBG 92]

$2 \le I(G) \le I'(G) \le n$ [BBL 94]

$I'(G) \ge \lceil 2\sqrt{n} \rceil - 1$ [BBL 94]

$I'(G) \ge \Delta(G) + 1$ [BBL 94]

$I'(G) \ge \min\{\lceil \sqrt{2\,n\,k'(G)} \rceil, n\}$ [BBL 94]

$n + 1 \le I'(G) + I'(\overline{G}) \le 2\,n$ [BBL 94]

$n \le I'(G)\,I'(\overline{G}) \le n^2$ [BBL 94]

D. Algorithms

In the following sections we discuss the algorithms used to measure the infrastructure graphs. Strictly speaking, the descriptions refer to our Java programming language implementation, not our C++ implementation, but the methods are similar in both. The Appendix provides more details on our software.

1. Articulation Points

An articulation node is one such that if it were removed, it would disconnect the graph. The key point behind this algorithm is that if a node is an articulation node, there is no alternative path from one of its adjacencies to another.

The algorithm picks a node that has not yet been looked at (one per component of the graph). It then does a depth-first search on the graph. All adjacent nodes are put on the tree. Each node is given a visit order, related to its depth. If a node is already on the depth-first-search tree, it is given the lesser of its visit orders, and its parent is reset to the current parent. When we get to the end of a path, either because there are no more children, or because all of the children have already been put on the tree, we check the current node to see if it is an articulation node. A node is an articulation node if its own visitation order is less than or equal to the minimum child visitation order of its child in the path. If the parent is not an articulation node itself, its minimum child visitation value is set to that of the smallest visitation order of child nodes. The algorithm then backs out of the tree, checking each node to see if it is an articulation node.

A special case occurs when checking the tree root to be an articulation node. The root node is an articulation node if and only if when a child node search is completely searched, there is still exists one or more child node(s) of the root node left unvisited.

2. Breadth-First Search

A breadth-first search finds eccentricity (maximum path length over all minimum paths to other nodes), and mean path length (average minimum-path length to all other nodes) for each node. From this one computes the radius (minimum eccentricity) and diameter (maximum eccentricity) for the graph.

The algorithm works on equally weighted graphs only. To begin, the node in question is selected, and is given distance 0. All of its adjacencies are then selected, and given distance 1. Next, all of their adjacencies are given distance 2, unless they have already been given a distance. This continues until all adjacent nodes have been given a distance. Note, not all of the nodes will be given a distance in a disconnected graph. The eccentricity is then the last distance assigned, and the mean path length can be computed. Once this has been done on all nodes in the graph, the radius and diameter are assigned.

3. Bridges

This algorithm finds edges such that if removed, it would disconnect the graph.

As input, this algorithm uses the set of articulation nodes. The heuristic relies on the categorization of bridges, and looks at each independently. Those categories are bridges that are part of pendants, fray strings, and inner-bridges. Note that every bridge connects two articulation nodes, two pendants, or an articulation node and a pendant.

A pendant node is a node that is of degree one: i.e., it is a radial node. Its sole adjacent edge is a bridge, and its sole adjacent node either is an articulation node, or is also pendant. Binary pendant bridges are found by searching all degree one nodes and looking at their adjacent node. If the adjacent node also has degree one, then this is a binary pendant subgraph, and the edge connecting them is a bridge. If the adjacent node has degree greater than one, then the edge connecting them is a bridge, and it is an ordinary pendant.

A fray string is a subset of a graph composed of zero or more connected degree two articulation nodes, with at least one pendant (degree 1) node. (A pendant node attached to a node of degree greater than 2 is a degenerate fray string. These are found above, but are still categorized here.) Fray strings are found by starting with a degree-two articulation point and searching out both directions until a pendant node is found on one end and either a pendant node, or node with degree greater than two is found on the other end.

The inner-bridges nodes are articulation points that are connected, do not fit the criteria above, and have only one path from end to end. These are found by checking the adjacencies of each articulation node to find articulation nodes. Once a connected pair of articulation nodes is found, the path-finder algorithm is run to see how many paths exist between those nodes. If only one path exists, the edge is a bridge.

All bridges, except for pendant nodes, can be found using the last method mentioned here. However, it is much more efficient to find fray strings without having to run a path-finder algorithm. Also, we believe that insights to the graph can be gained by knowing the types and number of bridges it has.

4. Clustering Coefficient

The clustering coefficient algorithm examines the neighborhood of each node to determine how complete the graph of its neighbors is. The measure is one if every node adjacent to node v has an edge to every other node adjacent to v, and zero if there are no edges between the adjacent nodes.

The algorithm looks at one node at a time, finding the minimum cycle for each node. A list of its adjacencies is created. Then, the list is examined as such. The adjacent nodes of each node on the adjacent list are checked to see if they are on the adjacent list: i.e., we count how many adjacent nodes have adjacencies that are adjacent to the original node. The count of these adjacent edges is then divided by the number of possible adjacent edges, $\binom{k}{2}$, where k is the degree of the node.

5. Contraction Fraction

A contraction is a pair of non-adjacent nodes sharing exactly one neighbor. In this case, the one common neighbor could be removed and the pair would be connected directly without affecting the topology of the graph. Thus the contraction fraction is the number of contractions found in the graph divided by the number of pairs of nodes one could choose—$\binom{n}{2}$, where n is the order of the graph.

The algorithm starts by selecting a focus node. A list of its adjacencies is constructed, then examined. The adjacencies of the nodes on the adjacency list are then dealt with as follows: If the node is not an adjacency itself, it is not on the list of possibilities, and has not been already removed from the list of possibilities, then add it to the list of possibilities. If the node is already on the list, remove it, and note that it has been removed. Once all of the adjacencies of nodes on the adjacency list have been examined, we are left with nodes on the list of possibilities that have exactly one common neighbor with the focus node.

6. Cycles

This algorithm finds the minimum cycle length beginning and ending at each node of a graph.

The algorithm works by creating a stack of nodes ordered by their distance from the focus node, the node in question. Nodes are put on this stack as they are encountered in a breadth-first search. Each node on the stack knows its distance and which branch it came from. If nodes from different branches are adjacent, a cycle about the focus node exists. We make sure this is the minimum cycle by continuing to search for cycles until the accumulated distance to each node on the stack is more than half the length of the cycle found. In an equally weighted graph, the first cycle found will be the minimum.

Proof :

That a cycle about the node is acquired is obvious; a connection between branches always creates a cycle about the root. Thus the only thing left to prove is the termination criteria for the algorithm, both for unweighted and weighted graphs.

The assertion is made that a minimum cycle will be found in the first level of the search that produces a cycle for an unweighted graph. When a cycle is found, its length is computed as the sum of the distances to the two connected nodes (from different branches) and one, for the weight of the edge connecting them. The connected nodes will either be from the same level, or one from the current level, and one from the previous. Thus, if a cycle is found at level j, then it must have length $2j$ or $2j + 1$. Hence, there can be no smaller cycle at any level greater than j, and the algorithm can stop searching for a smaller cycle.

For a weighted graph, a depth-first search is used. Again, the length of the shortest cycle will be the sum of the distances to the connected nodes (from different branches) and the weight of the edge that connects them. Let the length of the minimum cycle found so far be c. If the smallest distance on the stack is at least ½ c then at best, the length of any remaining cycles is $\frac{1}{2} c + \frac{1}{2} c + e$, where e is the weight of the edge connecting them. Since we allow positive weights on edges only, any remaining cycles will have length at least c, and the search can be stopped. Thus for a weighted graph, we can stop searching for a smaller cycle once the smallest distance to a node left on the stack is at least ½ of the length of the smallest cycle found thus far.

7. Paths

This is a maximum flow/minimum cut algorithm. A general depth-first path-finding procedure which continues to find paths from a source node (*S*) to a target node (*T*) as long as there exists a path where all of the edges in the path edges have residual capacity for flow in the direction defining the path. By default, each edge is initialized with capacity equal to the weight of the path (assuming positive weights). The path can be traversed from *T* to *S* by following the parent trail backwards. When a path is found from *S* to *T*, the residual flow for each edge is reduced by the minimum residual on the path. In the case that the initial residual is the same for each node (in an unweighted graph), the results of this program will be the maximum number of unique paths from *S* to *T* that do not repeat an edge. The number of such paths will be no greater than the lesser of the degree of *S* or *T*.

8. Shortcuts

A given edge between two nodes is a shortcut if there is no path that is shorter than two edges between the pair of nodes other than the given edge. The shortcut fraction is the number of shortcuts in the graph divided by the number of edges.

The algorithm looks to see if each edge is part of a triangle. Essentially, if a line is part of a triangle, it is not a shortcut. To accomplish this, we look at each node, one at a time. From the focus node, we gain a list of adjacencies. Then we look at the adjacencies of the adjacencies. If the second order adjacencies are also adjacencies, then we have found a triangle. This will rule out the possibility of the edges from the focus node to the adjacency and from the focus node to the second order adjacency that was also an adjacency to the focus node. The third edge of the triangle could also be ruled out here, but is left to be dealt with when one of the adjacencies is the focus node due to bookkeeping issues. Otherwise, the edge is a short cut, and is counted as such.

E. Heuristics

In addition to the algorithms outlined above, we have developed several heuristics for estimating and bounding quantities that cannot be practically calculated exactly in large graphs. These are discussed below. We have also considered using genetic algorithms in these cases, but have not implemented them.

1. Domination Set

The heuristic finds the minimum set such that each node is either in the set, or adjacent to a node in the set. Finding the actual minimum domination set is a NP-hard problem, so instead the algorithm finds several minimal domination sets, the smallest of which is used as the upper bound for the minimum domination set. If a node is adjacent to, or is in the domination set, it is dominated.

In this algorithm, we use a data structure called a treap. This allows us to sort data by two values instead of one. We create a treap of undominated nodes organized by both the number of nodes they dominate, and node ID. All ties are dealt with randomly, i.e. if multiple nodes have the same domination number, then they all have the same possibility of being popped off the treap first. This gives the ability to come up with several different minimal domination sets with the same algorithm. The node with the largest domination number is popped from the treap of undominated nodes. It is checked to see if the domination number is the same as the number of nodes that are currently undominated when it would dominate. If not, the domination number is updated, and the node is put back on the treap to be reprioritized. If so, the node is added to the domination set, and its adjacencies are removed from the undominated nodes.

Once there is nothing left undominated, i.e. the undominated treap is empty, we make sure the domination set is minimal. To do this, we want to make sure there are no nodes on the set which do not need to be there. This occurs if a node and all of its adjacencies are dominated by something else in the domination set. In this case, the extra node is removed.

2. Independent Set

An independent set is a set of nodes that are not adjacent to each other. The goal is to find the minimum (or maximum) independent set. Generally, we want to find the maximal set of nodes that are not adjacent to one another. However, we can find a minimal set as well, where the minimum number of nodes is found such that every node is either in the set, or adjacent to a node on the set. This differs from a domination set by requiring that no nodes in the set are adjacent. We also use this algorithm to find an upper bound on the chromatic number.

An independent set is determined by putting the nodes on a stack in order of priority, pulling them off one at a time. If the node is adjacent to anything on the independent set, it is put in a discard pile. If not, it is added to the independent set. The key to this algorithm is the priority value. We used the average degree of adjacent uncovered nodes and we used degree as a priority. When looking for a maximal independent set, we used the degree, where the highest priority is the node with the lowest degree. For the minimal independent set, we used the highest average degree of uncovered adjacent nodes.

3. Chromatic Number

For finding the chromatic number, we found the largest independent set and colored it all one color (none of the nodes are adjacent, so this is legal). We then removed that set of nodes from the graph, and repeated the process. The number of iterations, different colors, is an upper bound on the chromatic number.

F. Validation

1. Test Graphs

The first step in validating our measurement software is to compare the results to manually calculated values for the seven small test graphs shown in Figure 53; the first four graphs are from Reference [BM 76], but the others we constructed ourselves. All of the properties measured for these graphs match the correct values calculated by hand.

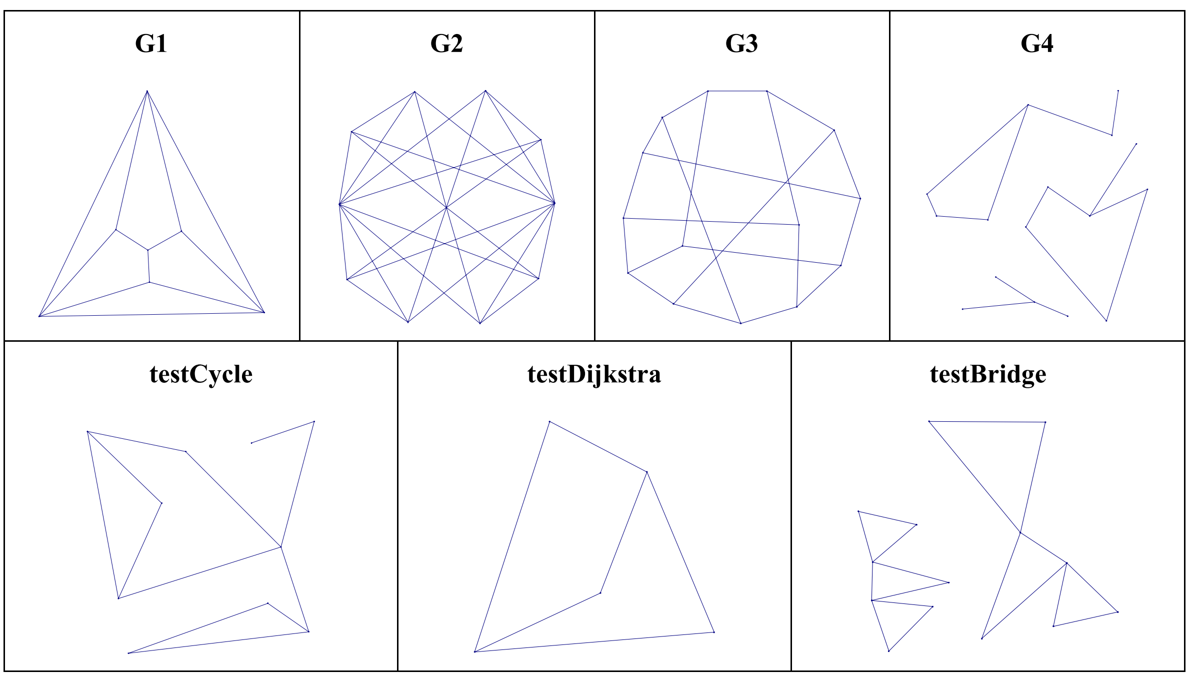

Figure 53. Layout of test graphs used in validation. The first four graphs come from Reference [BM 76].

2. Independent Measurements

Although the test graphs are useful for verifying the correctness of the measurement software, they do not represent a sufficient variety of graphs and graph "pathologies" to ensure that the measurements are correct. To address this issue, our team members separately and independently designed and implemented measurement algorithms in two different programming languages (C++ and Java). We compared to the measurements from the two programs verify that both gave the same results for all of the infrastructure graphs. Because the two programs were written by different teams, in different languages, using different designs, and (wherever possible) different algorithms, we have a high degree of confidence in the correctness of our results.

IV. Analysis of Intact Networks

In this section we analyze the characteristics of whole infrastructure graphs, where none of the vertices or edges have been deleted. Note that we do not include the analysis of the "interdependence" graph here because of its fundamentally composite nature and because we are still completing the computations measuring its properties—it will be discussed in a separate publication. The Appendix lists detailed results for the 43 graphs of the three data sets. This report also has a companion DVD-R disc containing complete detailed results for all of the graphs.

A. Individual Measurements

We now discuss the basic characteristics of the graphs. Figure 54, Figure 55, and Figure 56 illustrate the distributions of the key measurements for the "raw,", "clean0," and "clean1" data sets, respectively. One can see that the cleaning process typically broadens the distributions. The bimodality in some distributions results from the presence of two types of graphs in the data sets: sparsely connected tree- or grid-like graphs (e.g., pipelines and road networks) and relatively highly connected graphs (e.g., communications networks).

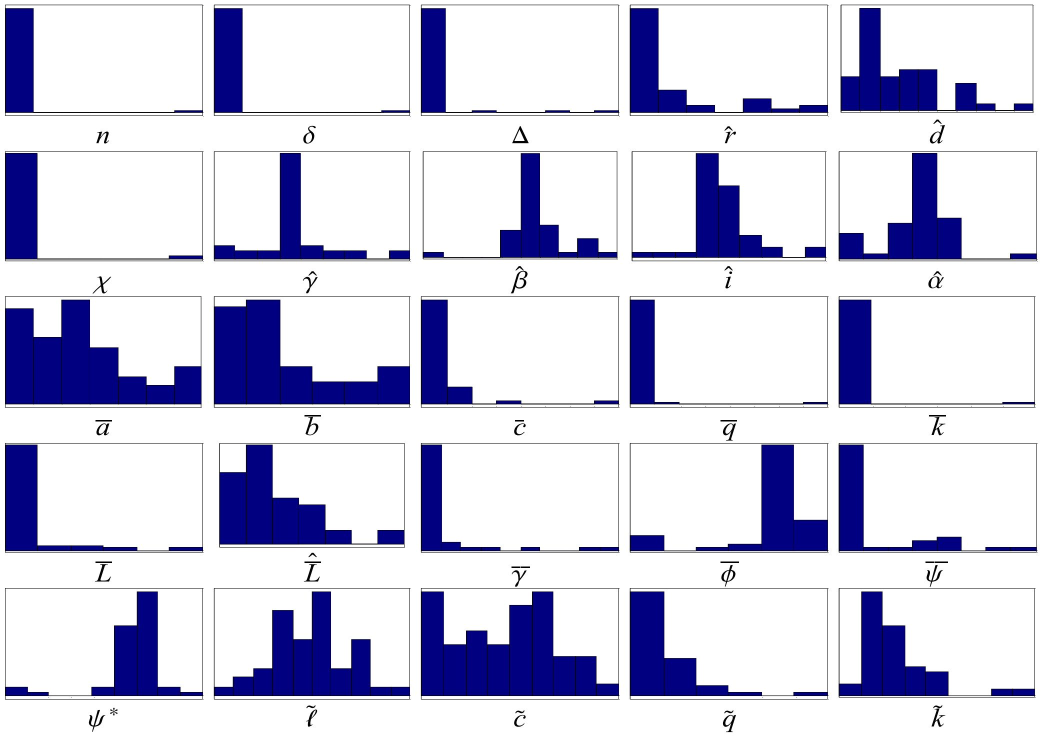

Figure 54. Distribution of key measurements in the "raw" data set.

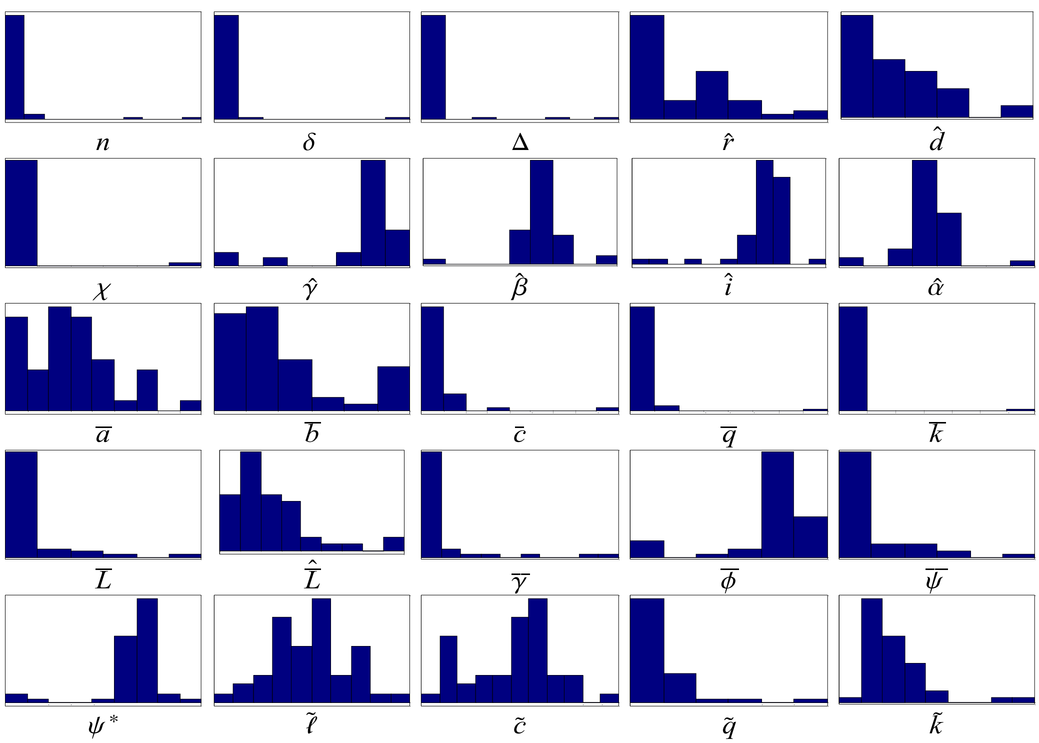

Figure 55. Distribution of key measurements in the "clean0" data set.

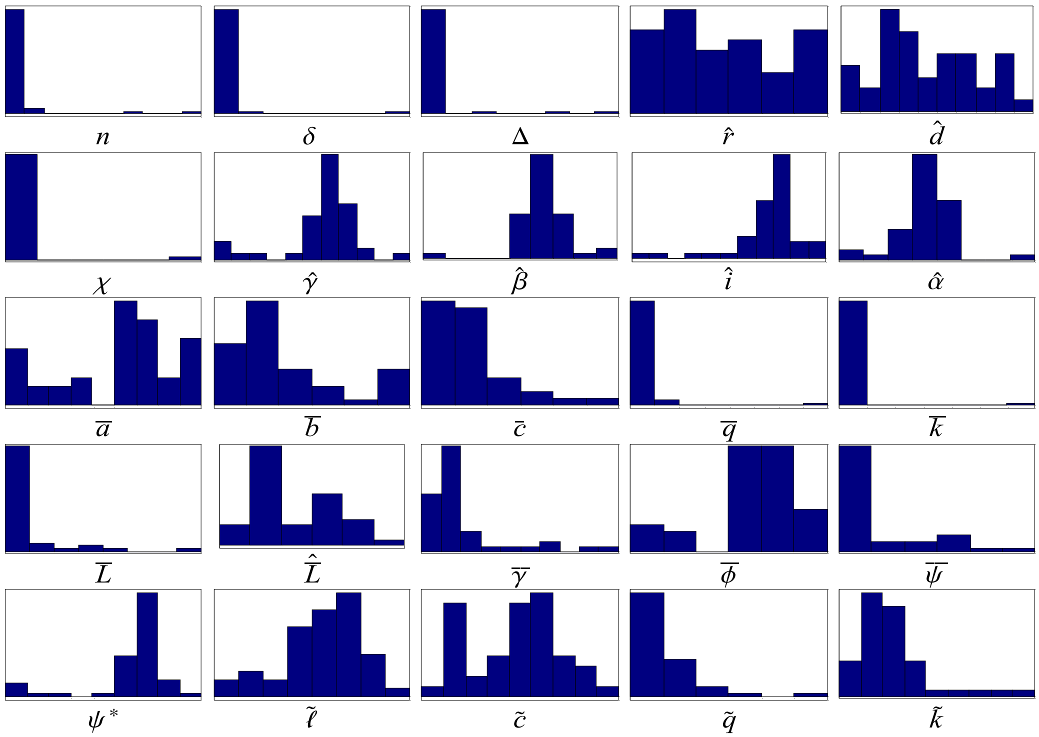

Figure 56. Distribution of key measurements in the "clean1" data set.

Figure 57 and Figure 58 show more precisely how individual measurements change during the cleaning process. Discarding the small components (the process that transforms "raw" to "clean0") has minor effects on most measurements because the larger components dominate average quantities and because the small components are typically not highly connected (so they do not affect measurements like chromatic number). The process of combining edges at degree-two vertices has a more profound effect on measurements. Measurements involving distance (normalized radius, characteristic path length, cycle length) generally become smaller, but not in a linear fashion. Measurements involving connectedness are also altered by the graph becoming more connected as weakly connected vertices are removed. The sizes of cut sets are hardly affected at all by this transformation of the graph. Entropies are likewise left relatively unchanged. For the rest of our analysis, we will consider only the "clean1" data set.

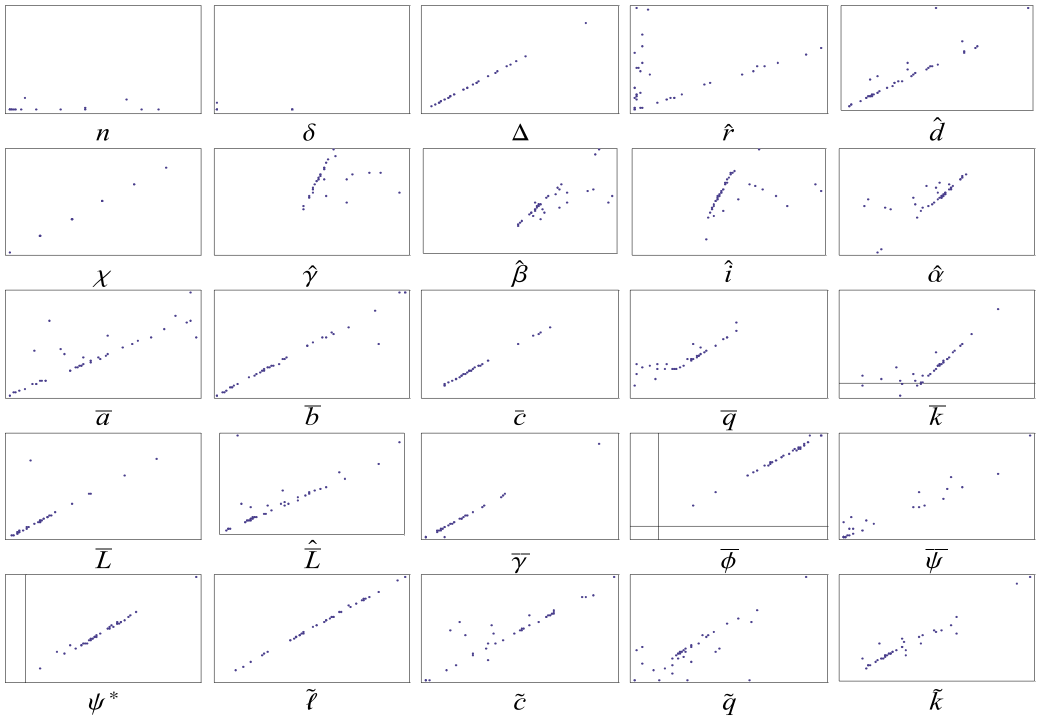

Figure 57. Effect of cleaning process on key measurements: the horizontal axis shows the value of the measurement in the "raw" data set and the vertical axis shows it in the "clean0" one.

Figure 58. Effect of cleaning process on key measurements: the horizontal axis shows the value of the measurement in the "clean0" data set and the vertical axis shows it in the "clean1" one.

B. Dependencies between Measurements

Because some of the measurements quantify similar properties of the graphs, we can expect there to be dependencies between certain of the variables. It is useful to measure this because some quantities are expensive (in terms of computational effort) to measure while others are trivial. It may be the case that certain easily measured quantities can be used as surrogates for the more difficult to measure. Since dealing with so many variables can be daunting, it would also be nice to settle on a smaller representative subset of variables for future studies.

1. Correlation Matrices

In Table 30 we display the correlation matrix between 24 graph measurements.

Table 30. Correlation matrix of key measurements for "clean1" data set. Entries in bold type have correlation coefficients greater than 0.70.

	delta	DDelta	rHat	dHat	chi	gammaHat	betaHat	iHat	alphaHat	aBar	bBar	cBar
delta	1.00	-0.04	-0.17	-0.21	**0.77**	-0.52	-0.62	-0.67	0.57	-0.40	-0.28	-0.18
DDelta	-0.04	1.00	-0.35	-0.42	0.52	-0.61	0.57	-0.17	-0.60	-0.38	-0.21	-0.23
rHat	-0.17	-0.35	1.00	**0.88**	-0.46	0.47	-0.07	0.33	0.10	0.51	0.38	0.30
dHat	-0.21	-0.42	**0.88**	1.00	-0.53	0.53	-0.15	0.35	0.19	0.44	0.37	0.33
chi	**0.77**	0.52	-0.46	-0.53	1.00	**-0.74**	-0.23	-0.53	0.18	-0.59	-0.43	-0.29
gammaHat	-0.52	-0.61	0.47	0.53	**-0.74**	1.00	-0.10	**0.74**	0.15	**0.78**	0.47	0.47
betaHat	-0.62	0.57	-0.07	-0.15	-0.23	-0.10	1.00	0.34	**-1.00**	0.27	0.46	0.05
iHat	-0.67	-0.17	0.33	0.35	-0.53	**0.74**	0.34	1.00	-0.31	0.68	0.52	0.35
alphaHat	0.57	-0.60	0.10	0.19	0.18	0.15	**-1.00**	-0.31	1.00	-0.24	-0.45	-0.03
aBar	-0.40	-0.38	0.51	0.44	-0.59	**0.78**	0.27	0.68	-0.24	1.00	**0.83**	0.59
bBar	-0.28	-0.21	0.38	0.37	-0.43	0.47	0.46	0.52	-0.45	**0.83**	1.00	0.44
cBar	-0.18	-0.23	0.30	0.33	-0.29	0.47	0.05	0.35	-0.03	0.59	0.44	1.00
qBar	**0.91**	-0.02	-0.24	-0.28	**0.71**	-0.68	-0.56	-0.75	0.51	-0.60	-0.46	-0.27
kBar	**0.85**	0.11	-0.29	-0.36	**0.73**	**-0.77**	-0.42	**-0.77**	0.37	-0.64	-0.48	-0.30
LBarHat	-0.16	-0.42	**0.88**	**0.93**	-0.51	0.50	-0.11	0.33	0.14	0.48	0.46	0.22
gammaBar	0.64	0.58	-0.27	-0.38	**0.80**	**-0.72**	-0.13	-0.59	0.07	-0.58	-0.44	-0.46
phiBar	-0.52	-0.47	0.29	0.38	-0.65	**0.78**	0.16	0.64	-0.11	**0.74**	0.61	0.55
psiBar	-0.11	0.07	0.40	0.29	-0.15	0.11	0.18	0.27	-0.17	0.16	0.28	-0.24
psiStar	-0.68	0.20	-0.04	-0.01	-0.36	0.40	0.60	0.60	-0.58	0.45	0.44	0.20
elTilde	-0.47	-0.48	**0.78**	**0.89**	**-0.76**	**0.74**	-0.04	0.51	0.09	0.56	0.45	0.32
cTilde	-0.45	-0.39	0.64	0.59	-0.61	**0.73**	-0.06	0.55	0.12	0.57	0.20	0.21
qTilde	0.49	0.07	-0.26	-0.27	0.41	-0.54	-0.44	-0.61	0.42	**-0.70**	**-0.72**	-0.40
kTilde	0.35	-0.07	0.17	0.01	0.15	-0.30	-0.16	-0.27	0.14	-0.21	-0.12	-0.37

Table 30 (continued). Correlation matrix of key measurements for "clean1" data set. Entries in bold type have correlation coefficients greater than 0.70.

	qBar	kBar	LBarHat	gammaBar	phiBar	psiBar	psiStar	elTilde	cTilde	qTilde	kTilde
delta	**0.91**	**0.85**	-0.16	0.64	-0.52	-0.11	-0.68	-0.47	-0.45	0.49	0.35
Ddelta	-0.02	0.11	-0.42	0.58	-0.47	0.07	0.20	-0.48	-0.39	0.07	-0.07
rHat	-0.24	-0.29	**0.88**	-0.27	0.29	0.40	-0.04	**0.78**	0.64	-0.26	0.17
dHat	-0.28	-0.36	**0.93**	-0.38	0.38	0.29	-0.01	**0.89**	0.59	-0.27	0.01
chi	**0.71**	**0.73**	-0.51	**0.80**	-0.65	-0.15	-0.36	**-0.76**	-0.61	0.41	0.15
gammaHat	-0.68	**-0.77**	0.50	**-0.72**	**0.78**	0.11	0.40	**0.74**	**0.73**	-0.54	-0.30
betaHat	-0.56	-0.42	-0.11	-0.13	0.16	0.18	0.60	-0.04	-0.06	-0.44	-0.16
iHat	**-0.75**	**-0.77**	0.33	-0.59	0.64	0.27	0.60	0.51	0.55	-0.61	-0.27
alphaHat	0.51	0.37	0.14	0.07	-0.11	-0.17	-0.58	0.09	0.12	0.42	0.14
aBar	-0.60	-0.64	0.48	-0.58	**0.74**	0.16	0.45	0.56	0.57	**-0.70**	-0.21
bBar	-0.46	-0.48	0.46	-0.44	0.61	0.28	0.44	0.45	0.20	**-0.72**	-0.12
cBar	-0.27	-0.30	0.22	-0.46	0.55	-0.24	0.20	0.32	0.21	-0.40	-0.37
qBar	1.00	**0.98**	-0.23	0.61	-0.68	-0.14	**-0.84**	-0.54	-0.55	**0.71**	0.51
kBar	**0.98**	1.00	-0.30	0.66	**-0.75**	-0.14	**-0.82**	-0.62	-0.61	**0.74**	0.53
LBarHat	-0.23	-0.30	1.00	-0.30	0.32	0.48	-0.06	**0.89**	0.59	-0.25	0.21
gammaBar	0.61	0.66	-0.30	1.00	**-0.89**	0.22	-0.50	-0.53	-0.45	0.59	0.50
phiBar	-0.68	**-0.75**	0.32	**-0.89**	1.00	-0.19	0.68	0.52	0.47	**-0.82**	-0.66
psiBar	-0.14	-0.14	0.48	0.22	-0.19	1.00	-0.09	0.42	0.36	0.09	0.61
psiStar	**-0.84**	**-0.82**	-0.06	-0.50	0.68	-0.09	1.00	0.17	0.27	**-0.85**	**-0.73**
elTilde	-0.54	-0.62	0.89	-0.53	0.52	0.42	0.17	1.00	**0.72**	-0.35	0.00
cTilde	-0.55	-0.61	0.59	-0.45	0.47	0.36	0.27	**0.72**	1.00	-0.30	-0.01
qTilde	**0.71**	**0.74**	-0.25	0.59	**-0.82**	0.09	**-0.85**	-0.35	-0.30	1.00	0.66
kTilde	0.51	0.53	0.21	0.50	-0.66	0.61	**-0.73**	0.00	-0.01	0.66	1.00

2. Principal Components

Another straightforward way to study dependencies between measurements is to use principal components analysis (PCA). Table 31 shows the results of performing principal components analysis on 24 standardized measurements of the infrastructure graphs. (We did not consider the unnormalized or extensive variables.) The first component explains nearly half of the variation, while the first three explain three-quarters. There seem to be significant contributions from the first eight components. All but the fifth component have significant contributions from several variables—the fifth one depends solely on the graph's girth.

Table 31. Principal components analysis of key measurements (first twelve components only) for “clean1” data set.

Measurement	*Importance*	*Principal Component*											
		1st	*2nd*	*3rd*	*4th*	*5th*	*6th*	*7th*	*8th*	*9th*	*10th*	*11th*	*12th*
delta	21	0.23	0.18	-0.12	0.35	0.00	0.04	0.27	0.14	0.08	0.08	-0.07	0.12
Ddelta	3	0.11	-0.30	0.29	0.04	0.00	0.40	0.07	-0.25	-0.05	0.03	0.01	0.23
rHat	7	-0.17	0.29	0.17	0.15	0.00	0.30	-0.10	-0.04	0.31	-0.25	0.11	-0.60
dHat	15	-0.18	0.30	0.09	0.08	0.00	0.38	-0.19	0.06	-0.16	-0.10	0.12	0.21
g	1	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
chi	8	0.25	-0.07	-0.01	0.24	0.00	0.23	0.40	-0.04	-0.08	-0.22	-0.05	0.11
gammaHat	14	-0.27	0.12	-0.14	-0.11	0.00	-0.15	0.23	-0.14	-0.10	0.14	0.31	0.02
betaHat	23	-0.10	-0.34	0.33	0.11	0.00	-0.07	-0.25	-0.03	0.14	-0.09	-0.03	0.13
iHat	9	-0.25	-0.06	0.07	-0.06	0.00	-0.14	0.32	-0.21	-0.36	-0.67	0.17	0.04
alphaHat	20	0.09	0.34	-0.33	-0.14	0.00	0.07	0.24	0.02	-0.11	0.08	-0.03	-0.08
aBar	10	-0.26	0.03	0.00	0.30	0.00	-0.24	0.18	-0.14	0.30	0.12	0.32	0.20
bBar	19	-0.21	-0.03	0.12	0.49	0.00	-0.26	0.03	0.24	-0.11	0.17	0.04	0.10
cBar	4	-0.15	0.02	-0.21	0.39	0.00	0.04	-0.15	-0.72	-0.15	0.16	-0.33	-0.17
qBar	24	0.27	0.17	-0.07	0.21	0.00	-0.07	-0.05	0.06	0.05	-0.22	-0.13	0.09
kBar	22	0.28	0.12	-0.02	0.21	0.00	-0.08	-0.14	0.00	0.11	-0.25	-0.09	0.11
LBarHat	11	-0.17	0.32	0.19	0.13	0.00	0.22	-0.11	0.17	-0.10	-0.06	-0.02	0.31
gammaBar	13	0.25	0.01	0.25	0.09	0.00	0.19	0.30	-0.11	0.01	0.34	0.34	-0.09
phiBar	17	-0.27	-0.05	-0.25	0.09	0.00	-0.04	0.02	0.17	0.02	-0.08	-0.31	0.14
psiBar	6	-0.05	0.15	0.49	-0.05	0.00	-0.17	0.33	0.07	-0.25	0.13	-0.53	-0.19
psiStar	16	-0.20	-0.31	-0.03	-0.04	0.00	0.18	0.21	0.19	0.09	0.04	-0.19	-0.12
elTilde	18	-0.24	0.25	0.11	-0.09	0.00	0.14	-0.11	0.04	-0.24	0.25	-0.01	0.12
cTilde	5	-0.21	0.18	0.07	-0.26	0.00	0.07	0.27	-0.20	0.62	-0.04	-0.27	0.32
qTilde	2	0.24	0.17	0.08	-0.25	0.00	-0.16	-0.17	-0.30	-0.11	0.04	-0.03	0.30
kTilde	12	0.13	0.25	0.36	-0.01	0.00	-0.41	-0.03	-0.08	0.12	-0.07	0.06	-0.11
Proportion of Variance (%)		43.9	19.0	11.2	6.3	4.1	4.1	3.5	2.7	1.4	1.2	0.8	0.4

We have also estimated the “importance” of the variables in Table 31 using the following procedure:

1. Start with the set of all variables.
2. Find the principal components for the current set of variables.
3. Discard the variable that has the largest projection onto the last principal component.
4. Go to step 2.

We hypothesize that this procedures removes the least important (i.e., the most redundant) variables first. This ranking can be used as a guide for variable selection.

3. Regression

Using the understanding of dependencies gathered above, we can make estimates for quantities that are expensive to measure. Perhaps the most computationally costly measurement is a minimum cut. A simple regression on several other measurements is a good predictor of this, however. In this illustrative case, the normalized domination and independence numbers along with the mean fraction of bridges and degree provide a good predictor of the mean size of a minimum cut in the graph (see Figure 59):

$$\mathsf{E}(\overline{q} \mid \hat{\gamma}, \hat{\beta}, \overline{b}, \overline{k}) \approx 5.36 - 5.31\hat{\gamma} - 5.87\hat{\beta} + 0.58\overline{b} + 0.35\overline{k} .$$

This simple regression never differs from the observation by more than about half an edge for $\overline{q} < 5$. (The outlier in the lower part of the plot is the “communications-gpc” graph.) For the three cases with $\overline{q} > 5$, this regression is marginally adequate, but a piecewise regression or more sophisticated smoothing technique could be used. Since most of the measured quantities correlate with a least one other quantity, we expect to be able to construct similar regressions for them.

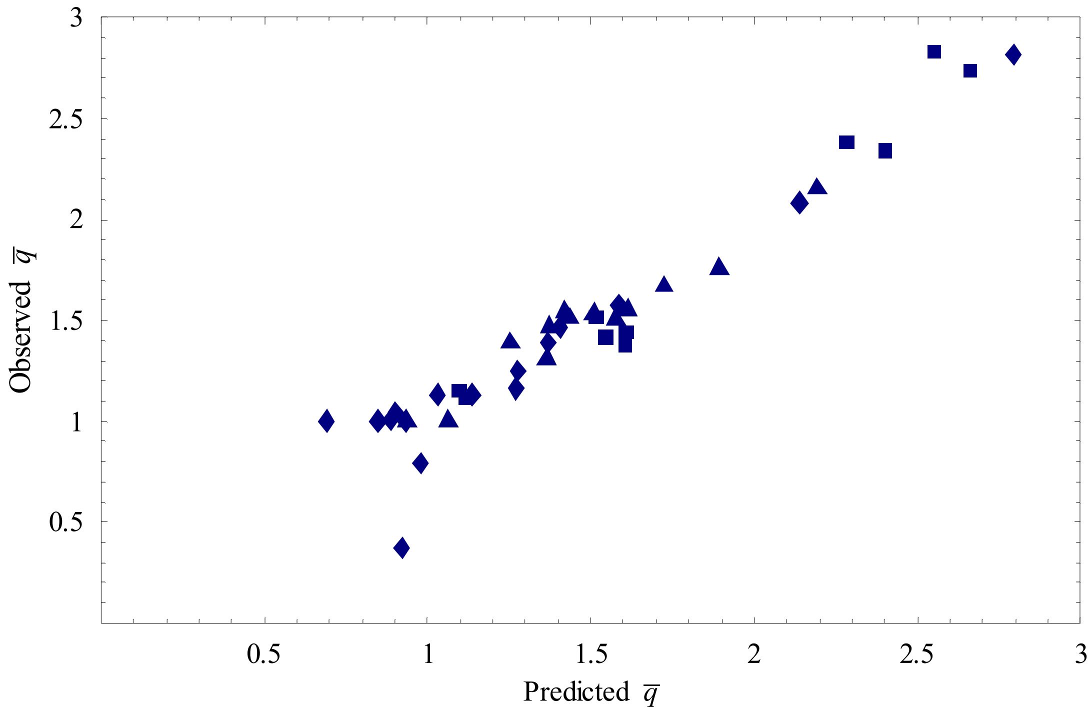

Figure 59. Regression results for the mean size of the minimum cuts. The boxes represent transportation networks, the diamonds communications networks, the triangles energy networks, and the stars logistics networks.

4. Specific Interesting Relationships

We next highlight some of the most interesting relationships in the "clean1" data set. These are plotted in Figure 60.

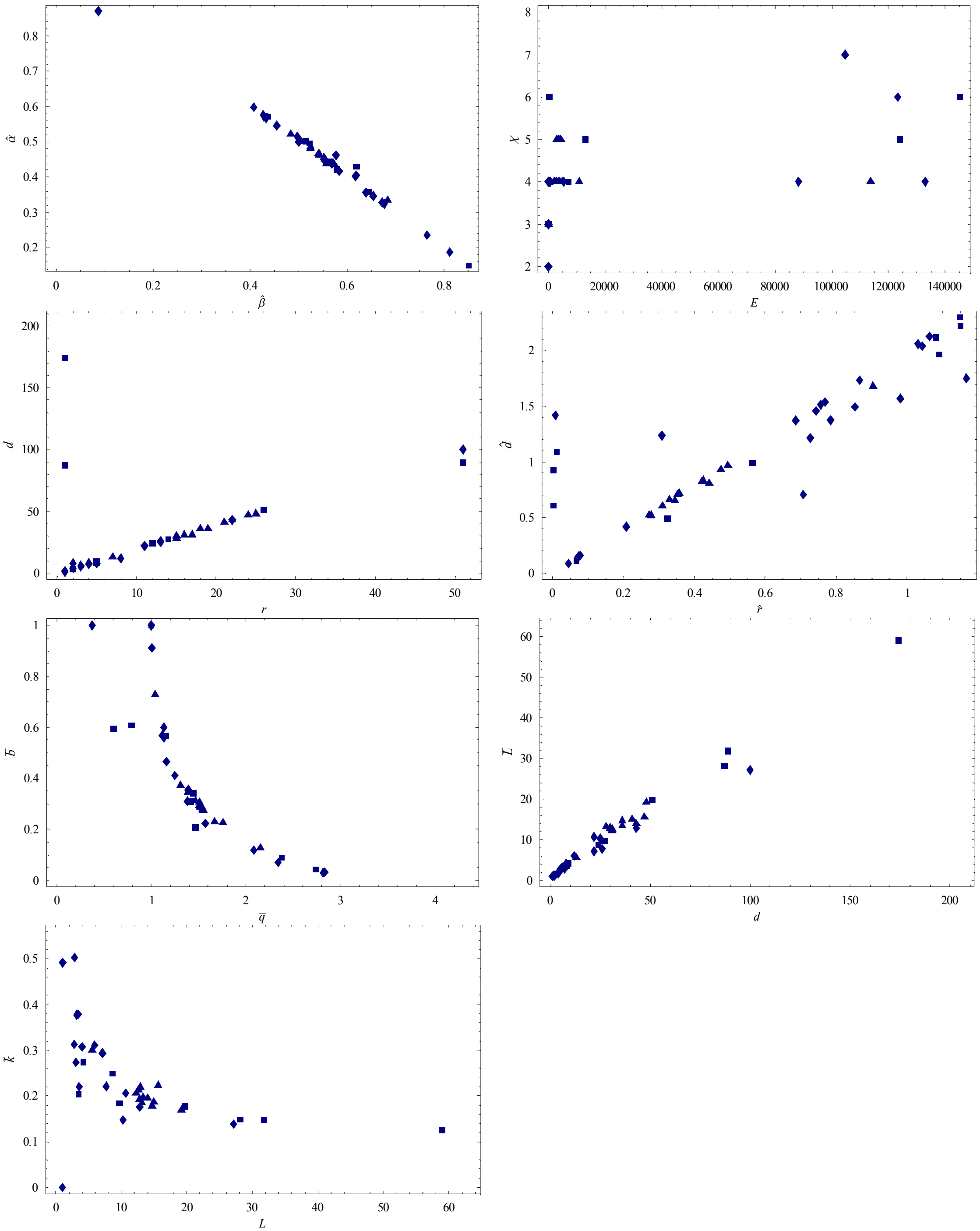

Figure 60. Selected interesting two-way dependencies in the measurements for the “clean1” data set. The boxes represent transportation networks, the diamonds communications networks, the triangles energy networks, and the stars logistics networks.

Covering & Independence Numbers. This is interesting in that from the graph it is apparent that the normalized covering number upper bound, and the normalized independence number lower bound are inversely proportional. Also of interest is the association of like infrastructures. We see in the graph a loose bundling of infrastructures, especially communications.

Chromatic Number & Size. This graph is interesting purely in the lack of correlation between the two measurements. Obviously, the size of the graph does not dictate what the chromatic number will be. The infrastructure types are not grouped together.

Radius & Diameter. This graph points out the 1:2 ratio we expect between radius and diameter given our general understanding of these concepts. Of more interest are the outliers. Also note that the like infrastructures are grouped together.

Normalized Radius & Normalized Diameter. In this graph we see that the relationship between radius and diameter is maintained, even the outliers, although the measurements have been normalized.

Bridge Fraction & Average Minimum Cut. The correlation makes sense in that as the number of bridges increases, the expected average minimum cut should decrease. Namely, the more bridges there are, the few number of edges must be cut to isolate one node from another. The infrastructures are not grouped by type.

Diameter & Characteristic Path Length. This is a near linear graph, which makes sense. If the average path length is large, then it is expected that the longest path length (diameter) is correspondingly large as well. The converse is also true.

Normalized Degree Distribution Entropy & Characteristic Path Length. Again, we have an interesting shape, with these two measures being inversely proportional. This can seemingly be rationalized by considering that the better something is connected, it is easier to get to, or away from. This means that as the degree increases, it becomes easier to get from node to node, decreasing the characteristic path length. The converse is also true.

C. Scaling Behavior

We have also looked for scaling behavior in the distributions of vertex properties. In Figure 61 we show the distribution of the degree of vertices on a log-log scale. For the non-transportation graphs, this distribution appears to be consistent with a power law

$$P(k) \propto k^{-D}$$

where the exponent depends on the particular type of infrastructure. Table 32 shows that the values of D are similar for the electric power infrastructures. Note that the quality of fit is generally not good enough to distinguish a power law from exponential behavior like

$$P(k) \propto e^{-D'k}.$$

Figure 62 and Figure 63 demonstrate some evidence for a scaling law in the distribution of shortest cycle length and minimum cut size in these graphs. Figure 64 illustrates that the distribution of shortest path lengths does possess a scale.

Table 32. Exponents for power-law scaling of degree distribution for selected graphs in data set "clean1."

graph	*exponent*
electric-ecar	2.36
electric-ercot	2.42
electric-frcc	2.23
electric-maac	2.66
electric-main	2.58
electric-mapp	2.46
electric-nepool	2.56
electric-nypp	2.61
electric-serc	2.74
electric-spp	2.33
electric-wscc	2.86
highways-mexico	1.68
highways-usa	1.22
internet-bell	2.78
internet-nlanr	1.17
microwave-fcc	2.72
petroleum-mapsearch	4.68
railways-canada	2.16
railways-mexico	0.78
railways-usa	4.92
roads-dallas	1.88
roads-portland	2.62
telephone-lerg	1.49
waterways-usa	2.76

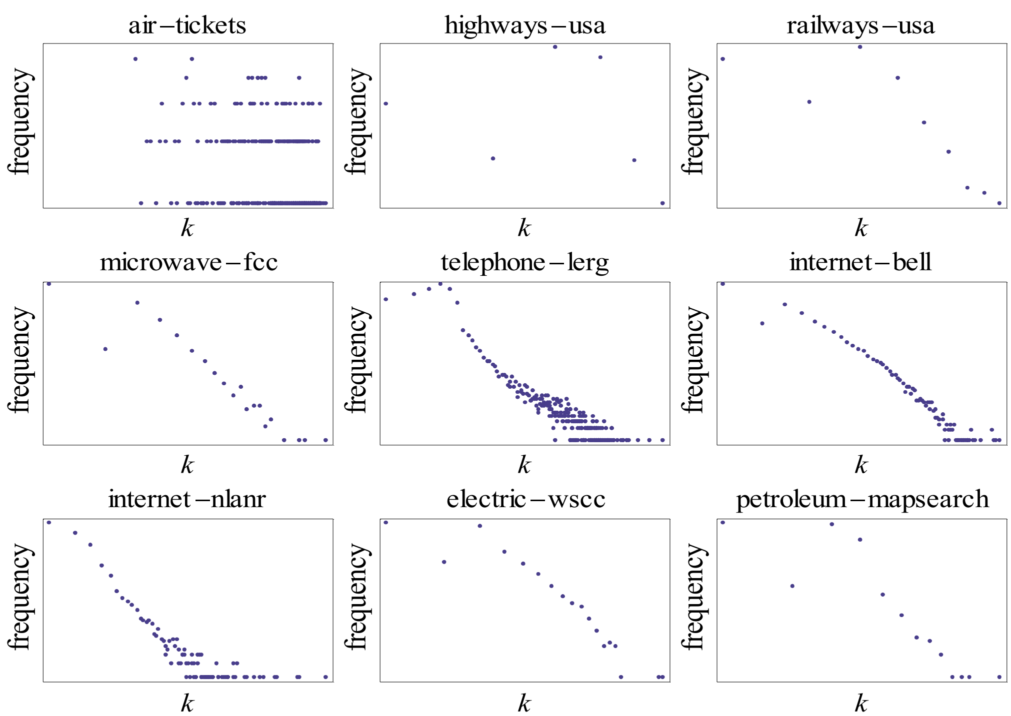

Figure 61. Degree distribution, plotted on a log-log scale, for some of the larger graphs in data set "clean1."

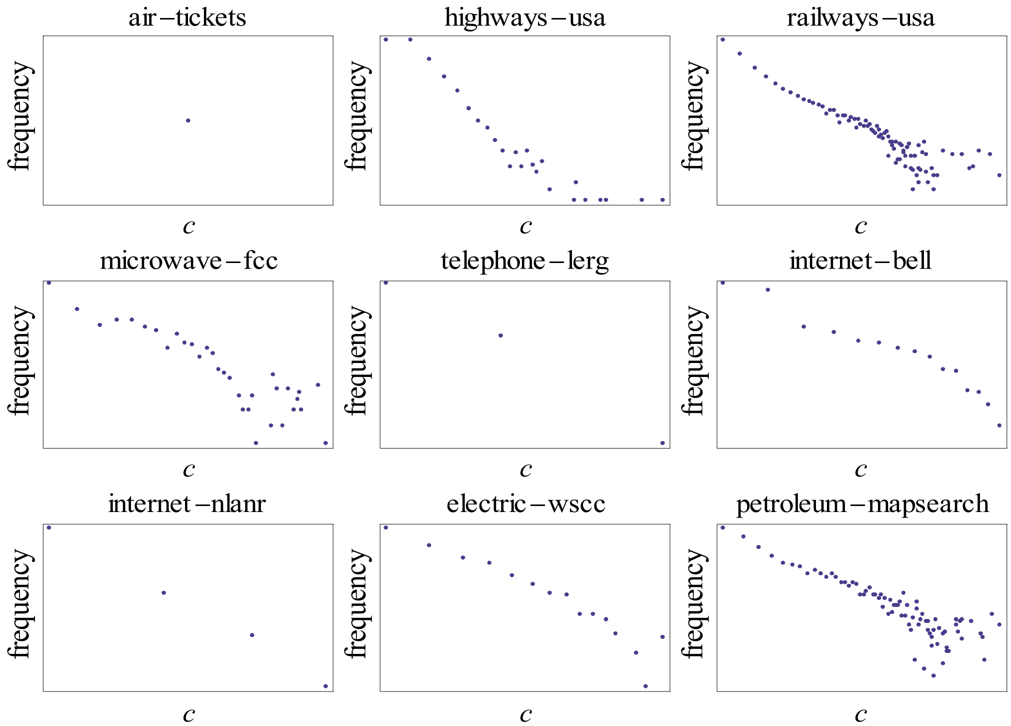

Figure 62. Shortest cycle length distribution, plotted on a log-log scale, for some of the larger graphs in data set "clean1."

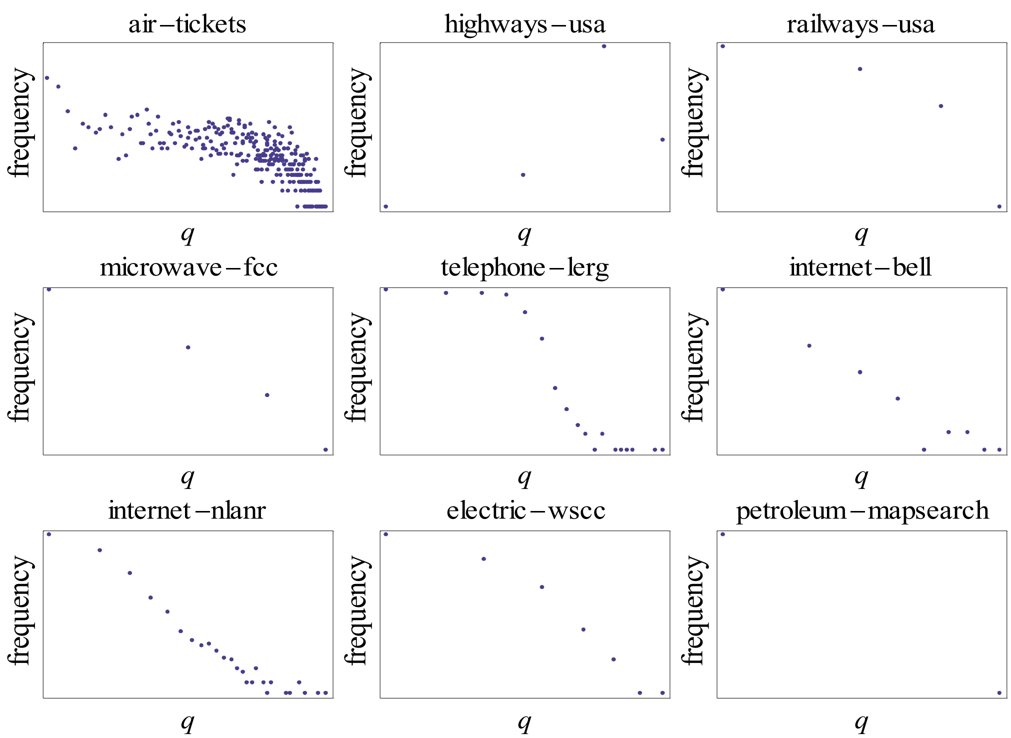

Figure 63. Minimum cut size distribution, plotted on a log-log scale, for some of the larger graphs in data set "clean1."

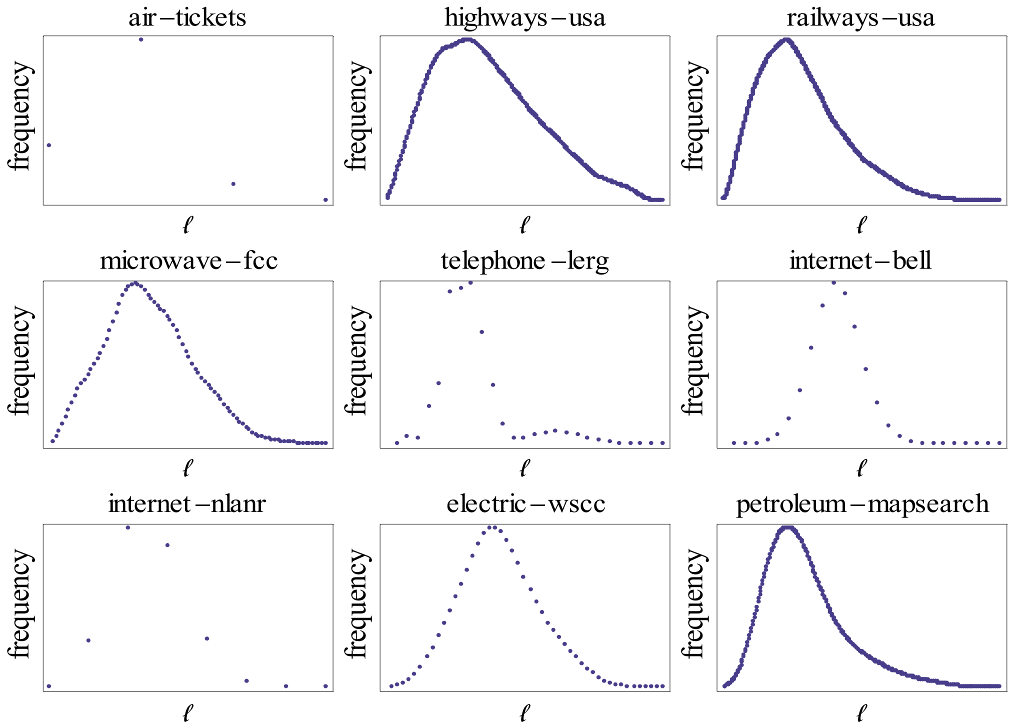

Figure 64. Shortest path length distribution, plotted on a linear-linear scale, for some of the larger graphs in data set "clean1."

D. Clustering

It is also natural to ask which graphs are most similar to which other graphs. We investigated this by applying agglomerative clustering algorithms based on various subsets of the measurements. Because we have dozens of measurements and no natural metric for constructing a dissimilarity matrix, we limited ourselves to considering the 25 intensive measurements that do not scale with a graph's order and we used a Manhattan distance for comparisons instead of a Euclidean one. We also use the "link average" method for combining clusters in order to minimize the effect of any single type of measurement. This still leaves us with 2^{25} possible subsets of variables to use. Using the full set of 25 did not produce a compelling hierarchy, so we focused on combinations of variables that seem to be most independent in our correlation and principal components analyses. Figure 65 shows the very interesting result of using the four entropies we measure (i.e., for degree, minimum cut, shortest path, and minimum cycle distribution). The dendrogram possesses tight coupling between graphs of similar types: the communications, electric power, air travel, and internet graphs are especially close together; there seems to be some mixing of road, highway, rail, and pipeline networks. The remarkable fact about this set of four variables is that they do not depend on the numerical value of any properties of the graph, just the distribution—in particular, it is insensitive to the value of the mean degree, minimum cut, shortest path, or minimum cycle.

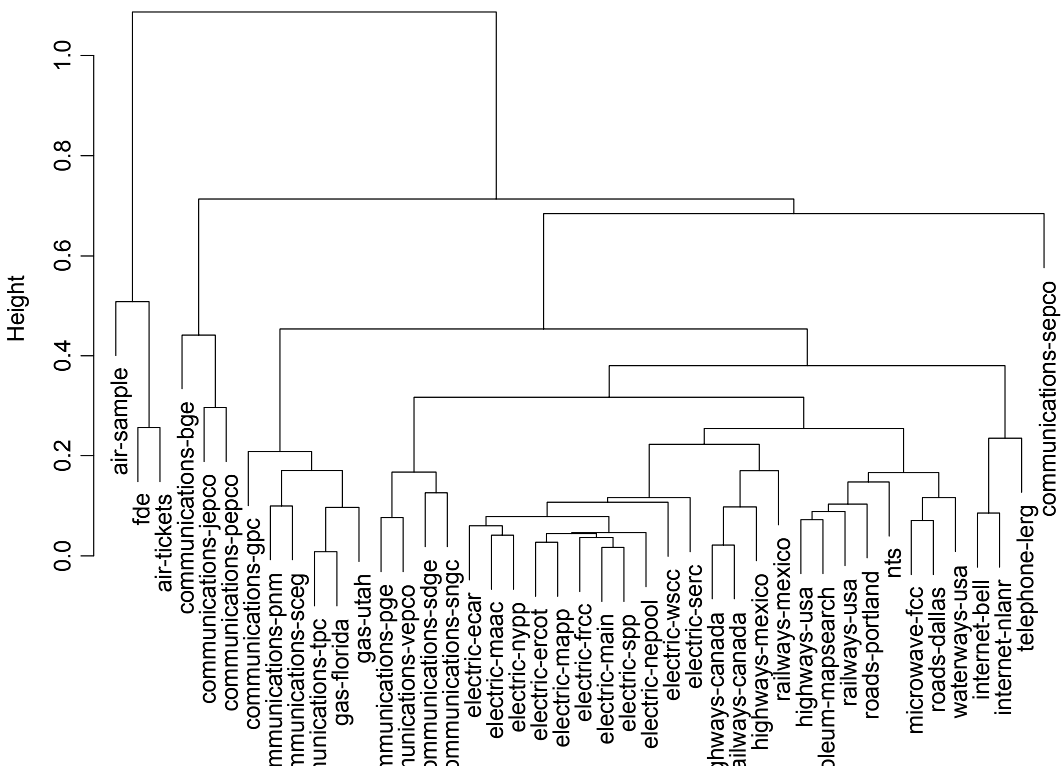

Figure 65. Hierarchical tree of graph in the "clean1" data set, generated using agglomerative clustering with a Manhattan distance and the average link method on the four entropy measurements.

E. Classification

1. Classification Tree

In addition to clustering graphs based on their properties, we can classify them based on their known types or subtypes. This results in a procedure for predicting the type of an unknown graph. Figure 66 and Figure 67 illustrate classification trees (CART) for graph type and subtype based on the intensive measurements of the graphs; these trees are among the ones with the fewest nodes that have no classification error. In both cases the communication networks contribute most to the complexity of the tree—if it were not for the heterogeneity of these, more parsimonious trees would be possible. Because of the small number of graphs we are dealing with, we have not pursued cross-validation studies to determine the robustness of these trees.

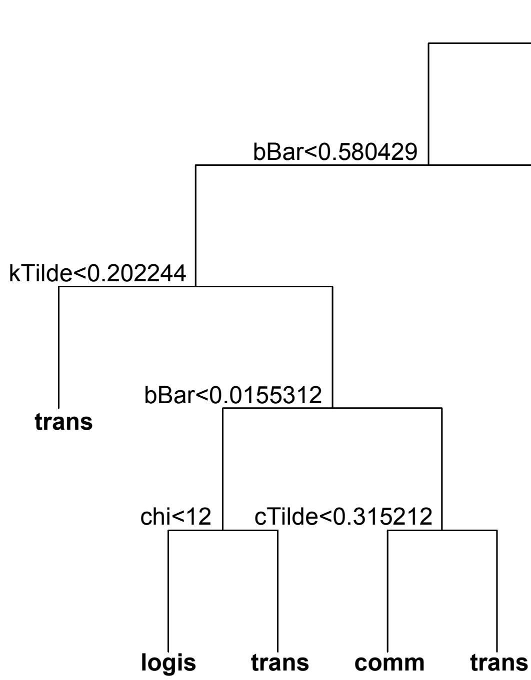

Figure 66. One of the simplest classification trees that do not misclassify infrastructure types in the "clean1" data set.

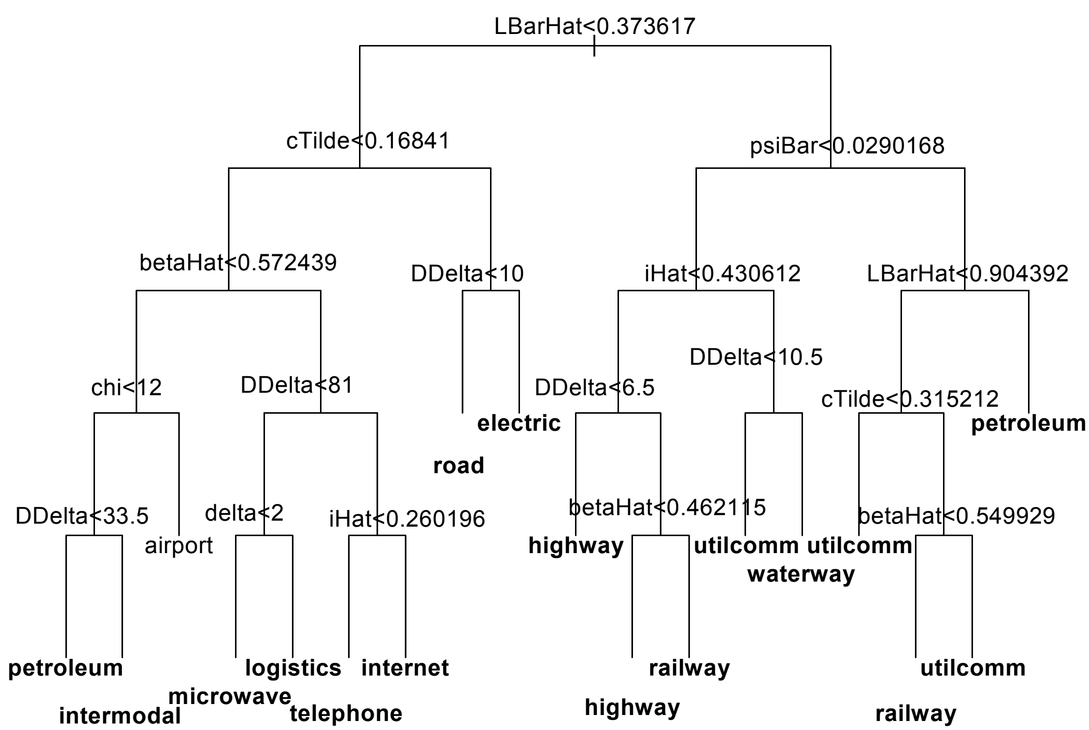

Figure 67. One of the simplest classification trees that do not misclassify infrastructure subtypes in the "clean1" data set.

2. Multinomial Logit

We have also developed a multinomial logit regression for classifying graphs. We cursorily determined a sufficient subset of the measurements (nine quantities in all) for predicting the type of graph perfectly for the "clean1" data set:

$$P_{\text{trans}} \propto \exp\left(-1334 - 344\chi + 16332\bar{a} - 1787\bar{b} + 752\bar{k} - 9434\hat{\imath} - 34554\hat{\beta} - 7777\hat{\bar{L}} + 48470\tilde{\ell} + 16657\psi^*\right)$$
$$P_{\text{comm}} \propto \exp\left(-924 - 6838\chi - 65778\bar{a} - 28509\bar{b} - 7674\bar{k} + 109629\hat{\imath} + 263567\hat{\beta} + 5389\hat{\bar{L}} - 94091\tilde{\ell} - 150097\psi^*\right)$$
$$P_{\text{energ}} \propto \exp\left(+20.58 + 0.42\chi + 53.78\bar{a} + 21.26\bar{b} - 0.81\bar{k} - 76.06\hat{\imath} - 128.57\hat{\beta} + 7.33\hat{\bar{L}} - 80.40\tilde{\ell} + 109.88\psi^*\right)$$
$$P_{\text{logis}} \propto \exp\left(-11.36 - 5.53\chi + 10.60\bar{a} - 7.72\bar{b} + 4.23\bar{k} + 5.74\hat{\imath} + 1.18\hat{\beta} - 0.79\hat{\bar{L}} - 31.27\tilde{\ell} + 1.84\psi^*\right)$$

Unfortunately, this regression contains 40 parameters, so it likely overfits the data. Since we have not investigated this approach in very much detail, we cannot rule out the possibility that a more parsimonious model would perform equally well.

F. Fuzzy Characterization

Fuzzy set theory can also be used to classify and categorize graphs. In a separate report [FS 00] we develop fuzzy membership functions for some different types of graphs and apply them to a selection of our data set.

G. Generative Models

A major goal of this work is to develop generative models of infrastructure graphs, so that one can create random graphs that are statistically indistinguishable from actual infrastructure graphs of a given type. If the algorithm for generating the graph is simple enough, then it may be possible to prove significant results about infrastructure graphs in general. This work is still underway and is being published separately [AGB 01].

H. Detecting Anomalous Components

Infrastructure graphs that are very large almost certainly contain mistakes in their representation. We have investigated techniques for identifying suspicious elements in the graph. Although this work is still underway and will be published separately, we would like to outline our approaches to this problem.

Given a large graph, we can compute for the vertices and edges the properties listed in Table 1. Considered the multidimensional empirical distribution functions for $(k, e, a, L, \bar{\gamma}, \psi, c)$ of the vertices or for $(k_1, e_1, a_1, L_1, \bar{\gamma}_1, \psi_1, c_1, k_2, e_2, a_2, L_2, \bar{\gamma}_2, \psi_2, c_2, b, \phi)$ of the edges, where the subscripts refer to the end vertices of the edge. Any subset of these variables could also be considered. One way to identify suspicious elements is to locate the outlying points in the distribution. We use the concept of *data depth* [LPS 99] to define how far an observation is from the edge of the distribution. In our simple implementation we compute the convex hull of the empirical distribution and consider the observations on this hull to be the most suspicious vertices or edges. These can then be manually checked for corruption. It is practical to perform this computation even for very large graphs.

Another approach is to consider the empirical distribution functions for $(k, \hat{e}, a, \hat{L}, \bar{\gamma}, \psi, c)$ of the vertices or for $(k_1, \hat{e}_1, a_1, \hat{L}_1, \bar{\gamma}_1, \psi_1, c_1, k_2, \hat{e}_2, a_2, \hat{L}_2, \bar{\gamma}_2, \psi_2, c_2, b, \phi)$ of the edges for a reference graph, such as one of the graphs discussed in this report, and determine which observations of a graph of interest lie outside the convex hull of the reference graph. This technique even applies when the graph of interest is small.

A third way to identify erroneous vertices or edges is to use supervised learning techniques on a graph (either a reference graph or the graph of interest) that has been intentionally contaminated with bad data. The learning method is trained with the contaminated data set to identify the known contamination. The graph of interested (without the added contamination) can then be classified to locate possible contamination.

The techniques described above may locate parts of the graph near where vertices or edges are missing. The Bayesian techniques used for generative graph models (discussed in the previous section) could also be adapted to compute probabilities that vertices or edges are missing in particular parts of the graph.

I. Comparison with Other Work

In Table 33 we compare our results with some previous work [Wa 99] that has been done on the WSCC power system. Some of the measurements are closest to our "clean1" data set while others are closer to our "raw" data set; this probably results from the level of aggregation present in the smaller graph used in Reference [Wa 99]. We expect agreement would be better if we could clean that representation so it has no artifacts such as small components or degree-two vertices.

Table 33. Comparison of our work with other results for the WSCC electric power system (our "electric-wscc" graph).

		data set		
measurement	***Ref. [Wa 99]***	***"clean1"***	***"clean0"***	***"raw"***
Order	4,941	8,381	12,202	12,226
Average Degree	2.67	2.6	2.41	2.4
Char. Path Length	18.7	19.2	29.3	29.3
Clustering Coefficient	0.08	0.17	0.06	0.06
Shortcut Fraction	0.79	0.76	0.9	0.9
Contraction Fraction	0.8	0.69	0.68	0.68

V. Analysis of Degraded Networks

We have studied extensively the effect of degrading infrastructure networks by removing selected vertices or edges and then measuring the properties of the resulting graph. The different types of degradation correspond to the different situations that infrastructures might face in the real world. Random deletions might occur as the result of equipment failures or a natural disaster (e.g., earthquake or hurricane). Degree-specific deletions might occur in situations where infrastructures are targeted by an unsophisticated adversary—a more sophisticated adversary might first consider attacking articulation points and bridges of high degree. Numerous additional scenarios are conceivable, but the six cases we consider span many of the possibilities. Table 34 lists the 87 degraded graphs that were measured for each infrastructure. Six types of degradation were considered:

1. *Random vertex deletion:* vertices were deleted from the graph at random so each vertex had an equal probability of being deleted.

2. *Random edge deletion:* edges were deleted from the graph at random so each edge had an equal probability of being deleted.

3. *Vertex deletion by degree:* vertices were deleted in order of decreasing degree; if several vertices had the same degree, then each of these had an equal probability of being deleted.

4. *Edge deletion by degree:* edges were deleted in order of decreasing total degree of their two vertices; if several edges had the same total degree, then each of these had an equal probability of being deleted.

5. *Vertex deletion for articulation points:* vertices were deleted by degree as described above, but articulation points were considered for deletion before any other vertices.

6. *Edge deletion for bridges:* edges were deleted by degree as described above, but bridges were considered for deletion before any other edges.

Table 34. Design for measurements of degraded networks. The entries in the table denote how many times each graph was measured. For degraded graphs, there are six types of deletions possible (i.e., random vertices, random edges, vertices by degree, edges by degree, articulation points, and bridges); in one case the measurements were repeated using different random-number seeds to determine where the deletions would occur.

	data set		
f	***"raw"***	***"clean0"***	***"clean1"***
0	1	1	1
0.005			6
0.010			6
0.015			6
0.020		6	3×6
0.025			6
0.030			6
0.035			6
0.040		6	6
0.045			6
0.050			6

It is also important to understand how our procedure for cleaning the graphs affects measurements in degraded networks. Figure 68 summarizes how some of the measurements differ between the “clean0” and “clean1” data sets. These affects mirror those seen in the intact networks.

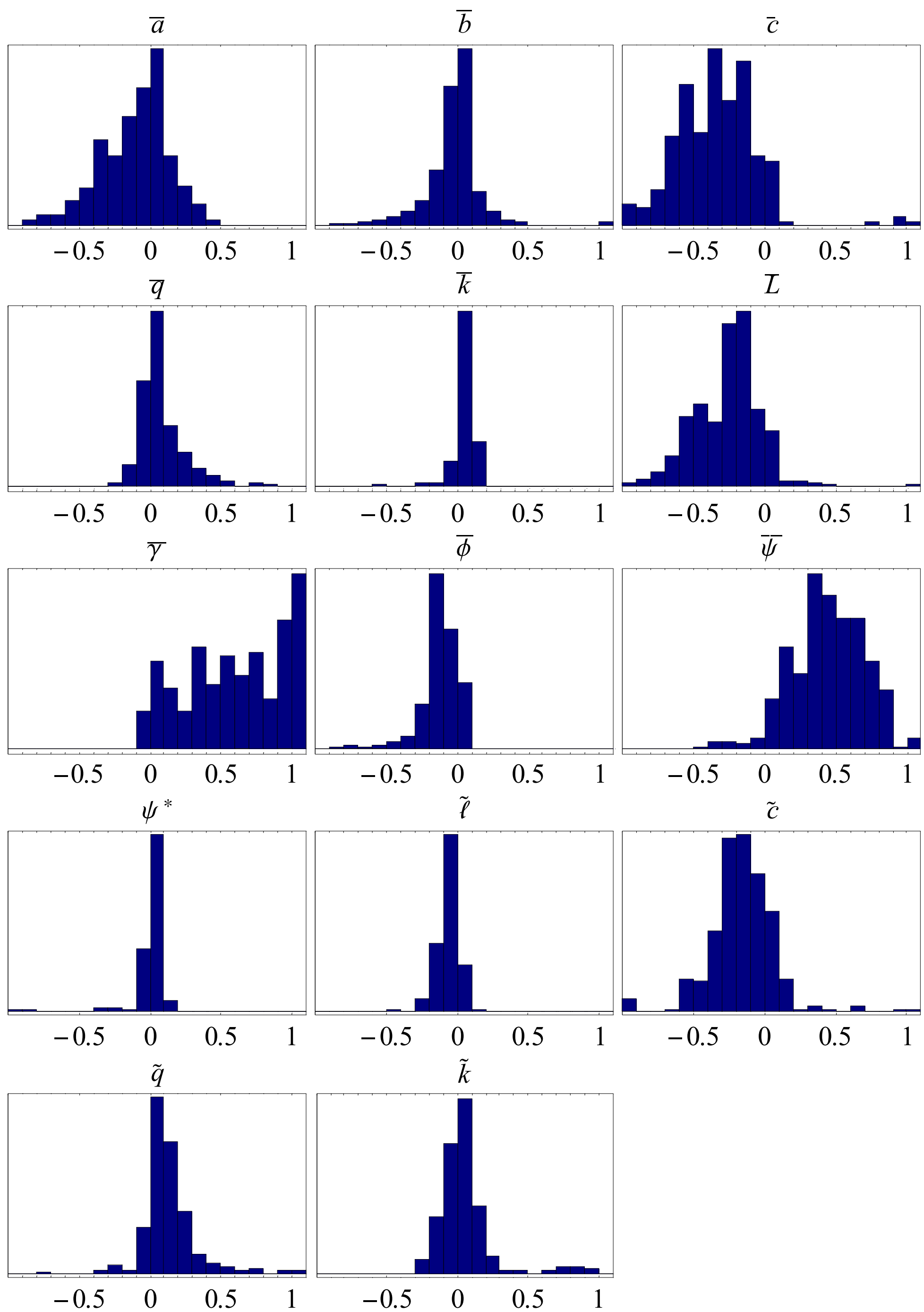

Figure 68. Histograms of the log ratio of the measurement in the “clean1” data set to the measurement of the “clean0” data set for degraded graphs.

A. Edge Deletions

Edge deletions generally affect a graph less that vertex deletions because deleting a vertex necessarily requires the edges connected to it to be deleted also. Figure 69 through Figure 81 show how this type of degradation affects the most interesting measurements for some of the key infrastructure graphs. The choice of random number seed makes little difference in the effects of degradation. We also notice that there are few cases where the three styles of degradation result in the same dependence of the measurement on fraction deleted. It is also clear that the nine example infrastructure networks considered respond to degradation differently. Here are some specific observations about individual cases:

- Using bridge deletion fragments the graphs much more that the other styles of deletion (Figure 69).

- Vertex chromatic number is insensitive to degradation (except for the "air-tickets" network) and style of deletion (Figure 70).

- Random edge deletion results in high variance for the measurements of vertex independence number, characteristic path length, and clustering coefficient (Figure 71, Figure 78, and Figure 79).

- Random deletions affect vertex independence number more than degree-based deletions (Figure 71).

- Different infrastructures show very different trends in lower vertex independence number—they increase, decrease, or stay relatively constant as a function of the fraction of edges deleted, and the three styles of deletion have different relative effects (Figure 72).

- Deleting bridges reduces the fraction of articulation points in a somewhat linear manner (Figure 73).

- Deleting bridges does not affect the average shortest cycle at vertices (Figure 75).

- The more tree-like a graph is, the more deletions reduce the average minimum cut between vertices (Figure 76).

- In tree-like networks, bridge deletions decrease the characteristic path length, but it increases in networks less like trees (Figure 78).

- Deletion of bridges generally increases clustering whereas random or degree deletions decrease it (Figure 79).

- The fraction of contractions decreases when bridges are deleted, but increases otherwise (Figure 80).

- The fraction of shortcuts changes differently with deletions for the "air-tickets" network compared to the other networks (Figure 81).

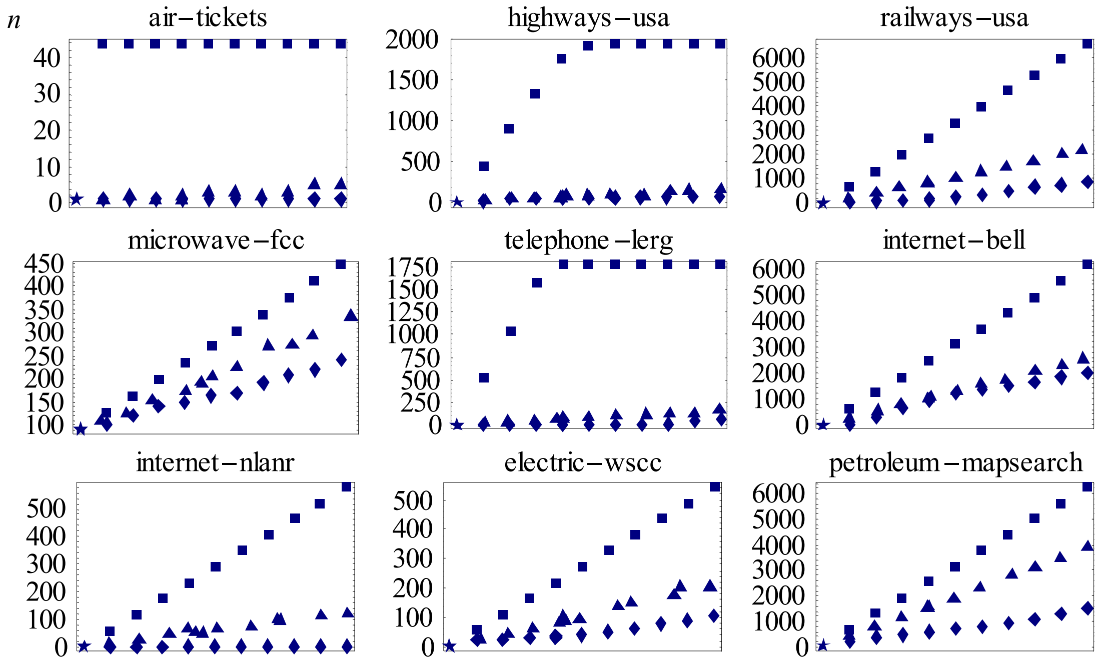

Figure 69. Measurements of the number of graph components (vertical axis) in degraded networks as a fraction of edges deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of bridges.

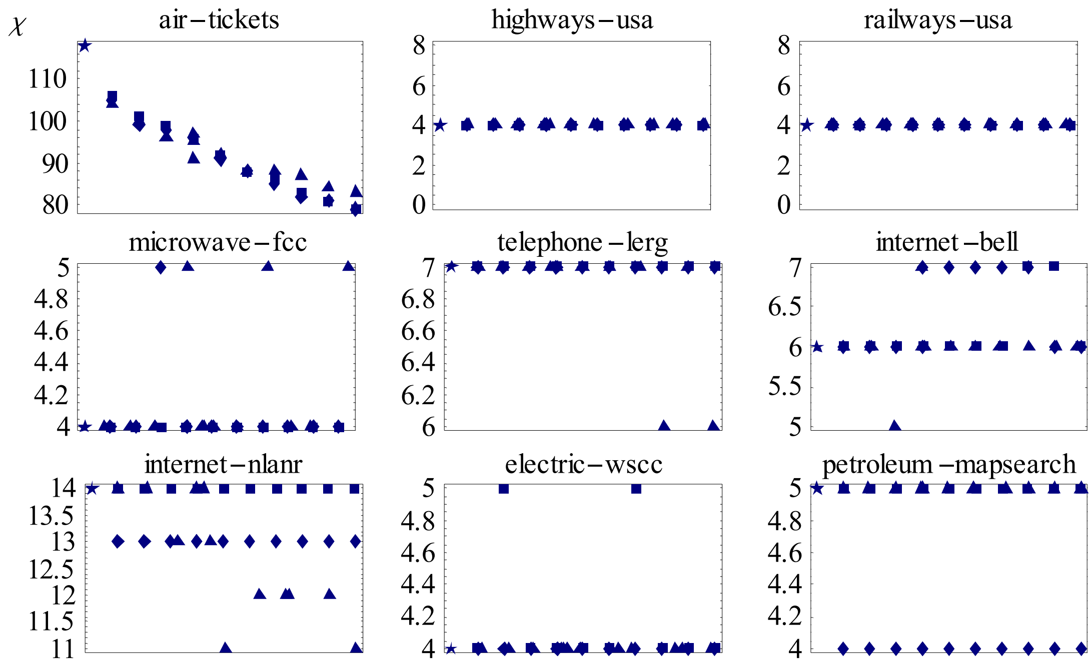

Figure 70. Measurements of the vertex chromatic number (vertical axis) in degraded networks as a fraction of edges deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of bridges.

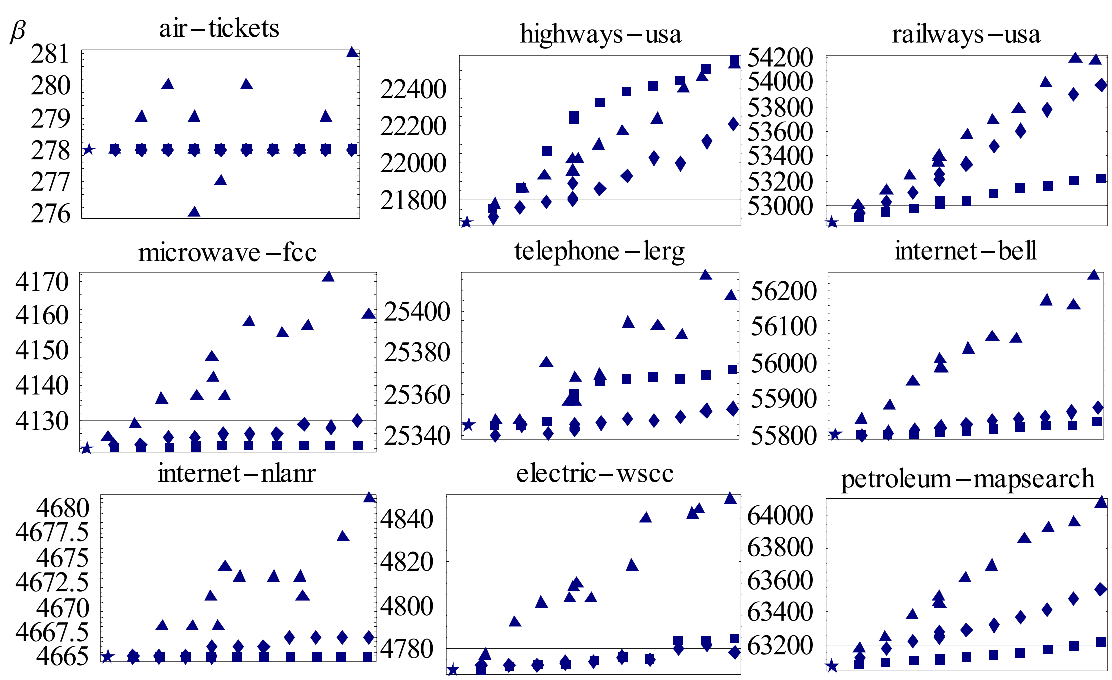

Figure 71. Measurements of the vertex independence number (vertical axis) in degraded networks as a fraction of edges deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of bridges.

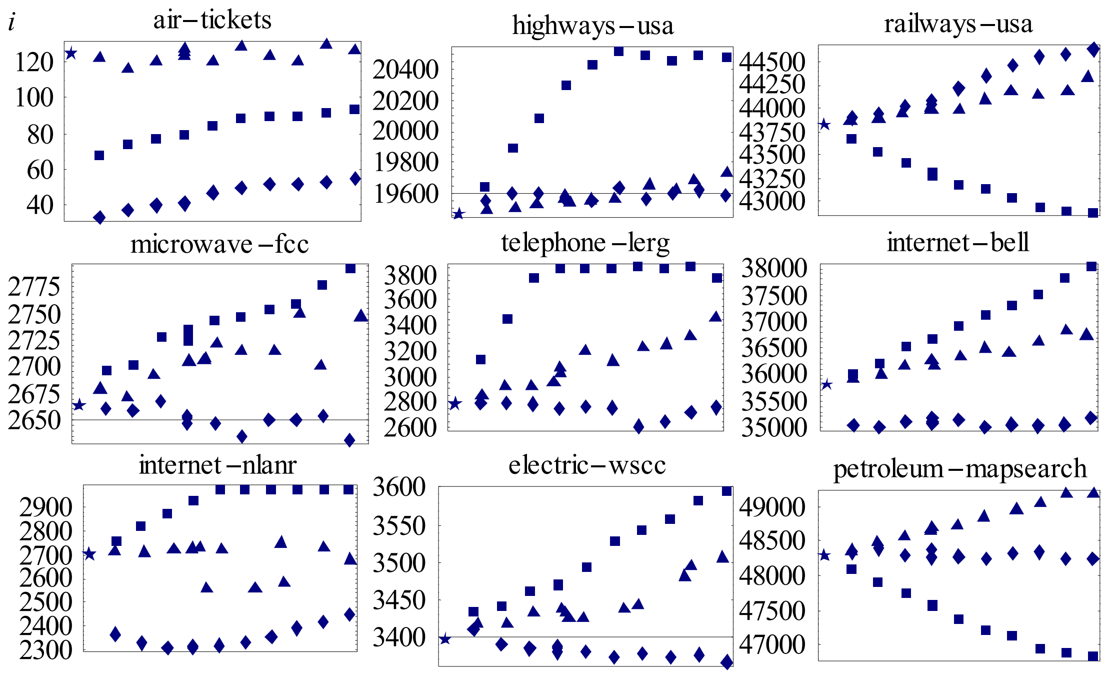

Figure 72. Measurements of the lower vertex independence number (vertical axis) in degraded networks as a fraction of edges deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of bridges.

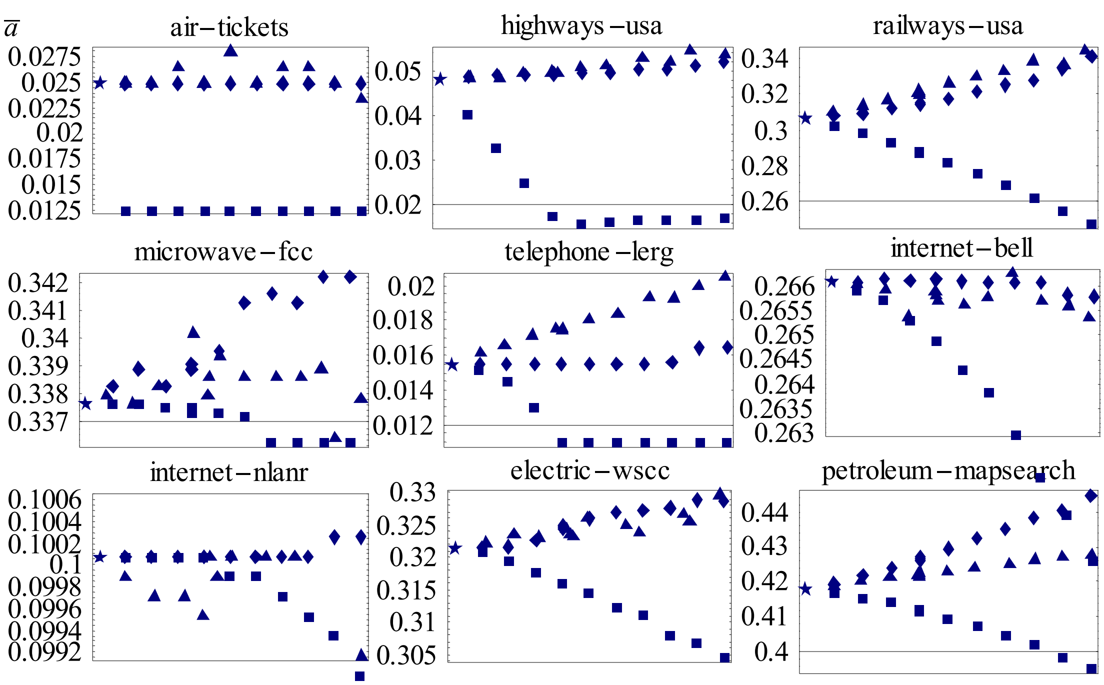

Figure 73. Measurements of the fraction of articulation points (vertical axis) in degraded networks as a fraction of edges deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of bridges.

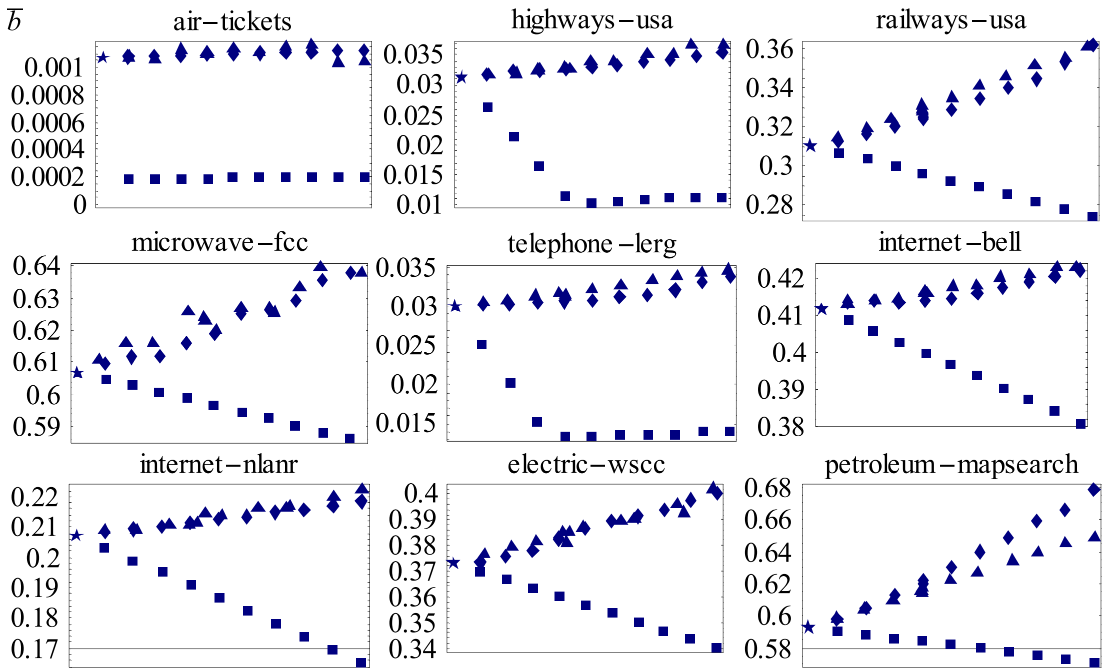

Figure 74. Measurements of the fraction of bridges (vertical axis) in degraded networks as a fraction of edges deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of bridges.

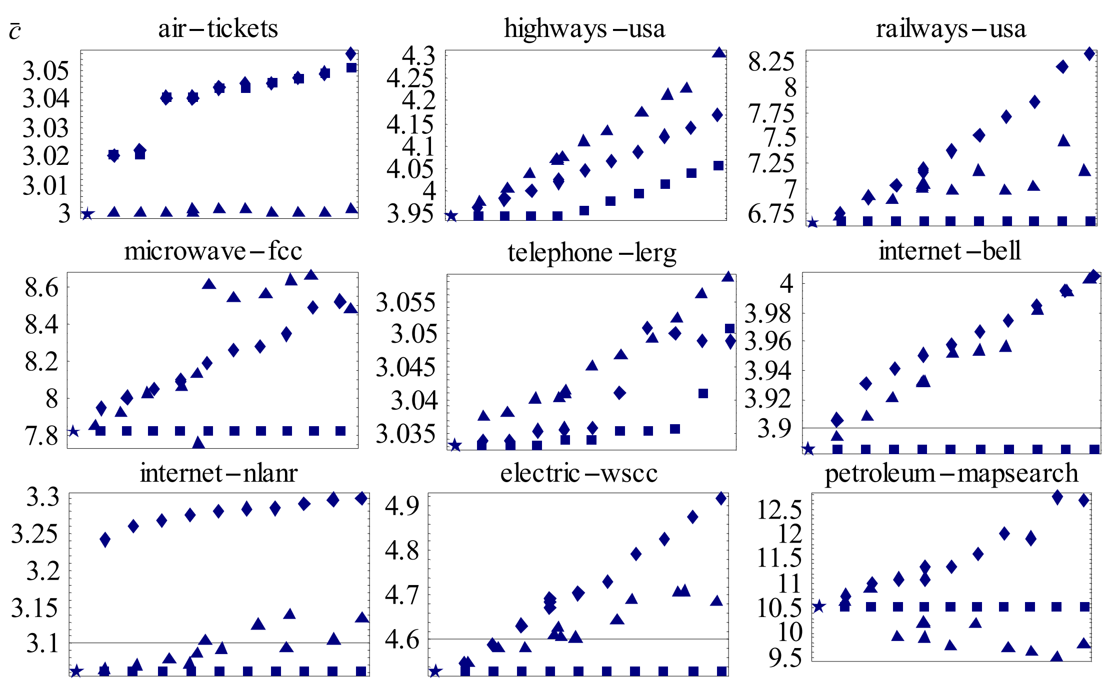

Figure 75. Measurements of the average shortest cycle (vertical axis) in degraded networks as a fraction of edges deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of bridges.

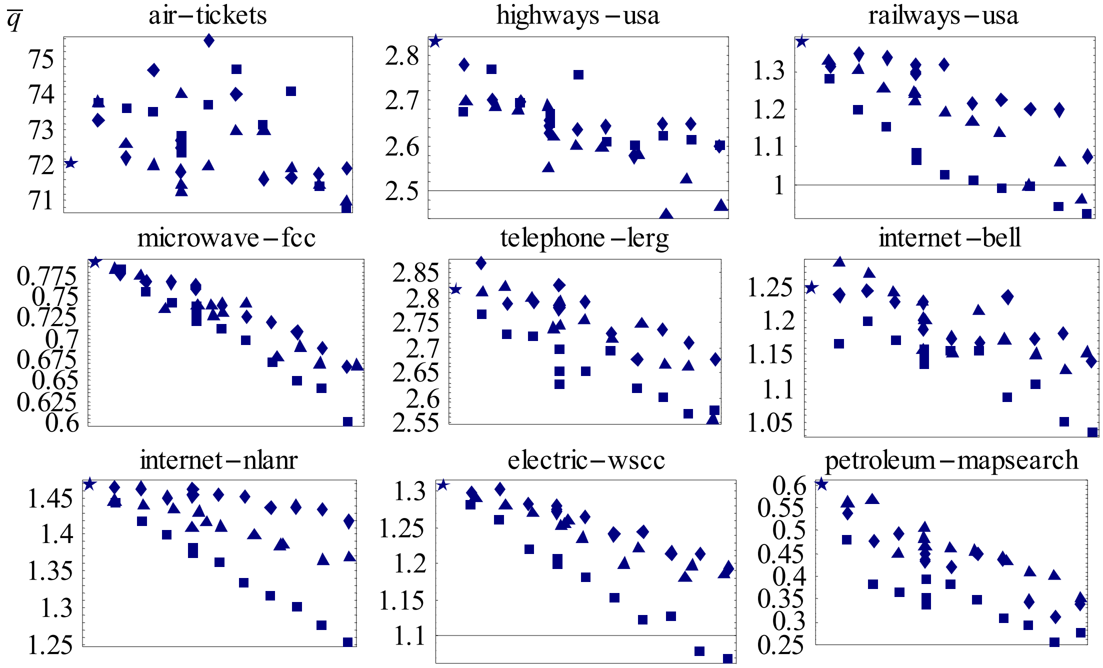

Figure 76. Measurements of the average minimum cut (vertical axis) in degraded networks as a fraction of edges deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of bridges.

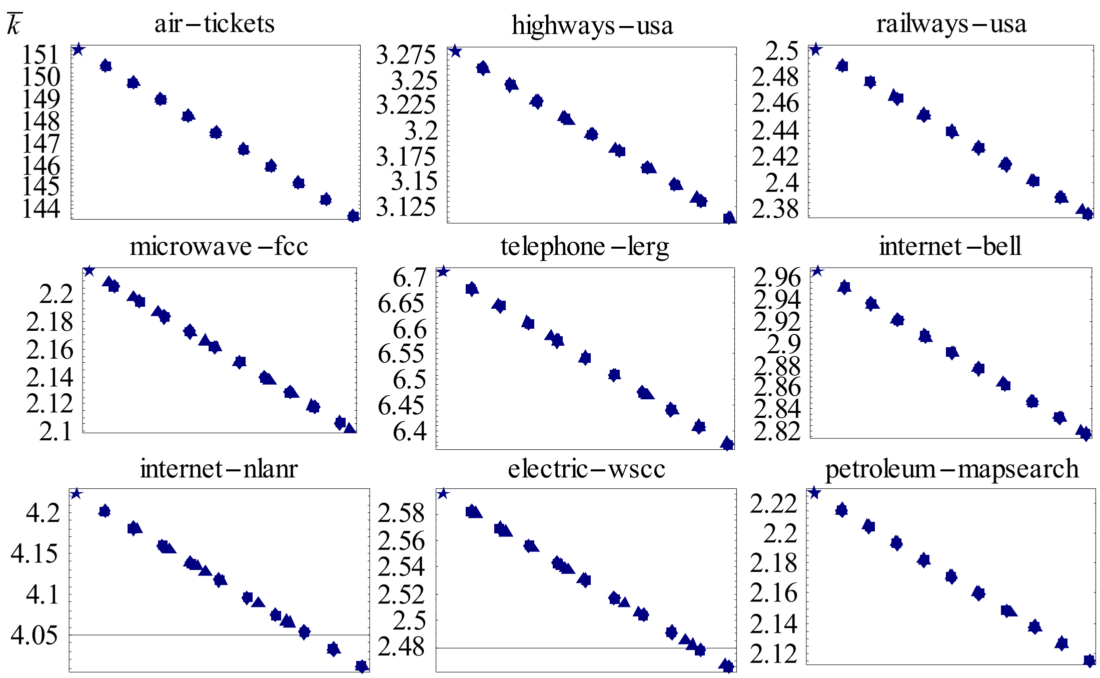

Figure 77. Measurements of the average degree (vertical axis) in degraded networks as a fraction of edges deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of bridges.

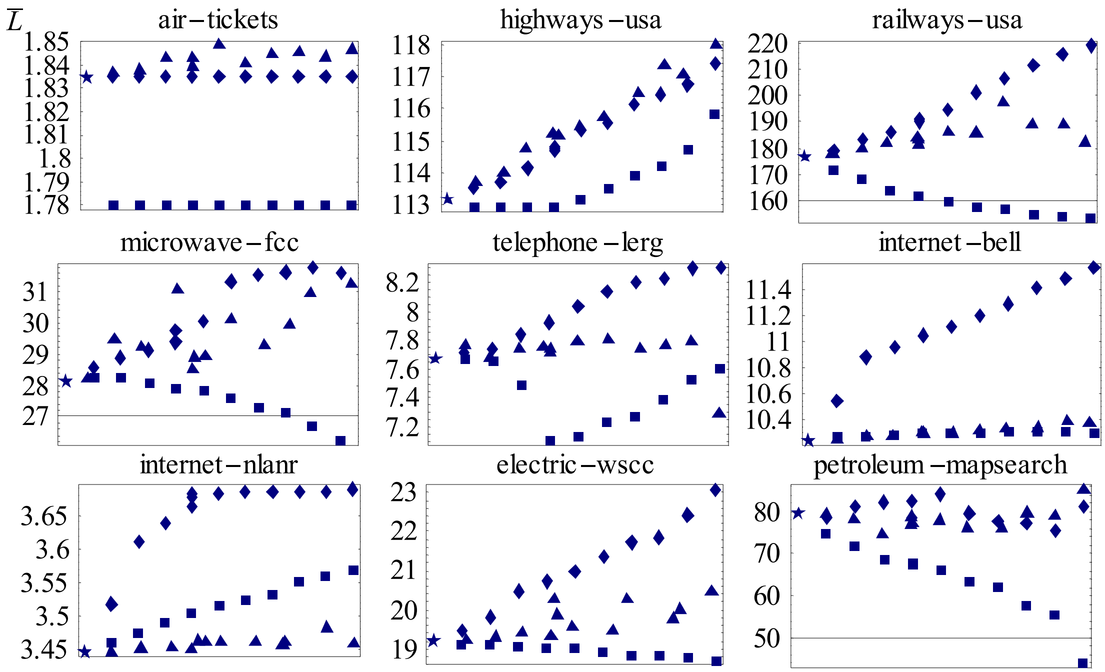

Figure 78. Measurements of the characteristic path length (vertical axis) in degraded networks as a fraction of edges deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of bridges.

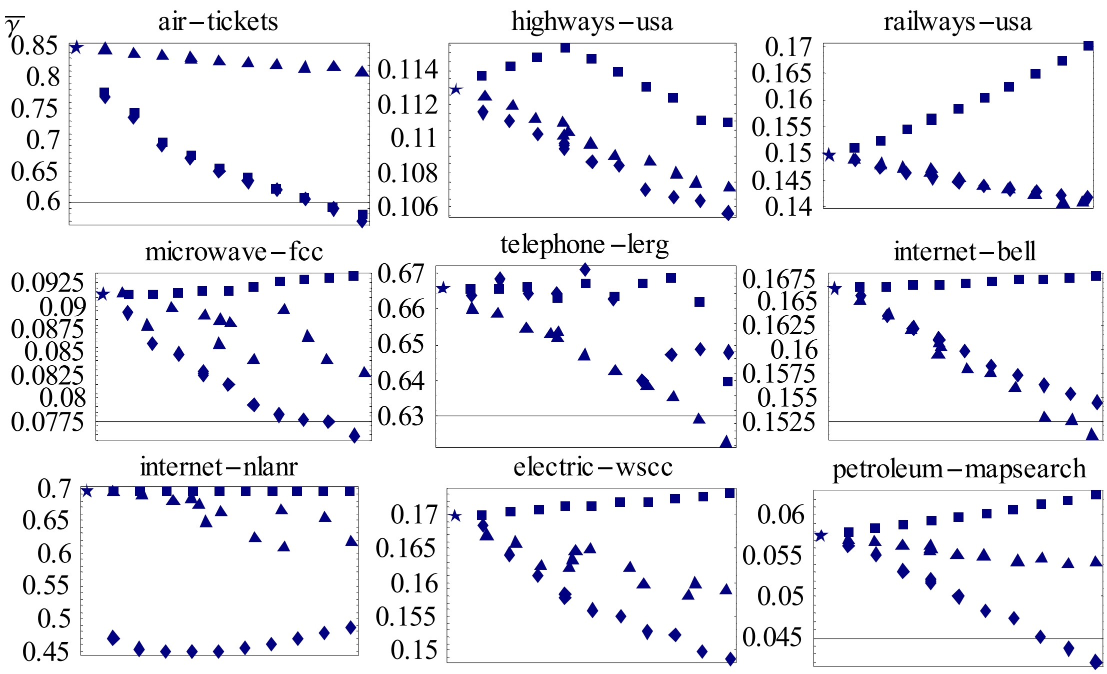

Figure 79. Measurements of the clustering coefficient (vertical axis) in degraded networks as a fraction of edges deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of bridges.

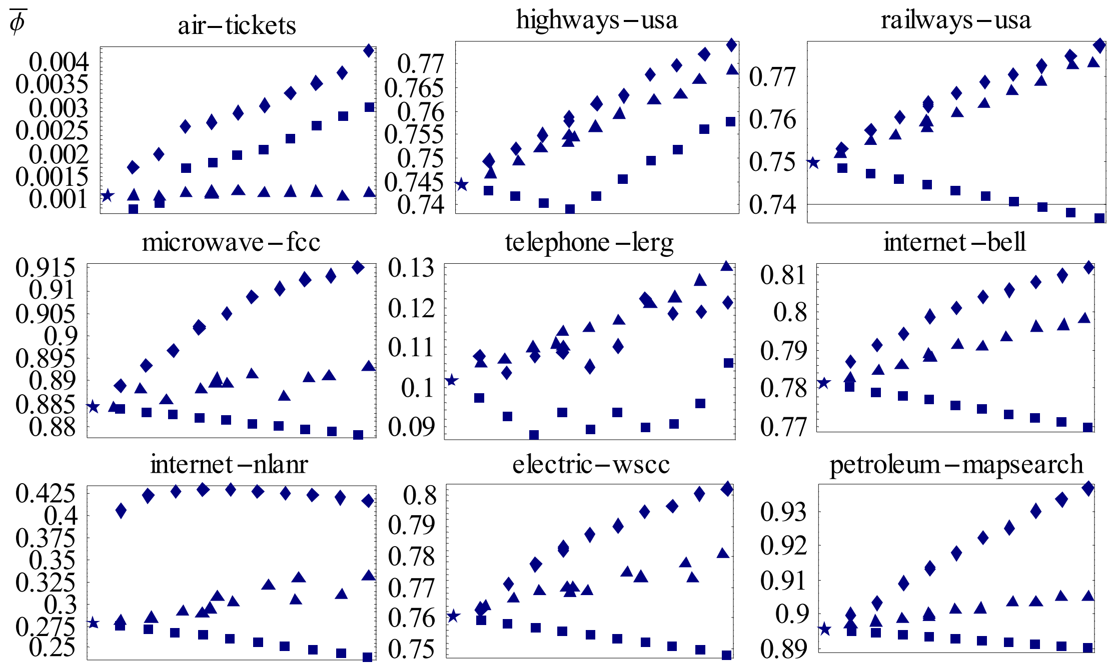

Figure 80. Measurements of the contraction fraction (vertical axis) in degraded networks as a fraction of edges deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of bridges.

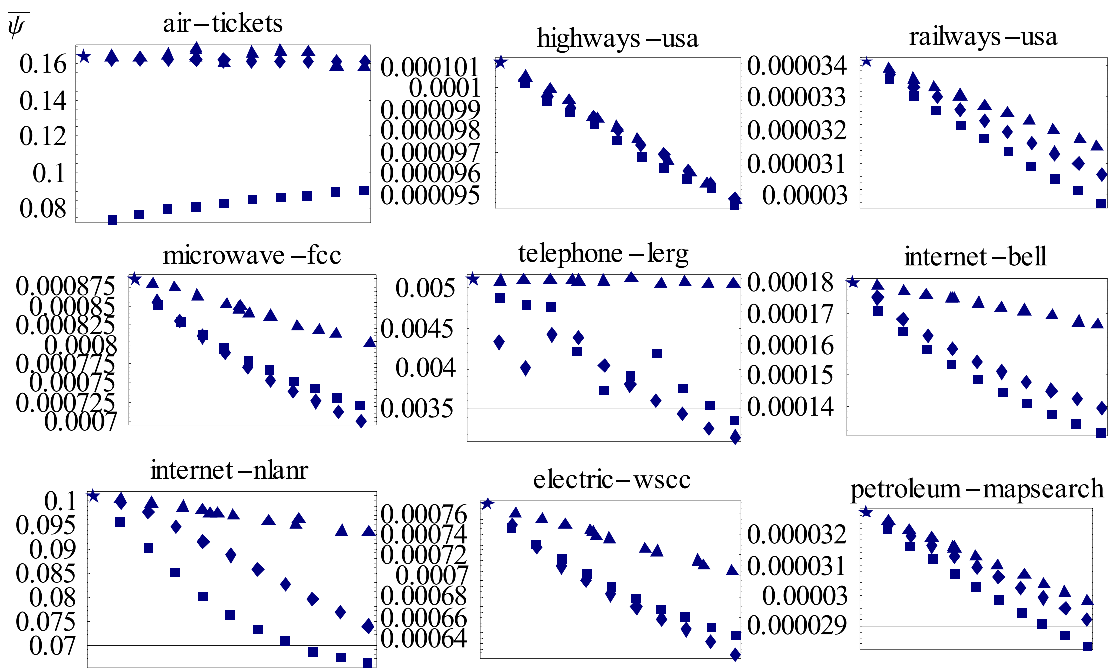

Figure 81. Measurements of the shortcut fraction (vertical axis) in degraded networks as a fraction of edges deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of bridges.

B. Vertex Deletion

We expect vertex deletions to degrade graphs similarly to edge deletions, but more severely because the edges associated with a vertex are also deleted. Figure 82 through Figure 94 demonstrate this for the more interesting measurements in some key infrastructures. We also notice again that the choice of random number seed makes little difference in the effects of degradation, that there are few cases where the three styles of degradation result in the same dependence of the measurement on fraction deleted, and that the nine example infrastructure networks considered respond to degradation differently. In most cases, the trends and relative effect of the styles of deletion are the same for vertices as for edges. Here are some specific exceptions of note:

- Vertex independence number decreases for vertex deletion whereas it increases for edge deletion (Figure 84).
- Deleting articulation points increases the fraction of bridges (Figure 87).
- Deleting vertices is more effective at reducing the size of average minimum cuts (Figure 89).
- Deleting vertices more strongly affects the shortcut fraction (Figure 94).

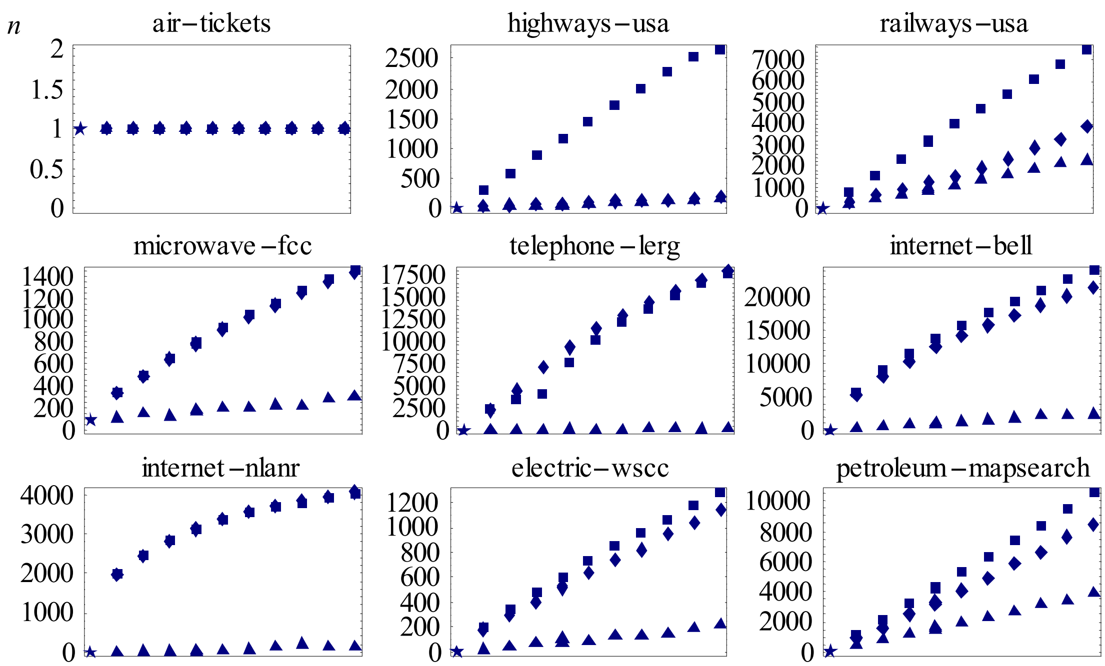

Figure 82. Measurements of the number of graph components (vertical axis) in degraded networks as a fraction of vertices deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of articulation points.

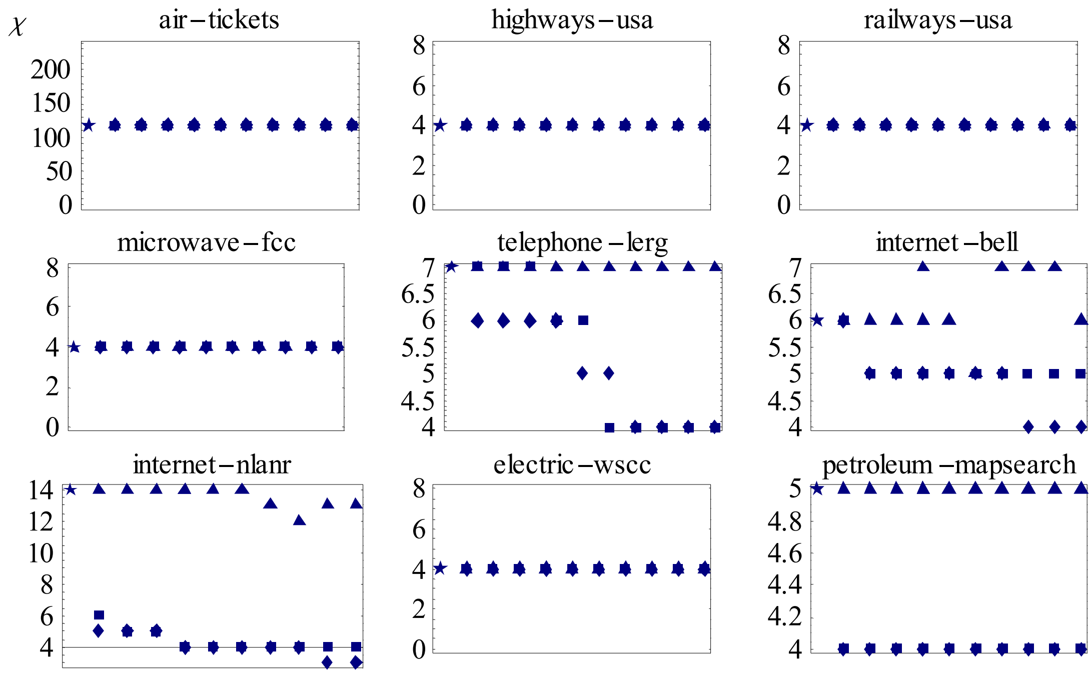

Figure 83. Measurements of the vertex chromatic number (vertical axis) in degraded networks as a fraction of vertices deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of articulation points.

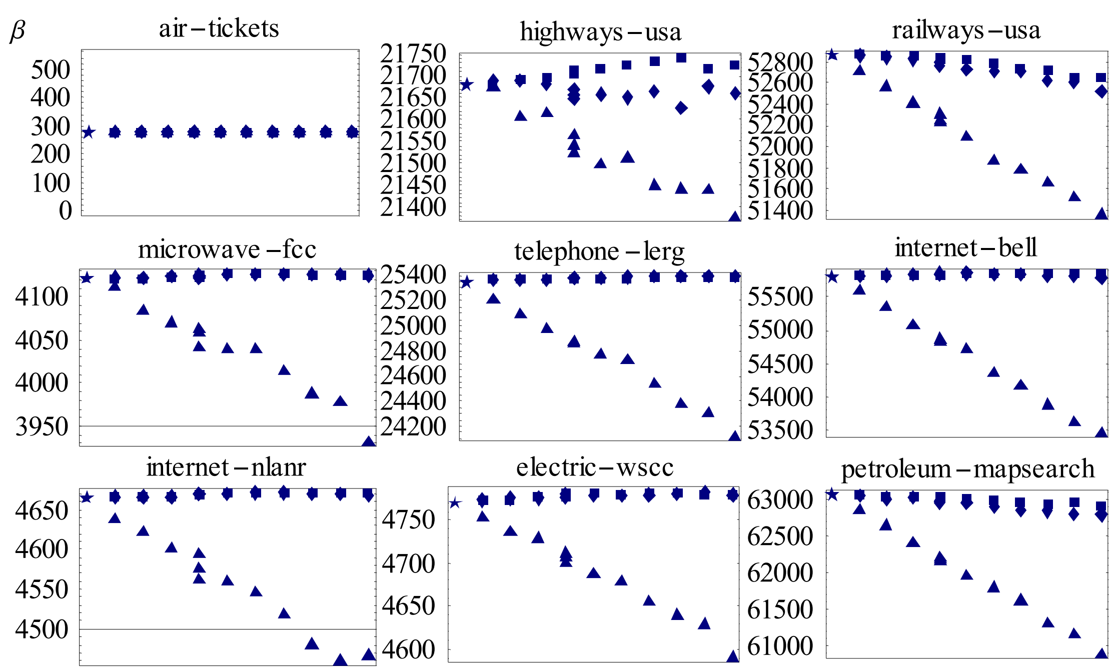

Figure 84. Measurements of the vertex independence number (vertical axis) in degraded networks as a fraction of vertices deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of articulation points.

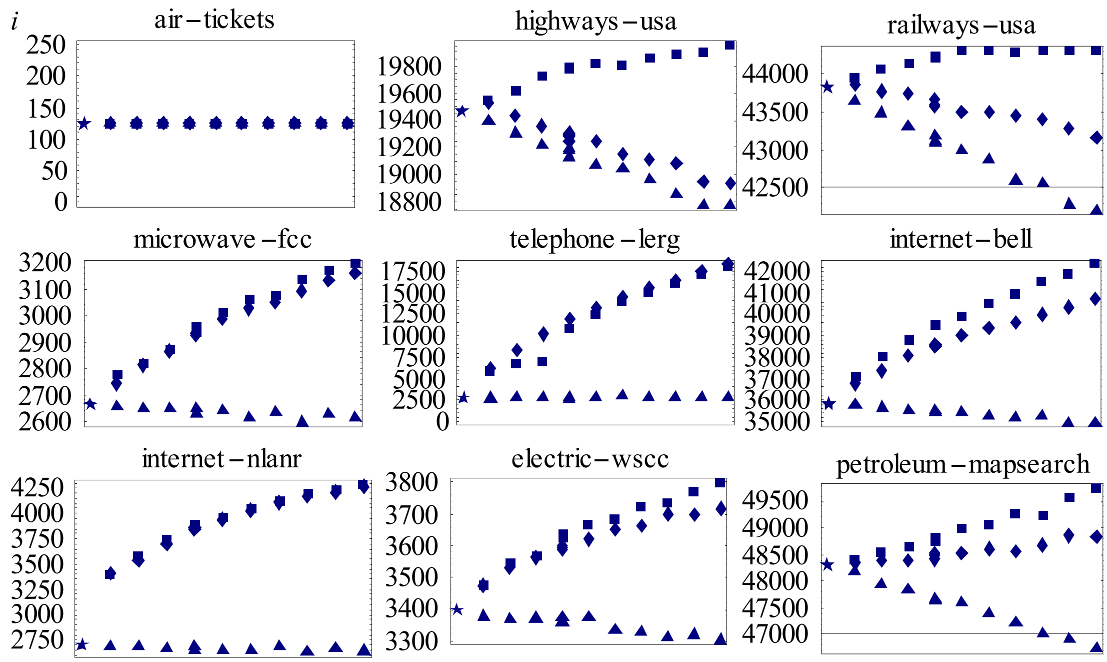

Figure 85. Measurements of the lower vertex independence number (vertical axis) in degraded networks as a fraction of vertices deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of articulation points.

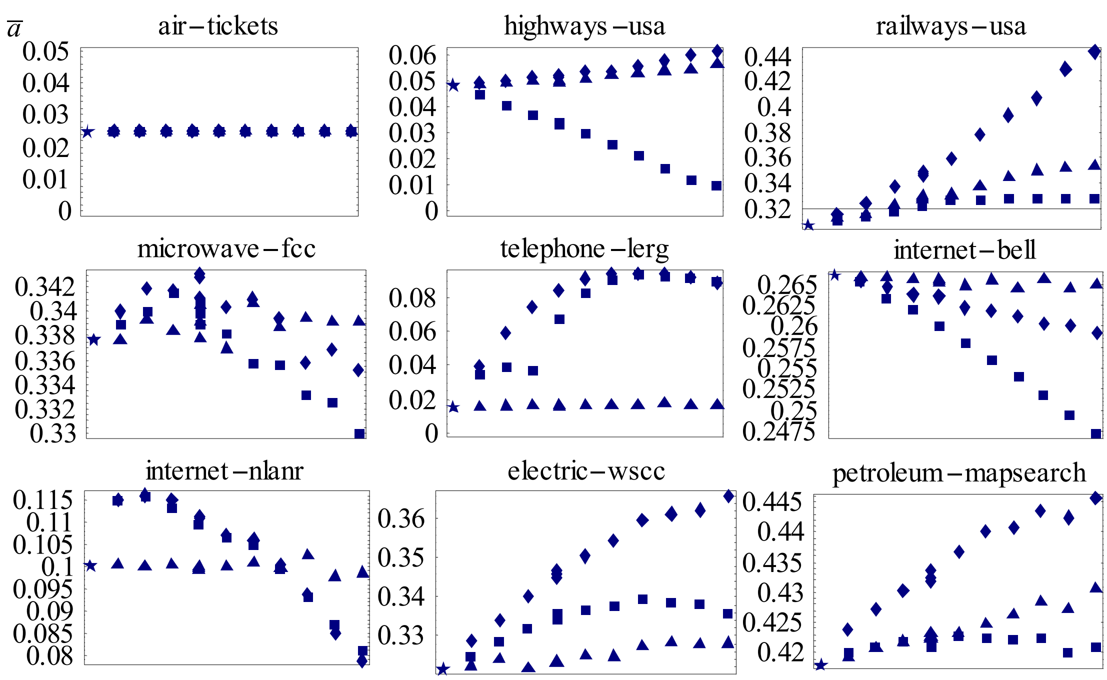

Figure 86. Measurements of the fraction of articulation points (vertical axis) in degraded networks as a fraction of vertices deleted (horizontal axis on a scale from 0% to 5%): the

star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of articulation points.

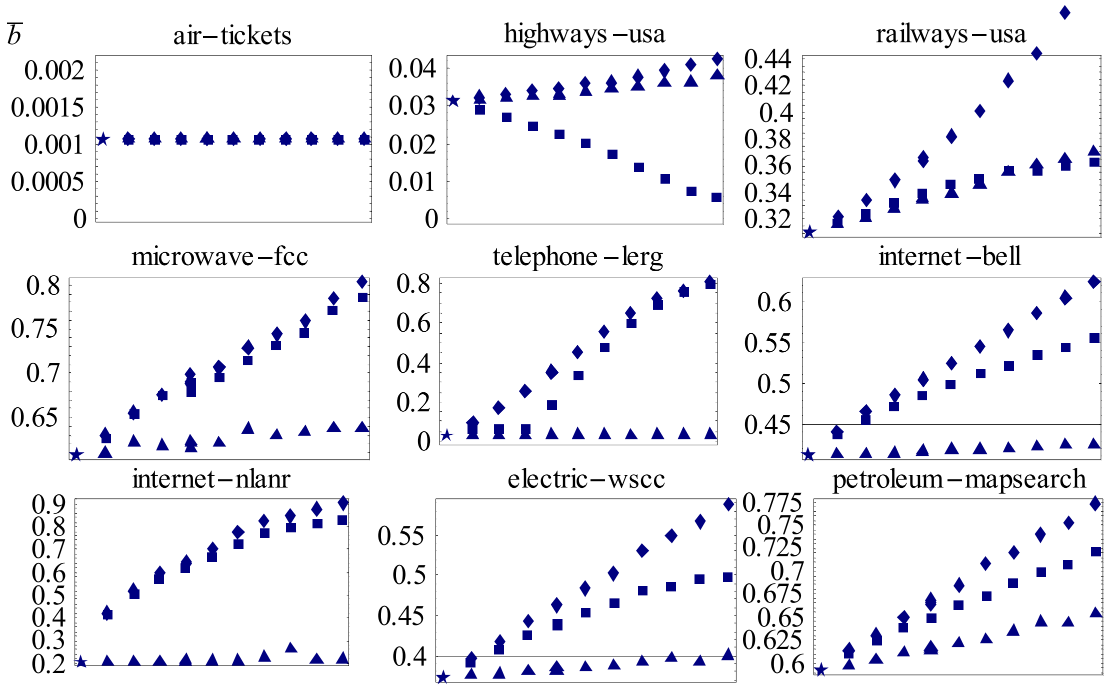

Figure 87. Measurements of the fraction of bridges (vertical axis) in degraded networks as a fraction of vertices deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of articulation points.

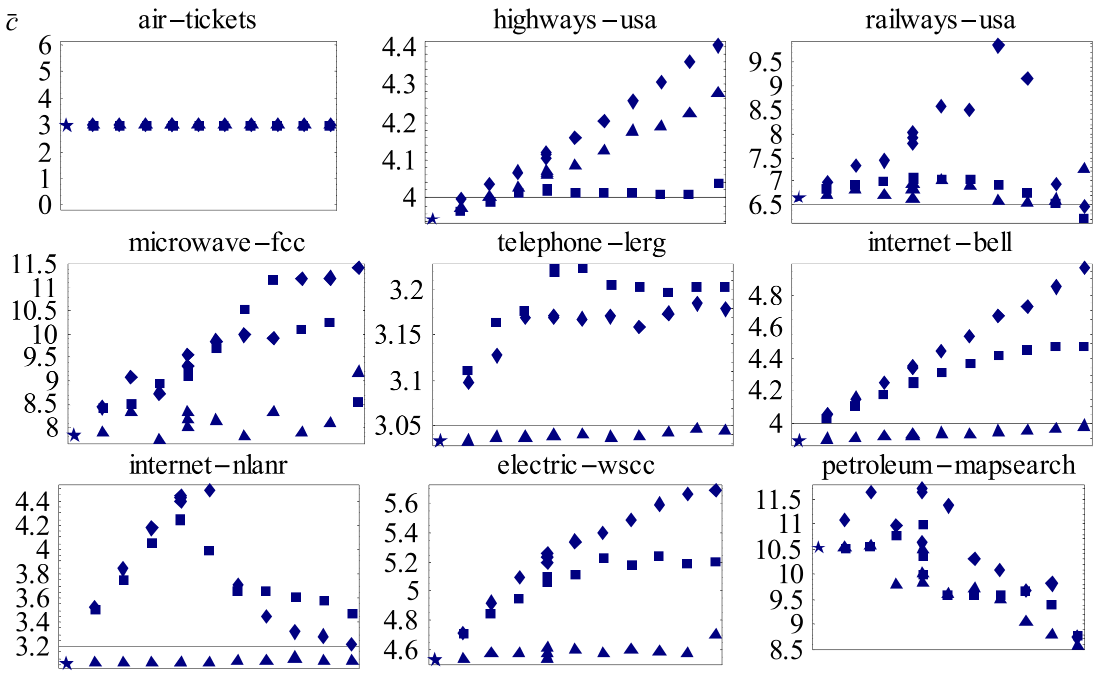

Figure 88. Measurements of the average shortest cycle (vertical axis) in degraded networks as a fraction of vertices deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of articulation points.

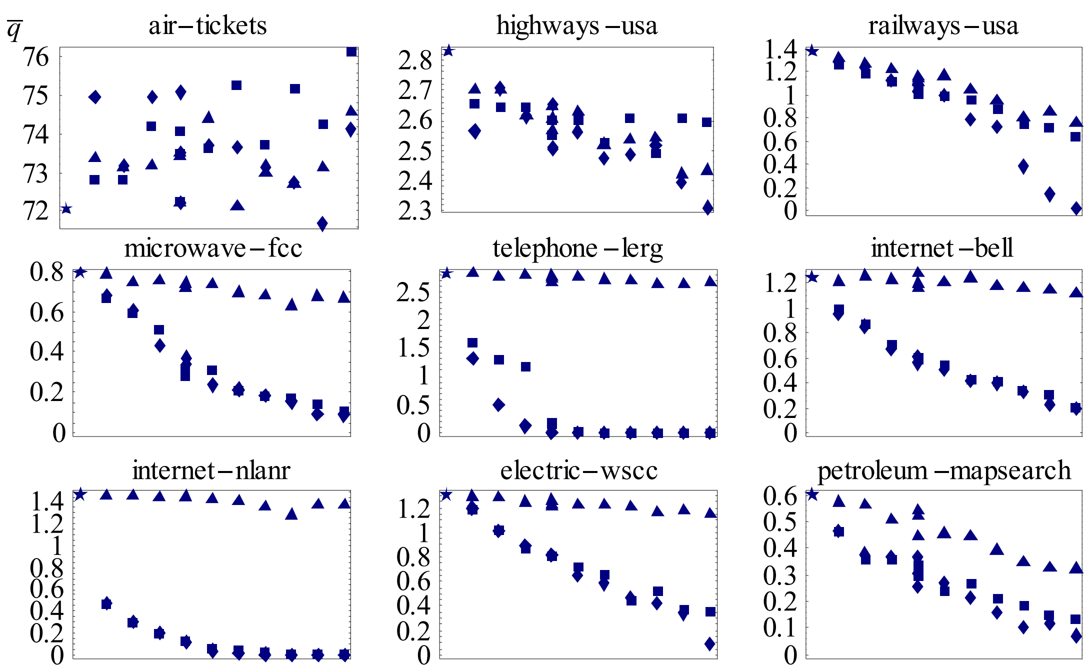

Figure 89. Measurements of the average minimum cut (vertical axis) in degraded networks as a fraction of vertices deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of articulation points.

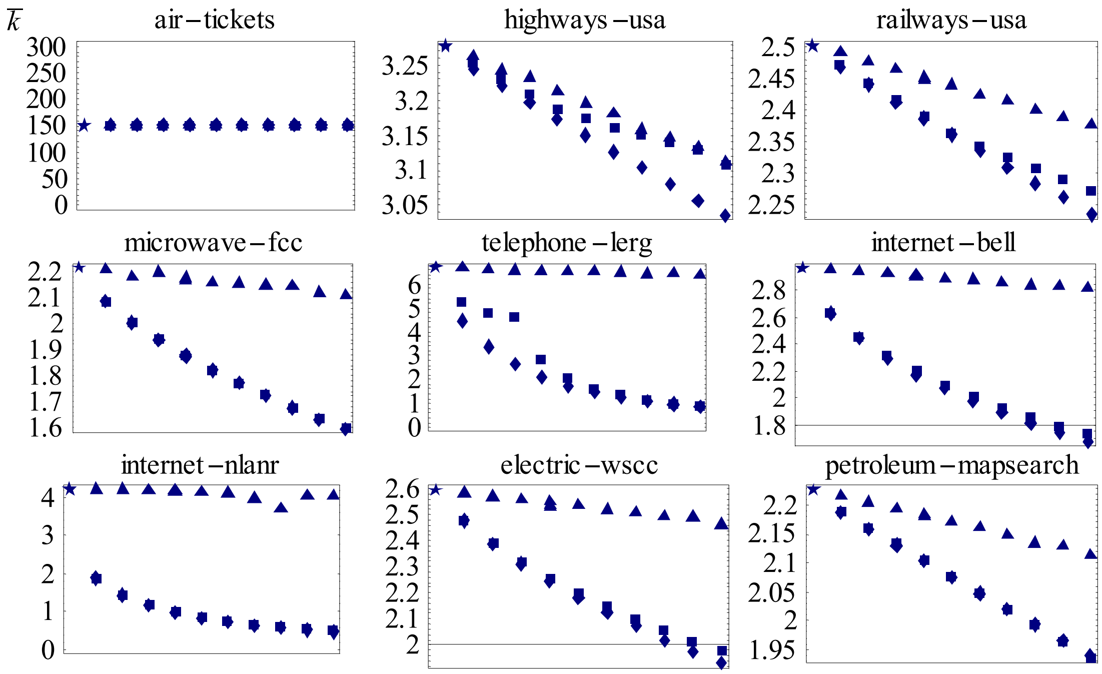

Figure 90. Measurements of the average degree (vertical axis) in degraded networks as a fraction of vertices deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of articulation points.

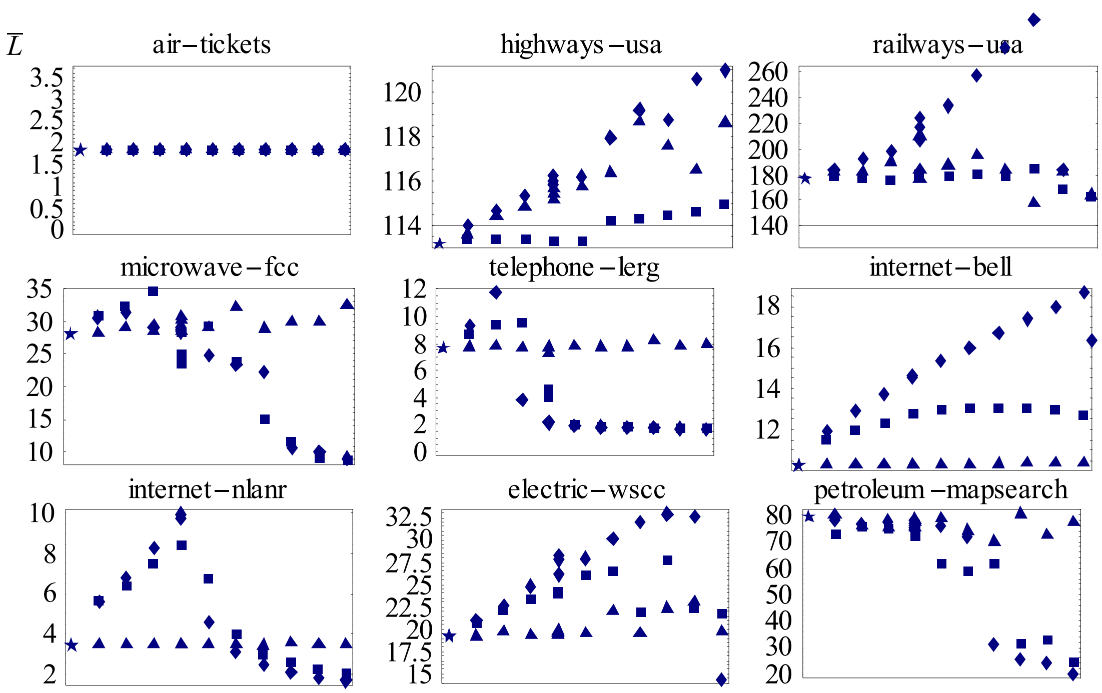

Figure 91. Measurements of the characteristic path length (vertical axis) in degraded networks as a fraction of vertices deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of articulation points.

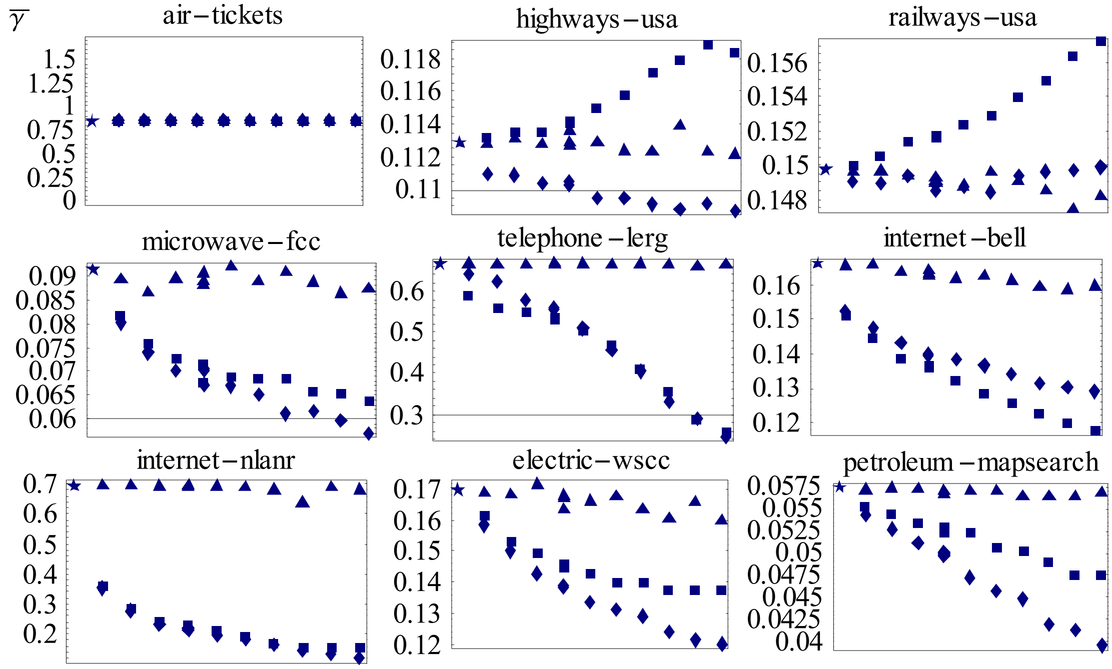

Figure 92. Measurements of the clustering coefficient (vertical axis) in degraded networks as a fraction of vertices deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of articulation points.

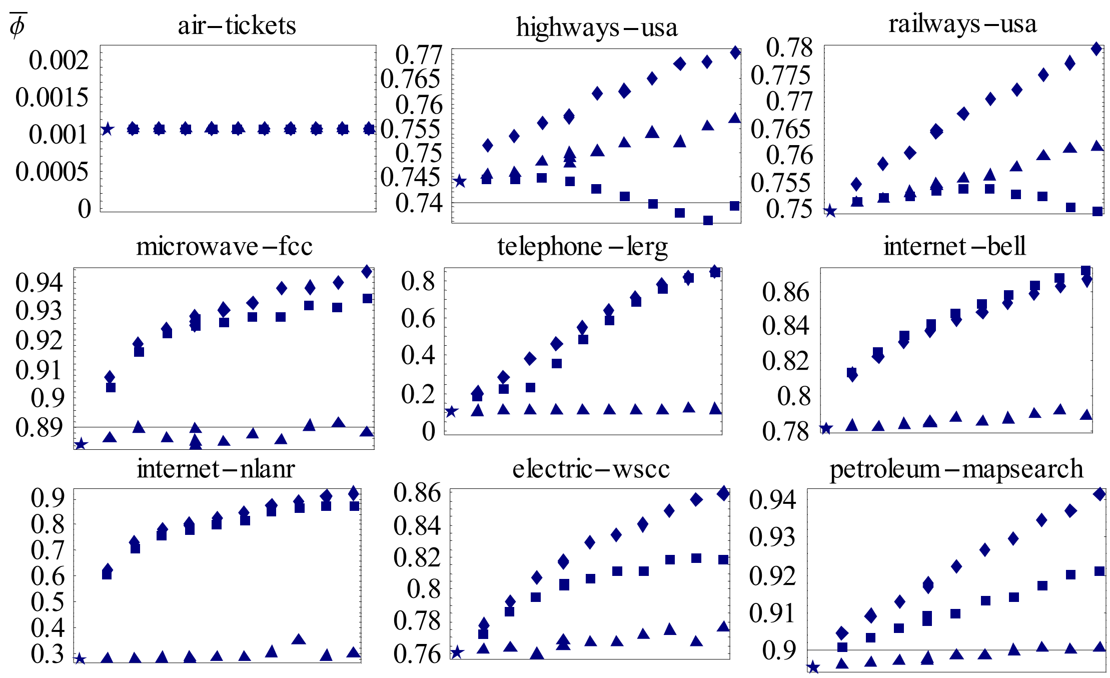

Figure 93. Measurements of the contraction fraction (vertical axis) in degraded networks as a fraction of vertices deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of articulation points.

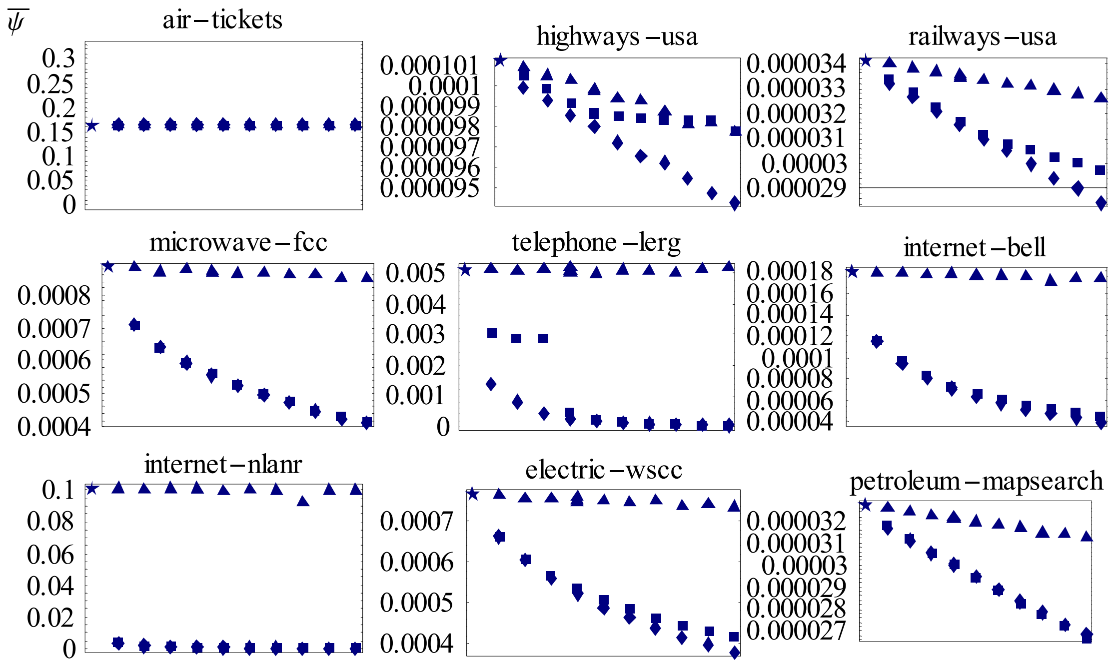

Figure 94. Measurements of the shortcut fraction (vertical axis) in degraded networks as a fraction of vertices deleted (horizontal axis on a scale from 0% to 5%): the star represents the intact network, the triangles represent deletion at random, the diamond represents deletion by degree, and the box represents deletion of articulation points.

VI. Conclusion

Based on our extensive reference database of infrastructure graphs, we have identified useful metrics for characterizing, clustering, and classifying infrastructure networks. We have developed practical algorithms for making these measurements that provide insight into the criticality of elements within the graphs and the overall robustness of the graphs. Our study of degraded networks illustrates how different types of degradation affect the character of the graphs.

We have continuing work ongoing in several areas:

- development of generative models for constructing and describing infrastructure graphs
- detection of suspicious graph elements using convex-hull-based data-depth computations
- construction of algorithms estimating bounds for NP-hard metrics related to global robustness (e.g., integrity and toughness)
- creation of heuristics for maximally-severe degradation of networks

We also plan to move beyond the generic and static analyses presented here to analyses that consider directed and weighted graphs where the attributes on the vertices and edges are treated. Further work will involve causality and the temporal structure of these graphs.

VII. Acknowledgements

We would like to thank Fred Roach for providing us with several of the data sources for the infrastructure network graphs. Richard Beckman gave us very helpful statistical advice. Thor Aspelund and Todd Graves are collaborating on generative methods. Meredith Blue provided useful comments on the manuscript. Ray Gordon sponsored this work, which was funded by the U.S. Department of Energy as part of LANL's Delphi LDRD effort.

VIII. References

[AGB 01] Thor Aspelund, Todd Graves, and Brian Bush, work in progress (2001).

[BBG 92] K. S. Bagga, L. W. Beineke, W. D. Goddard, M. J. Lipman, and R. E. Pippert, "A Survey of Integrity," *Discrete Applied Mathematics* **37/38**, 13-28 (1992).

[BBL 94] K. S. Bagga, L. W. Beineke, M. J. Lipman, R. E. Pippert, "Edge-Integrity: A Survey," *Discrete Mathematics* **124**, 3-12 (1994).

[Be 98] Bellcore, *Local Exchange Routing Guide* (1998).

[BM 76] J. A. Bondy and U.S.R. Murty, *Graph Theory with Applications,* (New York: Elsevier, 1976).

[BTS 98] U.S. Bureau of Transportation Statistics, *North American Transportation Atlas Data,* http://webcentral.bts.gov/ntda/nortad/ (1998).

[BTS 00] U.S. Bureau of Transportation Statistics, *National Transportation Atlas Data,* http://webcentral.bts.gov/ntda/ntatlas/desc.html (2000).

[CB 00] Bill Cheswick and Hal Burch, *Internet Mapping Project,* http://www.cs.bell-labs.com/who/ches/map/ (2000).

[CL 96] G. Chartrand and L. Lesniak, *Graphs and Digraphs,* 3rd edition (New York: Chapman & Hall, 1996).

[Ci 95] Robert Cimikowski, "On Heuristics for Determining the Thickness of a Graph," *Information Sciences* **85**, 87-98 (1995).

[De 92] Unclassified Department of Defense multimodal (air, sea, rail, road) network, provided by the Force Structure, Resource, and Assessment Directorate (J-8), Office of the Joint Chiefs of Staff (1992).

[DII 96] Yang Dai, Hiroshi Imai, Kazuo Iwano, Naoki Katoh, Keiji Ohtsuka, and Nobuhiko Yoshimura, "A New Unifying Heuristic Algorithm for the Undirected Minimum Cut Problems Using Minimum Range Cut Algorithms," *Discrete Applied Mathematics* **65**, 167-190 (1996).

[Dj 95] Hristo N. Djidjev, "A Linear Algorithm for the Maximal Planar Subgraph Problem" in Selim G. Akl, Frank Dehne, Jörge Rüdiger Sack, and Nicola Santoro (eds.), *Algorithms and Data Structures, 4th International Workshop, WADS '95, Kingston, Canada, August 16-18, 1995 Proceedings* (Berlin: Springer, 1995).

[EFG 81] Peter Eades, L. Foulds, and J. Giffin, "An Efficient Heuristic for Identifying a Maximum Weight Planar Subgraph" in Elizabeth J. Billington, Sheila Oates-Williams, and Anne Penfold Street (eds.), *Combinatorial Mathematics IX, Proceedings, Brisbande, Australia, 1981* (Berlin: Springer, 1981).

[EIA 99] Energy Information Administration, *Historical Data,* http://www.eia.doe.gov/neic/historic/historic.htm (1999).

[Fe 99] Federal Energy Regulatory Commission, *Records & Information Management System (RIMS),* http://rimsweb1.ferc.fed.us/rims.q?rp2~intro (1999).

[FHO 96] Gerd H. Fricke, Michael A. Henning, Ortrud R. Oellermann, Henda C. Swart, "An Algorithm to Find Two Distance Domination Parameters in a Graph," *Discrete Applied Mathematics* **68**, 85-91 (1996).

[FS 00] Christina Files and Patricia Saavedra, *Infrastructure Network Analysis using Fuzzy Logic,* Los Alamos National Laboratory Report LA-UR-01-285 (2000).

[Gi 85] Alan Gibbons, *Algorithmic Graph Theory* (New York: Cambridge, 1985).

[Gr 00] J. Grenier, personal communication to J. F. Roach (2000).

[Ha 98] Jochen Harant, "A Lower Bound on the Independence Number of a Graph," *Discrete Mathematics* **188**, 239-243 (1998).

[Ho 97] Dorit S. Hochbaum, "Approximating Covering and Packing Problems: Set Cover, Vertex Cover, Independent Set, and Related Problems" in Dorit S. Hochmaum (ed.), *Approximation Algorithms for NP-Hard Problems* (Boston: PWS Publishing, 1997).

[IAA 01] Infrastructure Assurance and Analysis Project (IAAP), Los Alamos National Laboratory, http://www.lanl.gov/orgs/d/d4//infra/assure.html (2001).

[KI 91] Naoki Katoh and Kazuo Iwano, "Efficient Algorithms for the Minimum Range Cut Problems" in F. Dehne, J.-R. Sack, and N. Santoro (eds.), *Algorithms and Data Structures, 2nd Workshop, WADS '91, Ottawa, Canada, August 14-16, 1991 Proceedings* (Berlin: Springer, 1991).

[KI 94] Naoki Katoh and Kazuo Iwano, "Efficient Algorithms for Minimum Range Cut Problems," *Networks* **24**, 395-407 (1994).

[KV 00] Michael Krivelevich and Van H. Vu, "Approximating the Independence Number and the Chromatic Number in Expected Polynomial Time," *LNCS* **1853**, 13-24 (2000).

[Le 92] Janny Leung, "A New Graph-Theoretic Heuristic for Facility Layout," *Management Science* **38**, 594-605 (1992).

[LPS 99] Regina Y. Liu, Jesse M. Parelius, and Kesar Singh, "Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference," *Annals of Statistics* **27**, 783-858 (1999).

[MOS 98] Petra Mutzel, Thomas Odenthal, Mark Scharbrodt, "The Thickness of Graphs: A Survey," *Graphs and Combinatorics* **14**, 59-73 (1998).

[Na 00] National Laboratory for Applied Network Research, *Autonomous System Information,* http://moat.nlanr.net/AS/ (2000).

[Oa 97] Oak Ridge National Laboratory's Multimodal Commodity Flow Survey 97 (CFS 97) network consisting of national highway and rail links and nodes, provided by Oak Ridge's Center for Transportation Analysis, directed by Michael Bronzini (1997).

[OAI 00] Office of Airline Information, Bureau of Transportation Statistics, *Origin & Destination City Pair Summary (Data Bank 6) magnetic tape, 4th Quarter 2000* (2000).

[Pe 95] PerCon Corporation, *Professional Series FCC Frequency Database,* http://www.perconcorp.com/products/pro/index.html (1995).

[Pe 98] Pennwell, *MapSearch Pipeline, Facility & Interconnect Inventory,* http://www.mapsearch.com/pipelineFacility.cfm (1998).

[PM 00] John Punin and Mukkai Krishnamoorthy, *DTD for XGMML 1.0,* Computer Science Department, Rensselaer Polytechnic Institute, http://www.cs.rpi.edu/~puninj/XGMML/ (2000).

[RR 99] Celso C. Ribeiro and Mauricio G. Resende, "Algorithm 797: Fortran Subroutines for Approximate Solution of Graph Planarization Problems Using GRASP," *ACM Transactions on Mathematical Software* **25**, 341-352 (1999).

[SCB 01] Deborah F. Swayne, Di Cook, and Andreas Buja, *XGobi Software Package,* http://www.research.att.com/areas/stat/xgobi/ (2001).

[Se 92] Robert Sedgewick, *Algorithms in C++* (Reading, Massachusetts: Addison-Wesley, 1992).

[Sk 98] Steven S. Skiena, *The Algorithm Design Manual* (New York: Springer-Verlag, 1998)

[SMT 98] M. Satratzemi, K. G. Margaritis, and C. Tsouros, "An Algorithm for Prescribed Multiple Domination in Arbitrary Graphs," *Computers Math. Applic.* **35**, 109-115 (1998).

[SW 97] Mechthild Stoer and Frank Wagner, "A Simple Min-Cut Algorithm," *Journal of the ACM* **44**, 585-591 (1997).

[Ta 92] Roberto Tamassia, database of flights between US cities posted on worldwide web at http://www.cs.sunysb.edu/~algorith/biblio/testdata/airplane-data (1992).

[TRA 01] TRansportation ANalysis SIMulation System (TRANSIMS) Project, Los Alamos National Laboratory, http://transims.tsasa.lanl.gov (2001).

[TS 93] C. Tsouros and M. Satratzemi, "Tree Search Algorithms for the Dominating Vertex Set Problem," *Intern. J. Computer Math.* **47**, 127-133 (1993).

[Un 01] C. Unal, Los Alamos National Laboratory, personal communication (2001).

[Wa 99] Duncan J. Watts, *Small Worlds: The Dynamics of Networks between Order and Randomness* (Princeton, New Jersey: Princeton University Press, 1999).

[Wo 01] Wolfram Research, *Mathematica Software Package,* http://www.wolfram.com/ (2001).

IX. Appendices

A. XGMML

The format we use for all of the infrastructure graphs is XGMML (eXtensible Graph Markup and Modeling Language) [PM 00]. The following excerpt from the DTD (document type definition) describes the graphs we have created.

```
<!-- Excerpt from DTD for XGMML 1.0 -->

<!-- Positive number type -->
<!ENTITY % number.type "NMTOKEN">

<!-- String type -->
<!ENTITY % string.type "CDATA">

<!-- Global Attributes -->
<!ENTITY % global-atts "id %number.type; #IMPLIED
```

```
                        name %string.type; #IMPLIED">

<!-- Safe Graph Attributes -->
<!ENTITY % graph-atts-safe "directed %boolean.type; '0' ">

<!-- Unsafe Graph Attributes (new attributes) -->
<!ENTITY % graph-atts-app-unsafe "Layout %string.type; #IMPLIED">

<!-- Graph Element -->
<!ELEMENT graph (att*,(node|edge)*,att*)>

<!-- Graph Attributes -->
<!ATTLIST graph
          %global-atts;
          %graph-atts-safe;
          %graph-atts-app-unsafe;>

<!-- Node Element -->
<!ELEMENT node (att*)>

<!-- Node Attributes -->
<!ATTLIST node
          %global-atts;>

<!-- Safe Edge Attributes (GML) -->
<!ENTITY % edge-atts-gml-safe "source %number.type; #REQUIRED
                               target %number.type; #REQUIRED">
<!-- Safe Edge Attributes (new attributes) -->
<!ENTITY % edge-atts-app-safe "weight %string.type; #IMPLIED">

<!-- Edge Element -->
<!ELEMENT edge (att*)>

<!-- Edge Attributes -->
<!ATTLIST edge
          %global-atts;
          %edge-atts-gml-safe;
          %edge-atts-app-safe;>

<!-- Value Attribute -->
<!ENTITY % attribute-value "value %string.type; #IMPLIED">

<!-- Att Element -->
<!ELEMENT att (#PCDATA | att)*>

<!-- Att Attributes -->
<!ATTLIST att
          %global-atts;
          %attribute-value;>
```

B. Graph Files

The INCCA archive DVD-R disc has a top-level "graphs" directory that contains subdirectories "raw," "clean0," and "clean1" containing the raw and preprocessed versions of the infrastructure graphs. The "test" subdirectory contains the graphs used to test the analysis software. The "raw/gobi" subdirectory contains XGobi-format [SCB 01] files for visualizing the layout of the graphs.

C. Results Files

The INCCA archive DVD-R disc has a top-level "analysis" directory that contains the subdirectories "raw," "clean0," and "clean1" containing the analysis results for the raw and preprocessed intact infrastructure graphs. There is also a "summary" subdirectory containing results for all of the runs on intact and degraded networks. These subdirectories, in turn, contain subdirectories with the results files in different formats—see Table 35 for details on these. References [Wo 01; SCB 01] provide specifications for the Mathematica and XGobi file formats.

Table 35. Files in analysis subdirectories.

subdirectory	*file*	*description*	*format*
text/	graph.txt	graph-level measurements for all the networks	tabbed text
	component.txt	component-level measurements for all the networks	tabbed text
text/cut/	*.txt	distribution of cut set size for each network	tabbed text
text/cycle/	*.txt	distribution of shortest cycle length for each network	tabbed text
text/degree/	*.txt	distribution of vertex degree for each network	tabbed text
text/distance/	*.txt	distribution of shortest path length for each network	tabbed text
text/edge/	*.txt	edge-level measurements for each network	tabbed text
text/pair/	*.txt	pairwise vertex measurements for each network	tabbed text
text/vertex/	*.txt	vertex-level measurements for each network	tabbed text
gobi/	graph.dat	graph-level measurements for all the networks	XGobi
	component.dat	component-level measurements for all the networks	XGobi
	cut.dat	distribution of cut set size for each network	XGobi
	cycle.dat	distribution of shortest cycle length for each network	XGobi
	degree.dat	distribution of vertex degree for each network	XGobi
	distance.dat	distribution of shortest path length for each network	XGobi
gobi/vertex/	*.dat	vertex-level measurements for each network	XGobi
math/	graph.mx	graph-level measurements for all the networks	Mathematica
	component.mx	component-level measurements for all the networks	Mathematica
math/cut/	*.mx	distribution of cut set size for each network	Mathematica
math/cycle/	*.mx	distribution of shortest cycle length for each network	Mathematica
math/degree/	*.mx	distribution of vertex degree for each network	Mathematica
math/distance/	*.mx	distribution of shortest path length for each network	Mathematica
math/edge/	*.mx	edge-level measurements for each network	Mathematica
math/pair/	*.mx	pairwise vertex measurements for each network	Mathematica
math/vertex/	*.mx	vertex-level measurements for each network	Mathematica

Additional directories contain results for degraded networks—see Table 36 for a full listing. The "comparison" subdirectory contains side-by-side comparisons that show the effect on preprocessing.

Table 36. Directories containing analysis results.

run subdirectory	*db*	*style*	*f*
raw	raw	no deletions	
clean0	clean0	no deletions	
clean0-edge020	clean0	edges randomly	2.0 %
clean0-edge020-cut	clean0	bridges first	2.0 %
clean0-edge020-deg	clean0	edges by degree	2.0 %
clean0-edge040	clean0	edges randomly	4.0 %
clean0-edge040-cut	clean0	bridges first	4.0 %
clean0-edge040-deg	clean0	edges by degree	4.0 %
clean0-node020	clean0	vertices randomly	2.0 %
clean0-node020-cut	clean0	articulation points first	2.0 %
clean0-node020-deg	clean0	vertices by degree	2.0 %
clean0-node040	clean0	vertices randomly	4.0 %
clean0-node040-cut	clean0	articulation points first	4.0 %
clean0-node040-deg	clean0	vertices by degree	4.0 %
clean1	clean1	no deletions	
clean1-edge005	clean1	edges randomly	0.5 %
clean1-edge005-cut	clean1	bridges first	0.5 %
clean1-edge005-deg	clean1	edges by degree	0.5 %
clean1-edge010	clean1	edges randomly	1.0 %
clean1-edge010-cut	clean1	bridges first	1.0 %
clean1-edge010-deg	clean1	edges by degree	1.0 %
clean1-edge015	clean1	edges randomly	1.5 %
clean1-edge015-cut	clean1	bridges first	1.5 %
clean1-edge015-deg	clean1	edges by degree	1.5 %
clean1-edge020	clean1	edges randomly	2.0 %
clean1-edge020-cut	clean1	bridges first	2.0 %
clean1-edge020-deg	clean1	edges by degree	2.0 %
clean1-edge020a	clean1	edges randomly	2.0 %
clean1-edge020a-cut	clean1	bridges first	2.0 %
clean1-edge020a-deg	clean1	edges by degree	2.0 %
clean1-edge020b	clean1	edges randomly	2.0 %
clean1-edge020b-cut	clean1	bridges first	2.0 %
clean1-edge020b-deg	clean1	edges by degree	2.0 %
clean1-edge025	clean1	edges randomly	2.5 %
clean1-edge025-cut	clean1	bridges first	2.5 %
clean1-edge025-deg	clean1	edges by degree	2.5 %
clean1-edge030	clean1	edges randomly	3.0 %
clean1-edge030-cut	clean1	bridges first	3.0 %
clean1-edge030-deg	clean1	edges by degree	3.0 %
clean1-edge035	clean1	edges randomly	3.5 %
clean1-edge035-cut	clean1	bridges first	3.5 %
clean1-edge035-deg	clean1	edges by degree	3.5 %
clean1-edge040	clean1	edges randomly	4.0 %
clean1-edge040-cut	clean1	bridges first	4.0 %
clean1-edge040-deg	clean1	edges by degree	4.0 %
clean1-edge045	clean1	edges randomly	4.5 %
clean1-edge045-cut	clean1	bridges first	4.5 %
clean1-edge045-deg	clean1	edges by degree	4.5 %
clean1-edge050	clean1	edges randomly	5.0 %
clean1-edge050-cut	clean1	bridges first	5.0 %
clean1-edge050-deg	clean1	edges by degree	5.0 %
clean1-node005	clean1	vertices randomly	0.5 %
clean1-node005-cut	clean1	articulation points first	0.5 %
clean1-node005-deg	clean1	vertices by degree	0.5 %
clean1-node010	clean1	vertices randomly	1.0 %
clean1-node010-cut	clean1	articulation points first	1.0 %
clean1-node010-deg	clean1	vertices by degree	1.0 %
clean1-node015	clean1	vertices randomly	1.5 %
clean1-node015-cut	clean1	articulation points first	1.5 %
clean1-node015-deg	clean1	vertices by degree	1.5 %
clean1-node020	clean1	vertices randomly	2.0 %

Table 36 (continued). Directories containing analysis results.

run subdirectory	*db*	*style*	*f*
clean1-node020-cut	clean1	articulation points first	2.0 %
clean1-node020-deg	clean1	vertices by degree	2.0 %
clean1-node020a	clean1	vertices randomly	2.0 %
clean1-node020a-cut	clean1	articulation points first	2.0 %
clean1-node020a-deg	clean1	vertices by degree	2.0 %
clean1-node020b	clean1	vertices randomly	2.0 %
clean1-node020b-cut	clean1	articulation points first	2.0 %
clean1-node020b-deg	clean1	vertices by degree	2.0 %
clean1-node025	clean1	vertices randomly	2.5 %
clean1-node025-cut	clean1	articulation points first	2.5 %
clean1-node025-deg	clean1	vertices by degree	2.5 %
clean1-node030	clean1	vertices randomly	3.0 %
clean1-node030-cut	clean1	articulation points first	3.0 %
clean1-node030-deg	clean1	vertices by degree	3.0 %
clean1-node035	clean1	vertices randomly	3.5 %
clean1-node035-cut	clean1	articulation points first	3.5 %
clean1-node035-deg	clean1	vertices by degree	3.5 %
clean1-node040	clean1	vertices randomly	4.0 %
clean1-node040-cut	clean1	articulation points first	4.0 %
clean1-node040-deg	clean1	vertices by degree	4.0 %
clean1-node045	clean1	vertices randomly	4.5 %
clean1-node045-cut	clean1	articulation points first	4.5 %
clean1-node045-deg	clean1	vertices by degree	4.5 %
clean1-node050	clean1	vertices randomly	5.0 %
clean1-node050-cut	clean1	articulation points first	5.0 %
clean1-node050-deg	clean1	vertices by degree	5.0 %

D. Software

The INCCA archive DVD-R disc contains the C++ and Java programming language software used to measure the graphs. It has the following subdirectory structure:

- **cpp/**. C++ source code and executable program for Linux-i86
- **java/src/**. Java source code
- **java/classes/**. Java class files

E. Selected Results

1. Graph Measurements “raw” Data Set

graph	type	subtype	name	V	E	n	delta	DDelta	r	rHat	d	dHat	g	kappa	kappa1	chi	gamma	gammaHat	beta	betaHat	i	iHat
1	1	1	air-sample	23	234	1	15	22	1	0.20851441	2	0.41702883	3	NA	NA	18	1	0.04347826	2	0.08695652	1	0.04347826
2	2	7	communications-bge	36	41	1	1	5	6	1	9	1.5	4	1	1	3	11	0.30555556	19	0.52777778	13	0.36111111
3	2	7	communications-gpc	401	99	303	0	4	1	0.04993762	13	0.64918902	3	1	1	3	332	0.82793017	353	0.88029925	334	0.83291771
4	2	7	communications-jepco	25	25	3	0	4	6	1.2	11	2.2	3	1	1	3	9	0.36	14	0.56	10	0.4
5	2	7	communications-pepco	38	44	7	0	5	4	0.64888568	8	1.29777137	3	1	1	3	17	0.44736842	21	0.55263158	18	0.47368421
6	2	7	communications-pge	475	322	207	0	15	1	0.04588315	36	1.65179328	3	1	1	3	289	0.60842105	371	0.78105263	309	0.65052632
7	2	7	communications-pnm	50	49	2	1	8	1	0.14142136	13	1.83847763	8	1	1	3	17	0.34	32	0.64	19	0.38
8	2	7	communications-sceg	28	26	2	0	7	7	1.32287566	13	2.45676907	NA	1	1	3	10	0.35714286	18	0.64285714	10	0.35714286
9	2	7	communications-sdge	51	45	11	0	8	5	0.70014004	10	1.40028008	5	1	1	3	22	0.43137255	35	0.68627451	27	0.52941176
10	2	7	communications-sepco	7	5	2	0	2	3	1.13389342	5	1.88982237	NA	1	1	2	3	0.42857143	4	0.57142857	4	0.57142857
11	2	7	communications-sngc	156	111	49	0	5	1	0.08006408	39	3.122499	3	1	1	3	85	0.54487179	107	0.68589744	91	0.58333333
12	2	7	communications-tpc	67	49	18	0	3	1	0.12216944	40	4.88677777	NA	1	1	3	34	0.50746269	43	0.64179104	37	0.55223881
13	2	7	communications-vepco	195	93	106	0	9	1	0.07161149	23	1.64706421	3	1	1	3	132	0.67692308	159	0.81538462	136	0.6974359
14	3	11	electric-ecar	4196	5889	2	1	13	1	0.01543769	37	0.57119446	3	1	1	5	1269	0.30243089	2112	0.50333651	1490	0.3551001
15	3	11	electric-ercot	4724	5669	7	0	16	20	0.29098798	40	0.58197596	3	1	1	5	1573	0.33298052	2578	0.54572396	1770	0.37468247
16	3	11	electric-frcc	2082	2605	4	0	12	20	0.43831833	40	0.87663666	3	1	1	4	688	0.33045149	1122	0.5389049	805	0.38664745
17	3	11	electric-maac	2836	3701	1	1	15	18	0.33800209	36	0.67600419	3	1	1	5	902	0.3180536	1539	0.54266573	1081	0.38117066
18	3	11	electric-main	3443	4261	9	0	16	1	0.01704243	51	0.86916387	3	1	1	5	1179	0.34243392	1833	0.53238455	1375	0.39936102
19	3	11	electric-mapp	4478	5449	12	0	12	1	0.01494369	48	0.71729729	3	1	1	4	1590	0.35506923	2351	0.52501117	1762	0.39347923
20	3	11	electric-nepool	2180	2674	6	0	17	17	0.3641	32	0.6853647	3	1	1	4	770	0.35321101	1195	0.54816514	880	0.40366972
21	3	11	electric-nypp	4117	5245	1	1	27	24	0.37404238	47	0.73249965	3	1	1	4	1342	0.32596551	2156	0.52368229	1522	0.36968667
22	3	11	electric-serc	5531	7201	4	0	23	1	0.01344616	55	0.73953862	3	1	1	4	1593	0.28801302	3019	0.54583258	1934	0.34966552
23	3	11	electric-spp	3730	4483	6	1	13	1	0.01637365	55	0.90055092	3	1	1	4	1249	0.33485255	2000	0.53619303	1453	0.38954424
24	3	11	electric-wscc	12226	14696	24	0	26	1	0.00904394	77	0.69638359	3	1	1	4	4118	0.33682316	6680	0.54637657	4826	0.39473254
25	4	13	fde	43	310	4	0	29	2	0.30499714	4	0.60999428	3	NA	NA	5	7	0.1627907	27	0.62790698	14	0.3255814
26	3	12	gas-florida	32	31	1	1	4	9	1.59099026	18	3.18198052	NA	1	1	2	11	0.34375	21	0.65625	11	0.34375
27	1	2	highways-canada	5232	4718	537	0	4	1	0.01382503	177	2.44703059	3	1	1	3	2071	0.39583333	2695	0.51509939	2383	0.45546636
28	1	2	highways-mexico	518	762	1	1	9	25	1.09843694	49	2.1529364	3	1	1	4	163	0.31467181	229	0.44208494	185	0.35714286
29	1	2	highways-usa	90415	124782	62	1	6	1	0.00332567	534	1.77591025	3	1	1	4	28694	0.31735885	41540	0.45943704	33474	0.37022618
30	2	10	internet-bell	122622	162133	605	0	130	1	0.00285572	43	0.12279601	3	1	1	6	NA	NA	72091	0.58791245	51001	0.41592047
31	2	10	internet-nlanr	6474	12572	1	1	1458	5	0.06214178	9	0.1118552	3	1	1	11	660	0.10194625	5411	0.83580476	2861	0.44192153
32	2	8	microwave-fcc	42313	10644	32456	0	32	1	0.00486142	174	0.84588698	3	1	1	4	35625	0.84193983	38102	0.90047976	36155	0.85446553
33	1	6	nts	101588	163932	476	0	51	1	0.00313746	206	0.64631765	3	1	1	6	28939	0.28486632	45803	0.45087018	33930	0.33399614
34	3	12	petroleum-mapsearch	160859	170818	3180	0	16	1	0.00249332	528	1.31647083	3	1	1	5	NA	NA	88594	0.55075563	69246	0.43047638
35	1	3	railways-canada	691	843	8	1	7	1	0.03804179	60	2.28250757	3	1	1	4	245	0.35455861	360	0.52098408	289	0.41823444
36	1	3	railways-mexico	179	207	2	1	5	1	0.07474351	47	3.51294494	3	1	1	3	60	0.33519553	90	0.5027933	72	0.40223464
37	1	3	railways-usa	133752	160168	304	0	9	1	0.00273432	1167	3.19095525	3	1	1	4	NA	NA	66020	0.4936001	55664	0.41617322
38	1	4	roads-dallas	9863	14750	4	1	8	1	0.01006921	111	1.11768253	3	1	1	4	3142	0.31856433	4548	0.46111731	3693	0.37442969
39	1	4	roads-portland	100852	126161	23	1	6	1	0.00314889	400	1.25955673	3	1	1	4	36502	0.3619363	51064	0.5063261	42772	0.42410661
40	2	9	telephone-lerg	33854	106253	1434	0	1133	1	0.00543494	29	0.15761335	3	1	1	7	2351	0.06944526	27849	0.82262067	3595	0.10619129
41	1	5	waterways-usa	6263	6870	6	1	20	1	0.01263598	185	2.33765557	3	1	1	4	2407	0.38432061	3344	0.53392943	2737	0.43701102
42	1	1	air-tickets	655	48624	13	0	515	2	0.07814647	4	0.15629293	3	1	1	118	22	0.03358779	290	0.44274809	137	0.20916031
43	3	12	gas-utah	160	159	1	1	8	22	1.73925271	44	3.47850543	NA	1	1	3	59	0.36875000	93	0.58125000	71	0.44375000

graph	name	alpha	alphaHat	aBar	bBar	cBar	qBar	kBar	LBar	LBarHat	gammaBar	phiBar	psiBar	psiStar	elTilde	cTilde	qTilde	kTilde
1	air-sample	20	0.86956522	0	0	3	19.36758893	20.3478	1.04545	0.21799139	0.937497	0	0	0	0.04282198	0	0.29518175	0.49142993
2	communications-bge	16	0.44444444	0.25	0.29268293	7.08333333	1.47301587	2.27778	4.28571	0.714285	0	1	0.101587	0.58715596	0.28047942	0.37808787	0.12524284	0.38704428
3	communications-gpc	48	0.11970075	0.13965087	0.96969697	3	0.01047382	0.493766	3.5119	0.17537592	0.0292398	0.969697	0.00138404	0.52857143	0.31853102	0.00735353	0.00515798	0.15237183
4	communications-jepco	11	0.44	0.08	0.08	9.95238095	1.55	2	4.45455	0.89091	0.0634921	0.88	0.106667	0.56140351	0.3503822	0.39727531	0.14721128	0.33256445
5	communications-pepco	17	0.44736842	0.10526316	0.09090909	4.32142857	1.38122333	2.31579	3.70968	0.60178956	0.0880952	0.818182	0.109531	0.6015625	0.27029622	0.44591451	0.20157795	0.46028066
6	communications-pge	104	0.21894737	0.18947368	0.47515528	4.71186441	0.36705308	1.35579	9.48689	0.43528837	0.118993	0.832298	0.00741728	0.70404722	0.27638066	0.15700422	0.06618379	0.2584987
7	communications-pnm	18	0.36	0.42	0.83673469	8	0.9077551	1.96	5.13043	0.72555237	0	1	0.0718367	0.64233577	0.29924055	0.11238939	0.06513864	0.32564983
8	communications-sceg	11	0.39285714	0.53571429	1	NA	0.92857143	1.85714	4.96154	0.93764293	0	1	0.121693	0.63888889	0.35198348	0	0.0433569	0.39829015
9	communications-sdge	17	0.33333333	0.25490196	0.51111111	5.83333333	0.77490196	1.76471	4.025	0.56361273	0	1	0.0768627	0.68531469	0.26384012	0.24599672	0.13978788	0.41949729
10	communications-sepco	3	0.42857143	0.57142857	1	NA	0.71428571	1.42857	2.2	0.83152184	0	1	0.190476	0.44444444	0.4380076	0	0.19650687	0.49113258
11	communications-sngc	51	0.32692308	0.33974359	0.55855856	33	0.54780811	1.42308	13.9434	1.11636545	0.0426357	0.945946	0.011249	0.54618474	0.36842951	0.16060489	0.09620015	0.27065152
12	communications-tpc	26	0.3880597	0.62686567	1	NA	0.53233831	1.46269	13.9792	1.7078311	0	1	0.0235188	0.51485149	0.45413806	0	0.08973331	0.26951697
13	communications-vepco	38	0.19487179	0.17435897	0.69892473	14.92307692	0.1852498	0.953846	7.95625	0.5697589	0.0174941	0.946237	0.00935765	0.64835165	0.32294493	0.09546941	0.05135201	0.2535225
14	electric-ecar	2100	0.50047664	0.12345091	0.10986585	6.75479143	1.92516643	2.80696	15.0669	0.2325981	0.0637916	0.84089	0.00146402	0.66280864	0.17767623	0.29579787	0.09647542	0.19439005
15	electric-ercot	2161	0.45745131	0.22967824	0.24325278	8.97367634	1.5347562	2.40008	16.3052	0.23723086	0.0228069	0.955724	0.00110195	0.67793526	0.17689276	0.29900947	0.0753311	0.1780598
16	electric-frcc	969	0.46541787	0.24303554	0.26333973	6.95916905	1.50895495	2.5024	14.4244	0.31612395	0.0543176	0.889443	0.00256749	0.66658677	0.19812414	0.29707174	0.07468206	0.20811513
17	electric-maac	1291	0.45521862	0.23695346	0.25236423	6.04502618	1.53761805	2.61001	14.7855	0.27764055	0.0718796	0.861119	0.00206118	0.66570258	0.1858115	0.26672826	0.07880943	0.21063872
18	electric-main	1624	0.47168167	0.2558815	0.27200188	7.49934124	1.50163184	2.47517	17.8192	0.30368245	0.0380145	0.918094	0.00152631	0.66755241	0.20649212	0.28562601	0.0826403	0.19455383
19	electric-mapp	2143	0.47856186	0.25055828	0.239677	8.52890995	1.55282702	2.43368	18.0148	0.26920765	0.0254283	0.941274	0.0010823	0.65355422	0.19029194	0.29857135	0.08241329	0.18085851
20	electric-nepool	976	0.44770642	0.33853211	0.35415108	5.87611336	1.35875016	2.45321	14.3229	0.30676281	0.0541576	0.908751	0.00253378	0.67322967	0.1896126	0.24788162	0.06468315	0.20522738
21	electric-nypp	1972	0.47898956	0.20986155	0.2036225	7.08292363	1.63491614	2.54797	16.5224	0.25750324	0.0360854	0.927359	0.00138254	0.67032904	0.18780961	0.28597261	0.07863172	0.18247665
22	electric-serc	2512	0.45416742	0.12294341	0.16706013	7.44654814	1.71023388	2.60387	19.4424	0.26142556	0.0528434	0.889738	0.00116594	0.69785918	0.19897511	0.29209361	0.08768865	0.18218712
23	electric-spp	1751	0.469437	0.21018767	0.22395717	8.64183486	1.56295544	2.40375	17.6387	0.28880995	0.0169695	0.965871	0.00132028	0.66555523	0.19688268	0.3056805	0.07772255	0.17913145
24	electric-wscc	5549	0.4538688	0.30770489	0.34887044	7.15661465	1.34101064	2.40406	29.2494	0.2645299	0.0545875	0.901402	0.00043797	0.67647728	0.20434085	0.22728532	0.07092183	0.16654338
25	fde	16	0.37209302	0	0	3.1	9.85714286	14.4186	1.61538	0.24634314	0.268898	0.0645161	0.0310078	0.03825137	0.12081633	0.1476762	0.37699222	0.75363607
26	gas-florida	11	0.34375	0.625	1	NA	1	1.9375	6.6129	1.16900661	0	1	0.0826613	0.56944444	0.38925806	0	0	0.34329189
27	highways-canada	2594	0.49579511	0.64468654	0.86286562	79	0.05345075	1.80352	9.78571	0.13528775	0.00033893	0.998728	0.00033674	0.49399657	0.35301098	0.0796745	0.01389749	0.09569234
28	highways-mexico	295	0.56949807	0.09073359	0.06955381	4.69957082	2.20552191	2.94208	14.3124	0.62885075	0.117023	0.744094	0.00980561	0.60340074	0.27922695	0.30264751	0.09725053	0.20428433
29	highways-usa	49482	0.54727645	0.0375159	0.02909073	7.32268363	2.09952229	2.76021	179.175	0.59587775	0.044399	0.875319	0.0000499	0.59208288	0.26174589	0.22916098	0.15988696	0.10426124
30	internet-bell	50601	0.41265841	0.39376295	0.4756527	5.43476146	1.15276851	2.64444	12.0491	0.03440887	0.0296517	0.936015	0.00008956	0.72027347	0.1054896	0.11292968	0.07548777	0.14335341
31	internet-nlanr	1057	0.16326846	0.09267841	0.19495705	3.52224708	1.49440474	3.88384	3.62429	0.04504396	0.399239	0.494353	0.0780689	0.89260535	0.07804508	0.14260873	0.08401679	0.18907717
32	microwave-fcc	4232	0.10001654	0.08715997	0.57478392	18.50065738	0.04173336	0.503108	53.4487	0.25983655	0.0234287	0.95838	0.00002448	0.66685939	0.26410621	0.05757373	0.01314161	0.08400873
33	nts	56600	0.55715242	0.05512462	0.04554937	4.76843195	2.41858238	3.22739	68.5389	0.21503835	0.123219	0.687742	0.0000681	0.6438943	0.20959826	0.16403004	0.17708942	0.13980555
34	petroleum-mapsearch	73531	0.45711462	0.44558899	0.6167617	17.26145013	0.47374302	2.12382	113.001	0.2817472	0.0283286	0.936377	0.00001979	0.59142162	0.2419355	0.14753437	0.11711469	0.10900262
35	railways-canada	339	0.49059334	0.31114327	0.32740214	6.21760391	1.27274481	2.43994	21.7366	0.82689923	0.0579727	0.8707	0.00576354	0.59895379	0.29678531	0.30458266	0.08568845	0.20258173
36	railways-mexico	89	0.4972067	0.32960894	0.36714976	8.54901961	1.34580378	2.31285	13.3295	0.99629361	0.0410256	0.903382	0.0192706	0.5847619	0.33393947	0.37282851	0.08870382	0.25441943
37	railways-usa	68793	0.5143325	0.27803696	0.28920883	19.88641583	1.38351648	2.395	292.32	0.79929738	0.0908936	0.818928	0.00002485	0.56115609	0.27788544	0.22417699	0.11105573	0.10134755
38	roads-dallas	5384	0.54587854	0.11315016	0.08162712	5.6763586	2.19662045	2.99098	38.3565	0.38621973	0.0619766	0.874237	0.00055304	0.5962536	0.23129609	0.23260439	0.12201763	0.16441622
39	roads-portland	50511	0.50084282	0.2831575	0.27784339	9.93797711	1.52371542	2.5019	129.788	0.40868837	0.0190688	0.964862	0.00003885	0.57234688	0.2382361	0.20987064	0.13797877	0.11071167
40	telephone-lerg	5938	0.17540025	0.01533054	0.0330626	3.11291271	2.40017271	6.27713	8.18661	0.04449376	0.569185	0.157492	0.00451982	0.57921127	0.11362691	0.06913777	0.21254751	0.22591528
41	waterways-usa	2968	0.4738943	0.48650806	0.61295488	18.11309812	1.0937261	2.19384	37.3894	0.47245156	0.0218094	0.94425	0.00059441	0.61665432	0.25891655	0.18278172	0.0314767	0.16286626
42	air-tickets	364	0.55572519	0.02021773	0.00106943	3	70.44770528	148.47000000	1.83489000	0.07169508	0.84740200	0.00106943	0.15816700	0.17480720	0.06024672	0.04904803	0.50534453	0.80416890
43	gas-utah	68	0.42500000	0.61250000	1	NA	1	1.98750000	19.10060000	1.51003502	0	1	0.01926100	0.60643564	0.35975775	0	0	0.25193284

2. Graph Measurements “clean0” Data Set

graph	*type*	*subtype*	*name*	*V*	*E*	*n*	*delta*	*DDelta*	*r*	*rHat*	*d*	*dHat*	*g*	*kappa*	*kappa1*	*chi*	*gamma*	*gammaHat*	*beta*	*betaHat*	*i*	*iHat*
1	1	1	air-sample	23	234	1	15	22	1	0.20851441	2	0.41702883	3	NA	NA	18	1	0.04347826	2	0.08695652	1	0.04347826
2	2	7	communications-bge	36	41	1	1	5	6	1	9	1.5	4	1	1	3	11	0.30555556	20	0.55555556	13	0.36111111
3	2	7	communications-gpc	65	62	3	1	4	4	0.49613894	13	1.61245155	NA	1	1	3	25	0.38461538	36	0.55384615	27	0.41538462
4	2	7	communications-jepco	23	25	1	1	4	6	1.25108648	11	2.29365855	3	1	1	3	7	0.30434783	12	0.52173913	8	0.34782609
5	2	7	communications-pepco	32	44	1	1	5	4	0.70710678	8	1.41421356	3	1	1	3	11	0.34375	15	0.46875	12	0.375
6	2	7	communications-pge	268	321	1	1	15	18	1.099525	36	2.19905	3	1	1	3	83	0.30970149	165	0.61567164	103	0.38432836
7	2	7	communications-pnm	47	47	1	1	8	7	1.02105494	13	1.89624489	8	1	1	3	16	0.34042553	30	0.63829787	18	0.38297872
8	2	7	communications-sceg	27	26	1	1	7	7	1.34715063	13	2.50185117	NA	1	1	3	9	0.33333333	17	0.62962963	9	0.33333333
9	2	7	communications-sdge	41	45	1	1	8	5	0.78086881	10	1.56173762	5	1	1	3	12	0.29268293	25	0.6097561	17	0.41463415
10	2	7	communications-sepco	6	5	1	1	2	3	1.22474487	5	2.04124145	NA	1	1	2	2	0.33333333	3	0.5	3	0.5
11	2	7	communications-sngc	107	110	1	1	5	22	2.12682028	39	3.77027231	3	1	1	3	37	0.34579439	59	0.55140187	43	0.40186916
12	2	7	communications-tpc	49	48	1	1	3	20	2.85714286	40	5.71428571	NA	1	1	3	17	0.34693878	26	0.53061224	20	0.40816327
13	2	7	communications-vepco	81	84	1	1	9	12	1.33333333	23	2.55555556	3	1	1	3	25	0.30864198	51	0.62962963	29	0.35802469
14	3	11	electric-ecar	4194	5888	1	1	13	19	0.293386	37	0.57133063	3	1	1	5	1270	0.30281354	2113	0.50381497	1489	0.355031
15	3	11	electric-ercot	4718	5669	1	1	16	20	0.29117295	40	0.5823459	3	1	1	5	1568	0.33234421	2573	0.5453582	1764	0.37388724
16	3	11	electric-frcc	2079	2605	1	1	12	20	0.43863446	40	0.87726893	3	1	1	4	685	0.32948533	1118	0.53775854	802	0.38576239
17	3	11	electric-maac	2836	3701	1	1	15	18	0.33800209	36	0.67600419	3	1	1	5	901	0.31770099	1539	0.54266573	1081	0.38117066
18	3	11	electric-main	3424	4250	1	1	16	26	0.44433085	51	0.87157206	3	1	1	5	1174	0.34287383	1822	0.53212617	1364	0.39836449
19	3	11	electric-mapp	4457	5439	1	1	12	25	0.37447143	48	0.71898515	3	1	1	4	1577	0.35382544	2336	0.52411936	1749	0.39241642
20	3	11	electric-nepool	2175	2674	1	1	17	17	0.36451826	32	0.68615202	3	1	1	4	764	0.35126437	1190	0.54712644	875	0.40229885
21	3	11	electric-nypp	4117	5245	1	1	27	24	0.37404238	47	0.73249965	3	1	1	4	1340	0.32547972	2157	0.52392519	1522	0.36968667
22	3	11	electric-serc	5526	7199	1	1	23	28	0.37666268	55	0.73987312	3	1	1	4	1591	0.28791169	3015	0.54560261	1931	0.34943902
23	3	11	electric-spp	3711	4468	1	1	13	28	0.45963443	55	0.90285335	3	1	1	4	1238	0.3336028	1990	0.5362436	1445	0.38938292
24	3	11	electric-wscc	12202	14695	1	1	26	40	0.3621133	77	0.69706811	3	1	1	4	4096	0.33568267	6657	0.5455663	4803	0.393624
25	4	13	fde	40	310	1	2	29	2	0.31622777	4	0.63245553	3	NA	NA	5	4	0.1	24	0.6	11	0.275
26	3	12	gas-florida	32	31	1	1	4	9	1.59099026	18	3.18198052	NA	1	1	2	11	0.34375	21	0.65625	11	0.34375
27	1	2	highways-canada	953	966	1	1	4	89	2.88299279	177	5.7335924	8	1	1	3	357	0.37460651	463	0.48583421	427	0.44805876
28	1	2	highways-mexico	518	762	1	1	9	25	1.09843694	49	2.1529364	3	1	1	4	161	0.31081081	229	0.44208494	185	0.35714286
29	1	2	highways-usa	90246	124659	12	1	6	4	0.01331515	534	1.77757231	3	1	1	4	28582	0.3167121	41449	0.45928905	33401	0.37011059
30	2	10	internet-bell	121665	161780	1	1	130	22	0.06307247	43	0.12327801	3	1	1	6	NA	NA	71373	0.58663543	50354	0.41387416
31	2	10	internet-nlanr	6474	12572	1	1	1458	5	0.06214178	9	0.1118552	3	1	1	11	660	0.10194625	5411	0.83580476	2861	0.44192153
32	2	8	microwave-fcc	9539	10233	91	1	32	1	0.01023879	174	1.78154921	3	1	1	4	3234	0.33902925	5638	0.59104728	3766	0.39480029
33	1	6	nts	100985	163780	11	1	51	2	0.00629364	206	0.64824442	3	1	1	6	28463	0.28185374	45279	0.44837352	33434	0.33107887
34	3	12	petroleum-mapsearch	142166	154791	64	1	16	11	0.02917392	528	1.40034824	3	1	1	5	NA	NA	77928	0.54814794	61120	0.42991995
35	1	3	railways-canada	654	813	1	1	7	30	1.17309283	60	2.34618566	3	1	1	4	231	0.35321101	337	0.51529052	273	0.41743119
36	1	3	railways-mexico	177	206	1	1	5	24	1.80395047	47	3.53273633	3	1	1	3	59	0.33333333	89	0.50282486	71	0.40112994
37	1	3	railways-usa	132432	159082	1	1	9	596	1.63775823	1167	3.20681855	3	1	1	4	NA	NA	65269	0.49284916	55078	0.41589646
38	1	4	roads-dallas	9857	14747	1	1	8	64	0.64462567	111	1.11802265	3	1	1	4	3134	0.31794664	4550	0.46160089	3690	0.37435325
39	1	4	roads-portland	100733	126059	1	1	6	202	0.63645175	400	1.2603005	3	1	1	4	36519	0.36253264	50947	0.50576276	42727	0.4241609
40	2	9	telephone-lerg	31768	105301	1	1	1133	15	0.08415818	29	0.16270581	3	1	1	7	915	0.02880257	25835	0.81323974	2160	0.06799295
41	1	5	waterways-usa	6251	6863	1	1	20	93	1.17627319	185	2.33989828	3	1	1	4	2408	0.38521837	3338	0.53399456	2731	0.4368901
42	1	1	air-tickets	643	48624	1	1	515	2	0.07887230	4	0.15774460	3	1	1	118	22	0.03421462	278	0.43234837	125	0.19440124
43	3	12	gas-utah	160	159	1	1	8	22	1.73925271	44	3.47850543	NA	1	1	3	59	0.36875000	93	0.58125000	71	0.44375000

graph	*name*	*alpha*	*alphaHat*	*aBar*	*bBar*	*cBar*	*qBar*	*kBar*	*LBar*	*LBarHat*	*gammaBar*	*phiBar*	*psiBar*	*psiStar*	*elTilde*	*cTilde*	*qTilde*	*kTilde*
1	air-sample	20	0.86956522	0	0	3	19.36758893	20.3478	1.04545	0.21799139	0.937497	0	0	0	0.04282198	0	0.29518175	0.49142993
2	communications-bge	16	0.44444444	0.25	0.29268293	7.08333333	1.47301587	2.27778	4.28571	0.714285	0	1	0.101587	0.58715596	0.28047942	0.37808787	0.12524284	0.38704428
3	communications-gpc	29	0.44615385	0.61538462	1	NA	0.35769231	1.90769	4.85714	0.60245407	0	1	0.0384615	0.56338028	0.32681653	0	0.08534926	0.28282313
4	communications-jepco	11	0.47826087	0.08695652	0.08	9.95238095	1.83794466	2.17391	4.45455	0.92883788	0.0634921	0.88	0.126482	0.56140351	0.3503822	0.38511165	0.08690903	0.27445721
5	communications-pepco	17	0.53125	0.125	0.09090909	4.32142857	1.95766129	2.75	3.70968	0.65578497	0.0880952	0.818182	0.155242	0.6015625	0.27029622	0.46701789	0.16335223	0.42423919
6	communications-pge	105	0.39179104	0.3358209	0.47352025	4.71186441	1.15489966	2.39552	9.49064	0.57973311	0.118993	0.831776	0.0233384	0.70464135	0.27637729	0.25174307	0.04338999	0.28708617
7	communications-pnm	18	0.38297872	0.42553191	0.82978723	8	1.02590194	2	5.3913	0.78640193	0	1	0.080481	0.64925373	0.29912552	0.11849504	0.01720596	0.33659177
8	communications-sceg	11	0.40740741	0.55555556	1	NA	1	1.92593	4.96154	0.95484882	0	1	0.131054	0.63888889	0.35198348	0	0	0.36911917
9	communications-sdge	17	0.41463415	0.31707317	0.51111111	5.83333333	1.20487805	2.19512	4.025	0.62859939	0	1	0.119512	0.68531469	0.26384012	0.2911519	0.08064644	0.38670346
10	communications-sepco	3	0.5	0.66666667	1	NA	1	1.66667	2.2	0.89814624	0	1	0.266667	0.44444444	0.4380076	0	0	0.3552453
11	communications-sngc	48	0.44859813	0.4953271	0.55454545	33	1.16769529	2.05607	13.9434	1.34795936	0.0426357	0.945455	0.0239817	0.5483871	0.36842706	0.20886768	0.05254734	0.22913669
12	communications-tpc	25	0.51020408	0.85714286	1	NA	1	1.95918	14	2	0	1	0.0442177	0.52	0.45422471	0	0	0.18577253
13	communications-vepco	32	0.39506173	0.38271605	0.66666667	14.92307692	1.07407407	2.07407	8.3875	0.93194444	0.0186869	0.940476	0.0521605	0.66015625	0.32296026	0.20348344	0.03299049	0.31774835
14	electric-ecar	2103	0.50143062	0.12350978	0.10971467	6.75479143	1.9239198	2.80782	15.0673	0.23265973	0.0637916	0.840863	0.00146542	0.66284274	0.1776762	0.29584897	0.09581002	0.19437538
15	electric-ercot	2161	0.45803306	0.22997033	0.24325278	8.97367634	1.5429536	2.40314	16.3052	0.23738166	0.0228069	0.955724	0.00110475	0.67793526	0.17689276	0.29924958	0.07414797	0.17716024
16	electric-frcc	966	0.46464646	0.24338624	0.26333973	6.95916905	1.51415212	2.50601	14.4244	0.31635195	0.0543176	0.889443	0.0025749	0.66658677	0.19812414	0.29734659	0.07336127	0.20703027
17	electric-maac	1291	0.45521862	0.23695346	0.25236423	6.04502618	1.54003007	2.61001	14.7855	0.27764055	0.0718796	0.861119	0.00206118	0.66570258	0.1858115	0.26672826	0.07911932	0.21063872
18	electric-main	1614	0.4713785	0.25554907	0.27011765	7.49934124	1.52298364	2.48248	17.8313	0.30473064	0.0381062	0.917882	0.00154227	0.66794768	0.20649137	0.2866643	0.07846884	0.19417043
19	electric-mapp	2139	0.47991923	0.25061701	0.2382791	8.52890995	1.56801852	2.44066	18.0289	0.27005232	0.0254657	0.941166	0.00109182	0.65380209	0.19029135	0.29945477	0.07865563	0.1799261
20	electric-nepool	978	0.44965517	0.33931034	0.35415108	5.87611336	1.36441512	2.45885	14.3229	0.30711521	0.0541576	0.908751	0.00254544	0.67322967	0.1896126	0.2482752	0.06257234	0.20364375
21	electric-nypp	1972	0.47898956	0.20986155	0.2036225	7.08292363	1.63126702	2.54797	16.5224	0.25750324	0.0360854	0.927359	0.00138254	0.67032904	0.18780961	0.28597261	0.07833239	0.18247665
22	electric-serc	2504	0.45313066	0.12305465	0.16682873	7.44654814	1.7166244	2.6055	19.4463	0.26159627	0.0528434	0.889707	0.00116805	0.69791381	0.19897508	0.29222797	0.08649652	0.18202658
23	electric-spp	1746	0.47049313	0.21018593	0.22269472	8.64766458	1.5838716	2.40798	17.6593	0.28988651	0.0170158	0.965756	0.00133166	0.66591602	0.19688131	0.30634701	0.07397237	0.17930036
24	electric-wscc	5553	0.45508933	0.30831011	0.34882613	7.15661465	1.35109023	2.40862	29.2499	0.26479445	0.0545875	0.901395	0.00043969	0.67649126	0.20434084	0.2275974	0.06944424	0.16548162
25	fde	16	0.4	0	0	3.1	11.41153846	15.5	1.61538	0.255414	0.268898	0.0645161	0.0358974	0.03825137	0.12081633	0.08812513	0.37684207	0.75230286
26	gas-florida	11	0.34375	0.625	1	NA	1	1.9375	6.6129	1.16900661	0	1	0.0826613	0.56944444	0.38925806	0	0	0.34329189
27	highways-canada	500	0.52465897	0.47534103	0.49689441	82.20886076	1.22747215	2.02728	67.2637	2.17888497	0	1	0.00238962	0.52878049	0.36010113	0.2659886	0.04698621	0.08373574
28	highways-mexico	295	0.56949807	0.09073359	0.06955381	4.69957082	2.20552191	2.94208	14.3124	0.62885075	0.117023	0.744094	0.00980561	0.60340074	0.27922695	0.30264751	0.09725053	0.20428433
29	highways-usa	49413	0.54753673	0.03706535	0.02822901	7.32328158	2.18212766	2.76265	179.25	0.59668509	0.044379	0.875292	0.00005007	0.59218067	0.26174588	0.2291139	0.15825436	0.10400974
30	internet-bell	50394	0.41420293	0.39566843	0.47450859	5.43476146	1.16141996	2.65943	12.0615	0.03457948	0.0297089	0.935876	0.00009095	0.72049351	0.10548957	0.11358339	0.06216228	0.14226347
31	internet-nlanr	1057	0.16326846	0.09267841	0.19495705	3.52224708	1.49277635	3.88384	3.62429	0.04504396	0.399239	0.494353	0.0780689	0.89260535	0.07804508	0.14260873	0.08378886	0.18907717
32	microwave-fcc	3930	0.41199287	0.3743579	0.5582918	18.52515144	0.8345195	2.14551	54.4417	0.55741706	0.0228959	0.957295	0.00047877	0.6739487	0.26410347	0.22378682	0.09361927	0.15794978
33	nts	56499	0.55947913	0.05494875	0.04492612	4.76916342	2.46892093	3.24365	68.5706	0.21577917	0.123097	0.687691	0.00006889	0.64398797	0.20959825	0.1638107	0.17407627	0.13802528
34	petroleum-mapsearch	65430	0.46023662	0.45027644	0.58769567	17.57072646	0.64359352	2.17761	118.897	0.31533562	0.0289529	0.933814	0.00002351	0.59663908	0.24193045	0.16144399	0.12781234	0.11117795
35	railways-canada	326	0.49847095	0.30733945	0.30258303	6.21760391	1.42018255	2.48624	22.0666	0.86287234	0.059707	0.865929	0.0062848	0.60125448	0.29675845	0.31636378	0.06620499	0.20496947
36	railways-mexico	88	0.49717514	0.33333333	0.36407767	8.54901961	1.37641243	2.32768	13.4318	1.00959591	0.0410256	0.902913	0.0197098	0.58587786	0.33393616	0.37600184	0.07964448	0.25430198
37	railways-usa	68207	0.51503413	0.27766703	0.28572057	19.91928206	1.39285714	2.40247	292.929	0.80494443	0.0909964	0.81832	0.00002522	0.56155379	0.27788532	0.22555633	0.10099939	0.1012383
38	roads-dallas	5374	0.54519631	0.11321903	0.08144029	5.6763586	2.19967011	2.99219	38.3607	0.38637956	0.0619766	0.874212	0.00055372	0.59629326	0.23129608	0.23262158	0.12124273	0.16438074
39	roads-portland	50475	0.5010771	0.28305521	0.27739392	9.93923404	1.53417969	2.50283	129.826	0.40904943	0.019051	0.964858	0.00003892	0.5723959	0.23823609	0.21001843	0.13686423	0.1107129
40	telephone-lerg	5866	0.18465122	0.01573911	0.02999972	3.11340206	2.78303943	6.62938	8.18995	0.04595008	0.566685	0.155174	0.00496329	0.57155435	0.11360666	0.0624248	0.19703821	0.21985539
41	waterways-usa	2962	0.47384418	0.48712206	0.6125601	18.11309812	1.09497787	2.19581	37.4198	0.47328933	0.0218204	0.944194	0.00059659	0.61684222	0.25891646	0.18308461	0.02863272	0.16297959
42	air-tickets	364	0.56609642	0.02488336	0.00106943	3	73.28648737	151.24100000	1.83489000	0.07236100	0.84740200	0.00106943	0.16413000	0.17640320	0.06024672	0.04344031	0.51036269	0.80803064
43	gas-utah	68	0.42500000	0.61250000	1	NA	1	1.98750000	19.10060000	1.51003502	0	1	0.01926100	0.60643564	0.35975775	0	0	0.25193284

3. Graph Measurements for "clean1" Data Set

graph	*type*	*subtype*	*name*	*V*	*E*	*n*	*delta*	*DDelta*	*r*	*rHat*	*d*	*dHat*	*g*	*kappa*	*kappa1*	*chi*	*gamma*	*gammaHat*	*beta*	*betaHat*	*i*	*iHat*
1	1	1	air-sample	23	234	1	15	22	1	0.20851441	2	0.41702883	3	NA	NA	18	1	0.04347826	2	0.08695652	1	0.04347826
2	2	7	communications-bge	26	31	1	1	5	5	0.98058068	8	1.56892908	3	1	1	3	8	0.30769231	15	0.57692308	9	0.34615385
3	2	7	communications-gpc	42	39	3	1	4	2	0.3086067	8	1.2344268	NA	1	1	2	16	0.38095238	26	0.61904762	21	0.5
4	2	7	communications-jepco	7	9	1	1	4	2	0.75592895	4	1.51185789	3	1	1	3	3	0.42857143	3	0.42857143	3	0.42857143
5	2	7	communications-pepco	22	34	1	1	5	4	0.85280287	7	1.49240501	3	1	1	4	7	0.31818182	10	0.45454545	9	0.40909091
6	2	7	communications-pge	205	258	1	1	15	11	0.76827333	22	1.53654665	3	1	1	4	57	0.27804878	131	0.63902439	79	0.38536585
7	2	7	communications-pnm	34	34	1	1	8	4	0.68599434	8	1.37198868	3	1	1	3	11	0.32352941	23	0.67647059	13	0.38235294
8	2	7	communications-sceg	17	16	1	1	7	3	0.72760688	5	1.21267813	NA	1	1	2	4	0.23529412	13	0.76470588	5	0.29411765
9	2	7	communications-sdge	26	30	1	1	8	4	0.78446454	7	1.37281295	3	1	1	3	8	0.30769231	17	0.65384615	13	0.5
10	2	7	communications-sepco	2	1	1	1	1	1	0.70710678	1	0.70710678	NA	NA	1	2	1	0.5	1	0.5	1	0.5
11	2	7	communications-sngc	47	50	1	1	5	8	1.16691993	12	1.7503799	3	1	1	3	18	0.38297872	26	0.55319149	22	0.46808511
12	2	7	communications-tpc	12	11	1	1	3	3	0.8660254	6	1.73205081	NA	1	1	2	5	0.41666667	7	0.58333333	5	0.41666667
13	2	7	communications-vepco	60	63	1	1	9	7	0.90369611	13	1.67829278	3	1	1	3	18	0.3	41	0.68333333	23	0.38333333
14	3	11	electric-ecar	2658	4352	1	1	13	16	0.31034339	31	0.60129031	3	1	1	5	769	0.28931527	1285	0.4834462	913	0.34349135
15	3	11	electric-ercot	2910	3861	1	1	16	15	0.278064	28	0.5190528	3	1	1	5	1010	0.34707904	1607	0.55223368	1169	0.40171821
16	3	11	electric-frcc	1480	2006	1	1	12	17	0.44189396	31	0.80580663	3	1	1	4	474	0.32027027	804	0.54324324	574	0.38783784
17	3	11	electric-maac	2060	2925	1	1	15	15	0.33048949	30	0.66097897	3	1	1	5	636	0.30873786	1140	0.55339806	797	0.3868932
18	3	11	electric-main	2484	3310	1	1	16	21	0.42135049	41	0.82263666	3	1	1	5	864	0.34782609	1344	0.5410628	1014	0.40821256
19	3	11	electric-mapp	3027	4009	1	1	12	19	0.3453404	36	0.65432918	3	1	1	4	1104	0.36471754	1588	0.52461183	1233	0.40733399
20	3	11	electric-nepool	1760	2259	1	1	17	15	0.35754847	30	0.71509694	3	1	1	4	605	0.34375	1010	0.57386364	706	0.40113636
21	3	11	electric-nypp	2608	3736	1	1	27	18	0.3524672	36	0.7049344	3	1	1	4	825	0.31633436	1368	0.52453988	978	0.375
22	3	11	electric-serc	3182	4855	1	1	23	24	0.42546237	47	0.83319714	3	1	1	4	825	0.2592709	1775	0.55782527	1064	0.33438089
23	3	11	electric-spp	2143	2900	1	1	13	22	0.47523848	43	0.9288752	3	1	1	4	712	0.33224452	1164	0.54316379	841	0.3924405
24	3	11	electric-wscc	8381	10874	1	1	26	25	0.27308138	48	0.52431625	3	1	1	4	2788	0.3326572	4770	0.56914449	3398	0.40544088
25	4	13	fde	38	308	1	3	29	2	0.32444284	3	0.48666426	3	NA	NA	6	2	0.05263158	22	0.57894737	9	0.23684211
26	3	12	gas-florida	21	20	1	1	4	5	1.09108945	9	1.96396101	NA	1	1	3	8	0.38095238	13	0.61904762	9	0.42857143
27	1	2	highways-canada	148	161	1	1	4	14	1.15079291	27	2.21938633	3	1	1	3	56	0.37837838	84	0.56756757	63	0.42567568
28	1	2	highways-mexico	445	689	1	1	9	22	1.04290002	43	2.03839549	3	1	1	4	131	0.29438202	190	0.42696629	159	0.35730337
29	1	2	highways-usa	53850	88263	12	1	6	2	0.00861861	329	1.41776106	3	1	1	4	15893	0.29513463	21950	0.40761374	19465	0.36146704
30	2	10	internet-bell	83071	123186	1	1	130	13	0.04510437	25	0.08673918	3	1	1	6	21937	0.26407531	55802	0.67173863	35815	0.43113722
31	2	10	internet-nlanr	5486	11584	1	1	1458	5	0.06750596	8	0.10800953	3	1	1	14	562	0.10244258	4665	0.85034634	2704	0.492891
32	2	8	microwave-fcc	6394	7088	91	1	32	1	0.01250586	87	1.08801012	3	1	1	4	2063	0.32264623	4122	0.64466688	2664	0.4166406
33	1	6	nts	82361	145156	11	1	51	1	0.00348449	174	0.60630122	3	1	1	6	21856	0.26536832	35542	0.4315392	25817	0.31346147
34	3	12	petroleum-mapsearch	111436	124061	64	1	16	1	0.00299562	309	0.92564769	3	1	1	5	NA	NA	63065	0.56593022	48299	0.43342367
35	1	3	railways-canada	579	738	1	1	7	26	1.08052313	51	2.11948767	3	1	1	4	202	0.34887737	298	0.51468048	240	0.41450777
36	1	3	railways-mexico	109	138	1	1	5	12	1.14939154	24	2.29878308	3	1	1	3	36	0.33027523	57	0.52293578	44	0.40366972
37	1	3	railways-usa	106316	132966	1	1	9	336	1.03048094	671	2.05789497	3	1	1	4	37629	0.35393544	52867	0.49726288	43824	0.41220512
38	1	4	roads-dallas	8144	13034	1	1	8	51	0.56513381	89	0.98621391	3	1	1	5	2558	0.31409627	3561	0.43725442	3040	0.37328094
39	1	4	roads-portland	88379	113705	1	1	6	147	0.49447324	287	0.96540013	3	1	1	4	32073	0.36290295	44699	0.50576494	37844	0.42820127
40	2	9	telephone-lerg	31225	104758	1	1	1133	13	0.07356854	26	0.14713708	3	1	1	7	825	0.02642114	25345	0.81168935	2787	0.0892554
41	1	5	waterways-usa	4709	5321	1	1	20	51	0.74320022	100	1.45725534	3	1	1	4	1789	0.37991081	2676	0.56827352	2051	0.43554895
42	1	1	air-tickets	643	48624	1	1	515	2	0.07887230	4	0.15774460	3	1	1	118	22	0.03421462	278	0.43234837	125	0.19440124
43	3	12	gas-utah	107	106	1	1	8	11	1.06341014	22	2.12682028	NA	1	1	3	39	0.36448598	66	0.61682243	50	0.46728972

graph	*name*	*alpha*	*alphaHat*	*aBar*	*bBar*	*cBar*	*qBar*	*kBar*	*LBar*	*LBarHat*	*gammaBar*	*phiBar*	*psiBar*	*psiStar*	*elTilde*	*cTilde*	*qTilde*	*kTilde*
1	air-sample	20	0.86956522	0	0	3	19.36758893	20.3478	1.04545	0.21799139	0.937497	0	0	0	0.04282198	0	0.29518175	0.49142993
2	communications-bge	12	0.46153846	0.30769231	0.35483871	4.8	1.39076923	2.38462	3.42	0.67071718	0.152083	0.806452	0.172308	0.64367816	0.28586391	0.39589789	0.13745411	0.37789849
3	communications-gpc	17	0.4047619	0.4047619	1	NA	0.37398374	1.85714	3.5625	0.54970568	0	1	0.0662021	0.59375	0.29723393	0	0.09781484	0.21979051
4	communications-jepco	4	0.57142857	0.28571429	0.22222222	3	1.57142857	2.57143	1.66667	0.62994205	0.533333	0.222222	0.190476	0.25	0.31842354	0.30744974	0.27814359	0.69467956
5	communications-pepco	12	0.54545455	0.18181818	0.11764706	3.22222222	2.08225108	3.09091	2.85714	0.60914429	0.4	0.382353	0.186147	0.49425287	0.28082672	0.30073274	0.21724251	0.50288304
6	communications-pge	73	0.35609756	0.28292683	0.46511628	3.44827586	1.16188427	2.51707	7.11765	0.49711824	0.290727	0.635659	0.0343855	0.71329365	0.24545151	0.18566438	0.04653315	0.29373095
7	communications-pnm	11	0.32352941	0.35294118	0.91176471	3	1.00534759	2	4.0303	0.69119075	0.105311	0.911765	0.12656	0.67619048	0.27237894	0.08462998	0.00526202	0.30688813
8	communications-sceg	4	0.23529412	0.29411765	1	NA	1	1.88235	2.8125	0.68213145	0	1	0.264706	0.69230769	0.2679415	0	0	0.31246811
9	communications-sdge	9	0.34615385	0.30769231	0.6	3.25	1.13230769	2.30769	3.22	0.63149396	0.146104	0.733333	0.196923	0.64646465	0.26483161	0.24255511	0.06104006	0.3772012
10	communications-sepco	1	0.5	0	1	NA	1	1	1	0.70710678	NA	1	0	0	0.00000026	0.00000026	0	0.00000026
11	communications-sngc	21	0.44680851	0.42553191	0.56	9.6	1.13043478	2.12766	5.93478	0.86567663	0.205128	0.82	0.0712303	0.60629921	0.30866487	0.25319545	0.05631138	0.31031836
12	communications-tpc	5	0.41666667	0.41666667	1	NA	1	1.83333	3.09091	0.89226886	0	1	0.227273	0.57692308	0.35113533	0	0	0.27332746
13	communications-vepco	20	0.33333333	0.35	0.73015873	6.73333333	1.04067797	2.1	5.5678	0.71879989	0.137198	0.873016	0.0841808	0.69953052	0.28674423	0.17953203	0.02295449	0.29936827
14	electric-ecar	1386	0.5214447	0.15951843	0.12683824	4.04843305	2.15262101	3.27464	12.5883	0.24416848	0.170308	0.677619	0.00288631	0.65793041	0.17829397	0.20680172	0.12778493	0.21315045
15	electric-ercot	1326	0.4556701	0.3137457	0.31261331	5.39308729	1.46658463	2.65361	13.1781	0.24429035	0.105578	0.831391	0.00227638	0.69501551	0.17589508	0.23502663	0.08543638	0.18467263
16	electric-frcc	684	0.46216216	0.27567568	0.29312064	5.02793296	1.50612368	2.71081	12.2637	0.3187797	0.148726	0.757727	0.00417558	0.66871525	0.19847517	0.24417092	0.08159588	0.20582496
17	electric-maac	928	0.45048544	0.26262136	0.27452991	4.19525692	1.55213568	2.83981	12.881	0.28380234	0.207767	0.703248	0.00326721	0.66727658	0.18573618	0.20509839	0.08651562	0.21908689
18	electric-main	1159	0.46658615	0.29750403	0.30574018	5.06245757	1.51062407	2.66506	14.9525	0.30001158	0.105598	0.821148	0.00248323	0.67811919	0.20306684	0.23084044	0.08596083	0.18596245
19	electric-mapp	1471	0.4859597	0.31020813	0.27887254	5.40303189	1.54198307	2.64883	14.698	0.26714806	0.11262	0.826889	0.00194439	0.6711637	0.18563293	0.24051748	0.09113638	0.17822604
20	electric-nepool	765	0.43465909	0.325	0.3457282	4.87728195	1.38580335	2.56705	12.7743	0.30449543	0.142846	0.806109	0.00346077	0.67999492	0.1875552	0.22877464	0.06761716	0.19266627
21	electric-nypp	1256	0.48159509	0.25115031	0.22992505	4.29640206	1.6695724	2.86503	13.3381	0.26118016	0.164275	0.741435	0.00270685	0.67532658	0.18675182	0.21353332	0.09557876	0.19639049
22	electric-serc	1394	0.43808925	0.18070396	0.22574665	3.75860421	1.75705839	3.05154	15.5937	0.27643886	0.21668	0.671061	0.00269059	0.70057108	0.20438099	0.17720188	0.11268599	0.22266362
23	electric-spp	992	0.46290247	0.28931405	0.28793103	4.84210526	1.53134714	2.70649	14.0089	0.30261674	0.110513	0.82931	0.0029632	0.67745791	0.20134858	0.23312032	0.08966396	0.19556922
24	electric-wscc	3658	0.43646343	0.32144136	0.37318374	4.53004587	1.30896991	2.59492	19.2053	0.20978439	0.169765	0.760622	0.00076879	0.68996626	0.18489263	0.1727162	0.07299427	0.16903905
25	fde	16	0.42105263	0	0	3.02631579	12.44238976	16.2105	1.59459	0.25867666	0.304017	0.038961	0.0227596	0.02328967	0.10770997	0.03345416	0.37020341	0.7456186
26	gas-florida	9	0.42857143	0.42857143	1	NA	1	1.90476	4.2	0.91651514	0	1	0.142857	0.6	0.34317694	0	0	0.27341154
27	highways-canada	65	0.43918919	0.43243243	0.56521739	6.96551724	1.15043045	2.17568	9.70068	0.79739098	0.0432099	0.925466	0.0241772	0.61737089	0.29390015	0.2866586	0.05626869	0.18375216
28	highways-mexico	256	0.5752809	0.09438202	0.06966618	3.98492462	2.33905254	3.09663	12.8198	0.6077168	0.157115	0.683599	0.0116712	0.58409321	0.27403195	0.25047236	0.09993063	0.175719
29	highways-usa	32185	0.59767874	0.04804085	0.03128151	3.94428032	2.83069977	3.27811	113.2	0.48781323	0.112882	0.744332	0.00010123	0.56992245	0.25270065	0.13025437	0.15913175	0.08140106
30	internet-bell	27246	0.3279845	0.26609768	0.41183251	3.88529106	1.24822695	2.9658	10.2393	0.03552594	0.166443	0.781363	0.00018019	0.74007116	0.09837849	0.10151527	0.08769932	0.14712559
31	internet-nlanr	818	0.14910682	0.10007291	0.20718232	3.06221646	1.46833721	4.22311	3.44613	0.04652686	0.694744	0.278401	0.100959	0.86380345	0.07639226	0.09490598	0.08726969	0.20362252
32	microwave-fcc	2291	0.35830466	0.33766031	0.60680023	7.82267442	0.79062192	2.21708	28.1337	0.35183621	0.0912139	0.884312	0.00088515	0.70853405	0.2335075	0.16410371	0.09620307	0.14811213
33	nts	47195	0.57302607	0.05342334	0.04277467	3.64577317	2.73529412	3.52487	58.9837	0.2055281	0.1791	0.602965	0.00009344	0.6383001	0.20432216	0.11296622	0.16552174	0.1262485
34	petroleum-mapsearch	49482	0.44403963	0.4179888	0.59273261	10.53617921	0.60191847	2.22659	79.5439	0.23828358	0.0574436	0.895568	0.00003273	0.60931902	0.225704	0.13913394	0.13074158	0.08978088
35	railways-canada	290	0.50086356	0.31606218	0.30894309	5.29261364	1.41677274	2.54922	19.7145	0.81930666	0.088489	0.822493	0.00748218	0.60688318	0.29223022	0.2882032	0.06855875	0.17700503
36	railways-mexico	54	0.49541284	0.28440367	0.34057971	4.58064516	1.44223581	2.53211	8.67593	0.83100338	0.128444	0.789855	0.037037	0.5844504	0.31426659	0.32297397	0.09737324	0.24937349
37	railways-usa	54662	0.51414651	0.30648256	0.31034249	6.666202	1.37918871	2.50134	176.861	0.54241634	0.149729	0.749695	0.00003412	0.57004798	0.25717161	0.16558612	0.11143502	0.07935939
38	roads-dallas	4656	0.57170923	0.12843811	0.08677306	4.58727221	2.38069624	3.20088	31.7756	0.35210718	0.105892	0.779039	0.00070287	0.58055998	0.22393958	0.19379136	0.12750465	0.14750894
39	roads-portland	44316	0.50143134	0.29432331	0.28553714	8.29348328	1.51495449	2.57312	76.0174	0.25570456	0.0400855	0.933064	0.0000463	0.57134507	0.21644248	0.19228369	0.14087548	0.09232806
40	telephone-lerg	5827	0.18661329	0.01546837	0.02999294	3.03310905	2.81738061	6.70988	7.67544	0.04343622	0.665619	0.101644	0.00510065	0.57072427	0.11086443	0.04414481	0.19995156	0.22073538
41	waterways-usa	2063	0.43809726	0.39583776	0.56812629	13.28571429	1.11637369	2.25993	27.1145	0.3951275	0.0422491	0.917309	0.00090347	0.63660056	0.24339118	0.18835539	0.03484885	0.13791308
42	air-tickets	364	0.56609642	0.02488336	0.00106943	3	72.06724722	151.24100000	1.83489000	0.07236100	0.84740200	0.00106943	0.16413000	0.17640320	0.06024672	0.04344031	0.50687532	0.80803064
43	gas-utah	43	0.40186916	0.42056075	1	NA	1	1.98131000	10.66040000	1.03057977	0	1	0.03385650	0.64429530	0.31731819	0	0	0.20593665