Internet Topology Data Kit

Young Hyun
CAIDA

ISMA 2011 AIMS-3 Workshop
Feb 9, 2011

Introduction: ITDK

* goals:
  * provide curated data for studying Internet topology
    * interface-, router-, and AS-level topology
  * employ best available measurement and analysis techniques
  * release 2-3 ITDKs per year

Introduction: ITDK

* motivation:
  * overwhelming amount of raw data
    * e.g., TB's of raw traceroute data over a decade
  * researchers often interested in derived data
    * e.g., AS level, not interface level
  * valuable for multiple researchers to study same dataset
    * build upon each other's work (explore different facets)
    * cross validation

History

* historical ITDK releases in 2002 and 2003

  * traceroute topology from skitter

* revived ITDK in 2010

  * three releases: Jan, Apr, and July 2010

  * traceroute topology from Ark

  * same goals but significantly different contents

Contents

* ITDK 2010-07 (July 2010 release, the latest)

  * router-level topology graphs

  * router-to-AS assignments

  * geographic locations of routers

  * DNS lookups of observed IP addresses

Contents: Topology

* router-level topology graphs

  * derived from IPv4 Routed /24 Topology Dataset

    * used two weeks of traceroutes to every routed /24
    * probed 8.25 million /24's from 45 monitors in 26 countries

  * resolved interfaces into routers by combining multiple techniques

    * iffinder (implements Mercator technique)
    * MIDAR (IP ID based technique)
    * kapar (extended APAR technique)

Contents: Topology

* router-level topology graphs

  * data:

    * interface addresses for each router
    * routers sharing each link
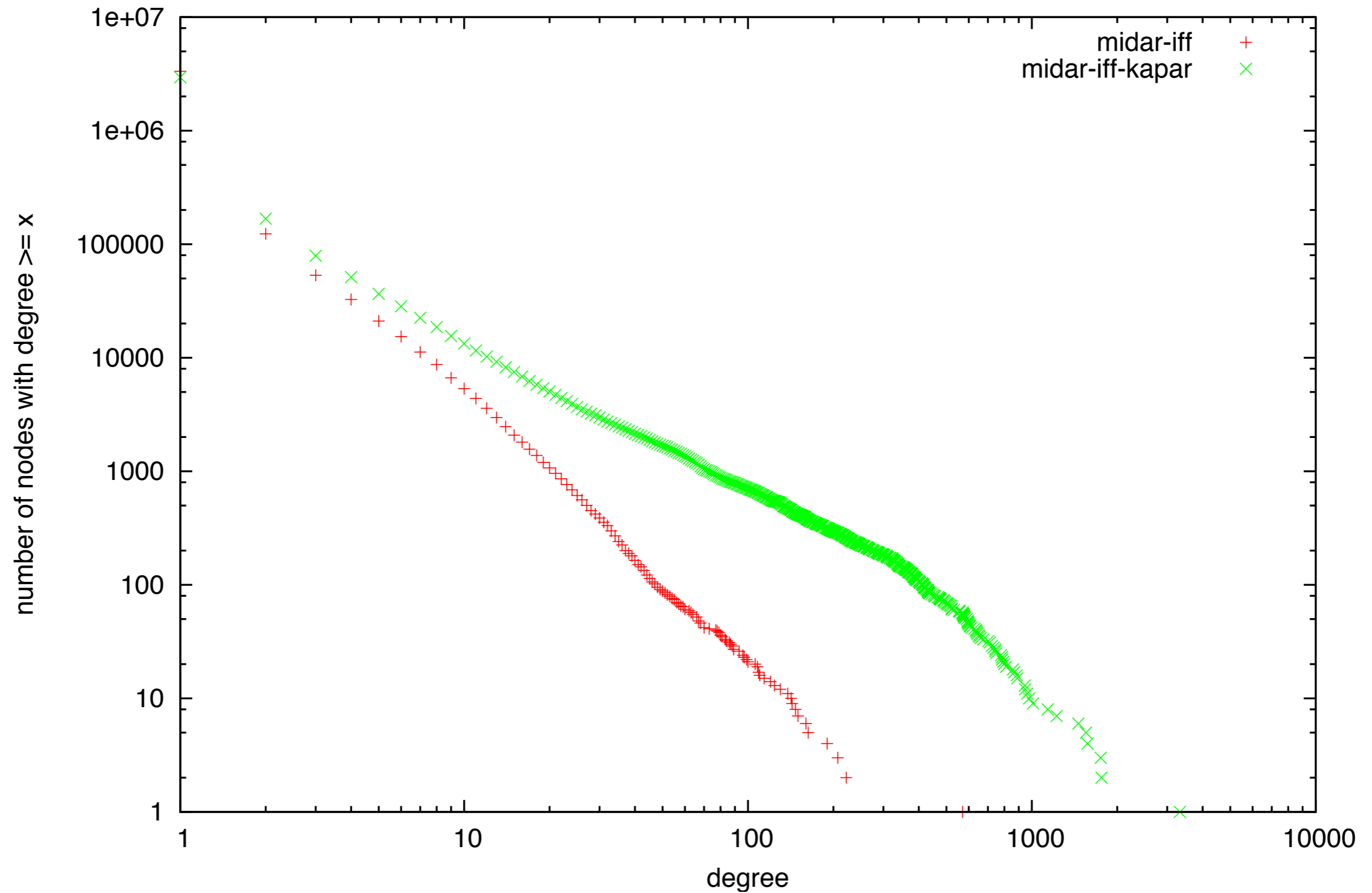   * may have >2 routers per link due to layer 2 and other causes

  * two graphs:

    * midar-iffinder: highest confidence alias resolution
    * midar-iffinder-kapar: more topology coverage but also more false positives

  * statistics:

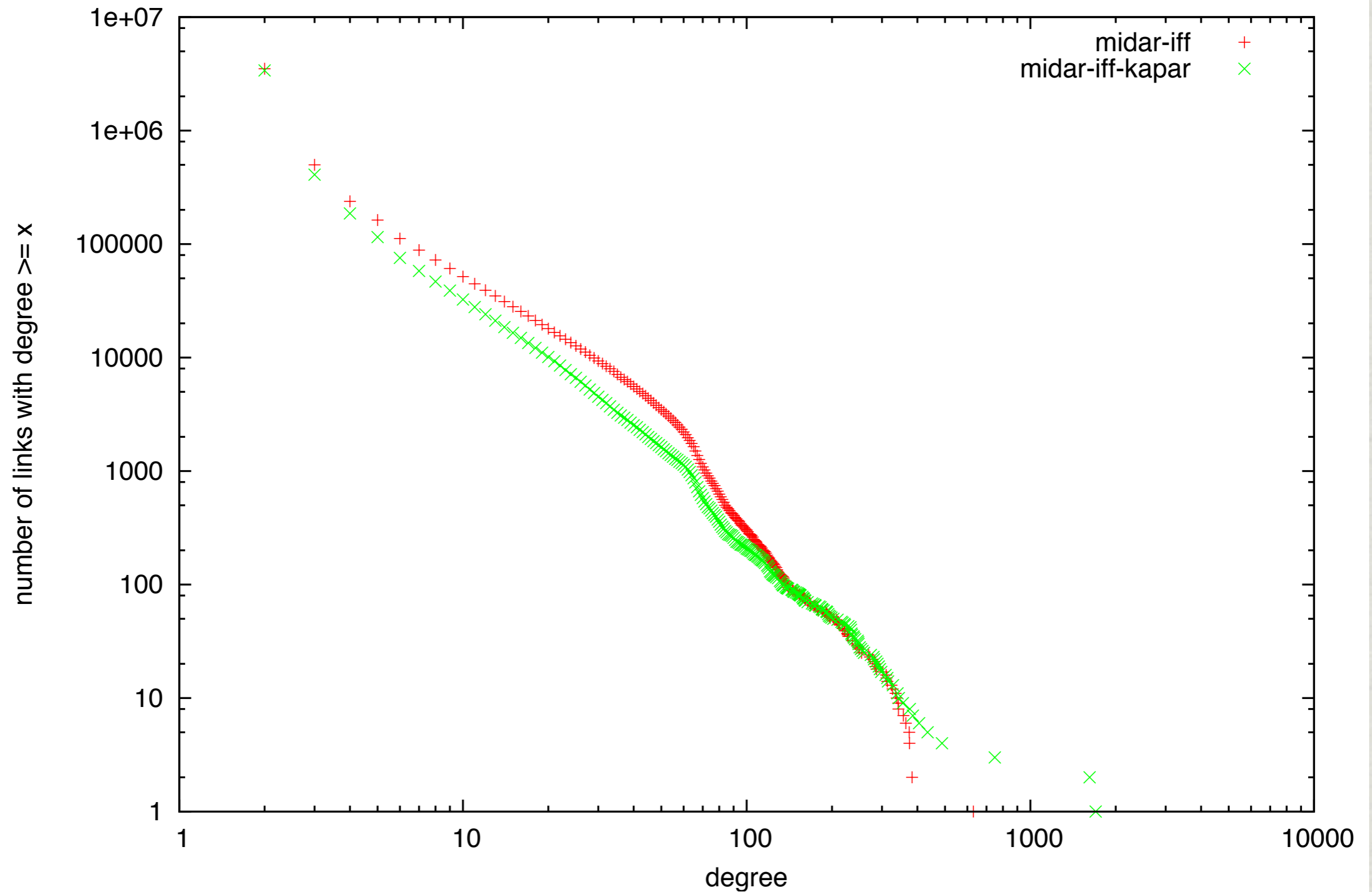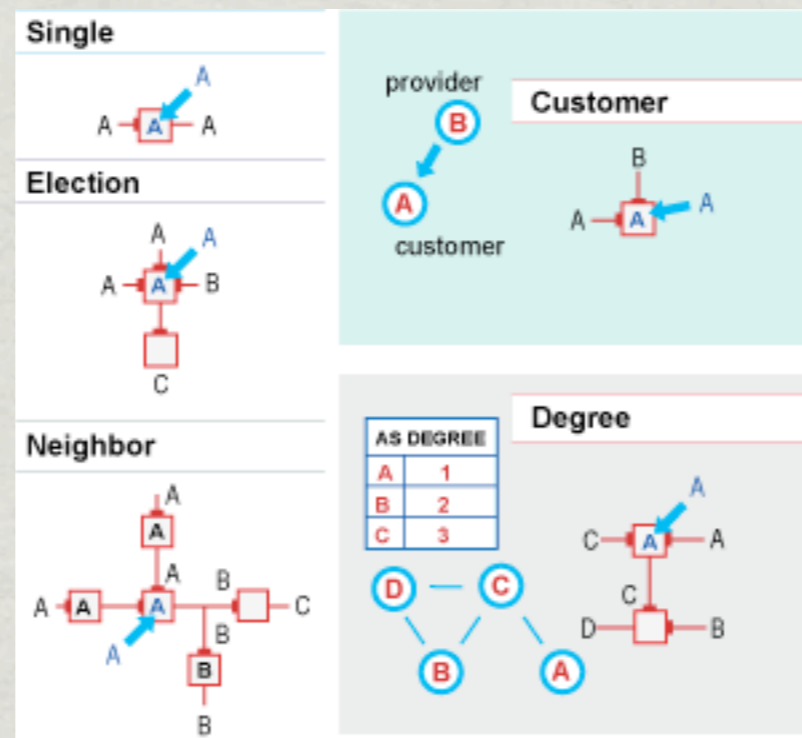|	nodes	links
midar-iffinder	3.3M	3.5M
midar-iffinder-kapar	3.0M	3.4M

Contents: Topology

ITDK 2010-07: node degree

Contents: Topology

ITDK 2010-07: link degree

Contents: AS Assignments

* goal: determine which AS owns each router

* algorithm:

  * Huffaker, et al, "Toward Topology Dualism: Improving the Accuracy of AS Annotations for Routers," in PAM 2010.

Contents: Geolocation

* geographic location (at city granularity) of routers in the router-level graphs

    * MaxMind's free GeoLite City database

* procedure:

    * map each interface on a router to a location

    * if all interfaces map to same location, then use that location

    * otherwise, no assigned location for router

Contents: DNS Lookups

* use HostDB, CAIDA's bulk DNS lookup service

* two datasets:

  * DNS lookups within days of observing an address in a traceroute path

  * DNS lookups during alias resolution runs

    • better matches alias resolution results

Future Work

* AS-level topology overlaid on router-level topology

* AS relationships

* IPv6 topology

Thanks!

For more information or to request data:
   www.caida.org/data/active/internet-topology-data-kit

   For questions: data-info@caida.org