

復旦大學

本科毕业论文



论文题目：自治系统社交网络的动态分析

姓 名：吴舒同 学 号：19307130284

院 系：计算机科学技术学院

专 业：计算机科学与技术

指导教师：王新 职 称：教授

单 位：复旦大学计算机科学技术学院

完成日期：2023 年 5 月 16 日

论文撰写人承诺书

本毕业论文是本人在导师指导下独立完成的，内容真实、可靠。本人在撰写毕业论文过程中不存在请人代写、抄袭或者剽窃他人作品、伪造或者篡改数据以及其他学位论文作假行为。

本人清楚知道学位论文作假行为将会导致行为人受到不授予/撤销学位、开除学籍等处理（处分）决定。本人如果被查证在撰写本毕业论文过程中存在学位论文作假行为，愿意接受学校依法作出的处理（处分）决定。

承诺人签名：吴舒同

日期： 2023 年 5 月 16 日

指导教师对论文学术规范的审查意见：

☐ 本人经过尽职审查，未发现毕业论文有学术不端行为。

☐ 本人经过尽职审查，发现毕业论文有如下学术不端行为：

指导教师签名： 日期： 20 年 月 日

指导教师评语：

答辩委员会（小组）评语：

签名：

20 年 月 日

签名：

20 年 月 日

学分

成绩

备注：

目录

摘要.....III

ABSTRACT.....IV

第一章 引言.....1

第二章 背景.....1

第三章 实验.....1

3.1 数据集.....1

3.2 实验环境.....2

3.3 数据总览.....3

 3.3.1 总量.....3

 3.3.2 连通分量.....5

 3.3.3 度分布.....6

3.4 宏观网络演变分析.....7

 3.4.1 节点量值.....7

 3.4.2 边量值.....11

 3.4.3 量值增减比.....14

 3.4.4 新增（减）边与点的关系.....16

 3.4.5 增减节点（边）的连接性.....21

3.5 自治系统层级的实例分析.....26

 3.5.1 各层级自治系统的度数变化.....26

 3.5.2 单个自治系统的自我中心网络.....32

3.6 以国家为单位的聚合分析.....35

 3.6.1 国内外协议总量.....35

 3.6.2 国家社群.....40

3.7 新冠疫情的影响分析及启示.....41

 3.7.1 总量.....41

 3.7.2 国家.....42

 3.7.3 运营商.....43

第四章 结论.....45

参考文献.....46

致谢.....48

摘要

自治系统是 BGP 协议中的基础单位。运营商使用它们签订商贸协议，给 BGP 网络路由引进人为规则。这些规则与政治、经济、社会因素相关，理解它可以帮助改善自治系统路由设施架构和协议设计。此前关于 BGP 社交网络的连通性研究很多，但缺少以经济贸易视角探索自治系统间关系的研究。本文使用 CAIDA 自治系统关系数据集，从贸易视角出发，挖掘了 2000 至 2023 年自治系统网络量值、波动、比率、集群等方面的规律，并结合案例分析现象背后的利益缘由。全文主要规律均由 R 进行可视化。除得到上述一系列结论外，我们也对新冠疫情对网络的影响进行了评估，揭示在当前互联网存量发展阶段的企业逻辑和未来趋势。

关键词：自治系统，BGP 协议，社交网络

ABSTRACT

An autonomous system(AS) is the basic unit in the BGP protocol. Internet service providers sign commercial agreements for them, introducing human-made rules to the BGP network routing. These rules subject to political, economic, and social influences. Understanding them can help improve the autonomous system routing infrastructure and protocol design. Many studies have been conducted on the connectivity of BGP social networks, but there is a lack of research exploring the relationships between autonomous systems from an economic and trade perspective. This article uses the CAIDA autonomous system relationship dataset to explore the rules of autonomous system networks' quantity, volatility, ratio, clusters, and other aspects from a trade perspective from 2000 to 2023. We also inspect the chain of interests behind the phenomenon through case studies. Charts in the article are visualized using R. In addition to obtaining the above results, we also evaluated the impact of the COVID-19 pandemic on the network, revealing the corporate logic and future trends of the Internet in the current stage of development.

Key words: autonomous system, BGP protocol, social networks

第一章 引言

由于自治系统研究对于互联网生态的重要性，不同学科，组织均对这一课题进行过研究。较早的研究如 [1]，文章中研究者使用的是一个较为流行的数据集: NLANR 组织收集和维持的 BGP 路由表信息。他们通过解析路由表得到自治系统数据，基于此剖析了 1994 至 2000 年自治系统的拓扑结构特点，收集得到了一系列符合幂定律 (power-law) 的统计学特征。较新的研究有 [2]，基于与前述类似的数据处理方法，这篇文章覆盖了 2000-2020 年的自治系统网络分析结果，得到了网络发展演变逻辑方面的若干结论，并对于国家级别的拓扑逻辑进行了分析。除了宏观层面的网络分析外，亦有文章如 [3] 从底层案例出发，生动地阐述了自治系统偏低层构成部分的行为特征。

以上若干文章的数据初始形态均为原始 BGP 路由数据，研究者对其进行处理后方得到自治系统层面数据，因此带来因数据处理而生的不确定性，这对结果的验证，其他研究者对其实验的复现均带来了额外的难度。在本篇论文中，使用的数据集直接为处理后的自治系统关系数据，因而相较于前述研究去除了这份不确定性。此外，前述论文均主要针对自治系统的网络特性，即连通性特征进行研究，着眼于宏观的拓扑结构，或部件间的互连指标。而对于自治系统的贸易特征，亦或没有提及，亦或加以淡化。本篇论文则以自治系统的贸易关系为着眼点，从数据处理到模型建立，均致力于从利益、贸易角度探索规律，在需要时也兼顾部分网络特征研究。同时，由于本篇论文成文时世界各国已完

整经历新型冠状病毒疫情对经济、政治等领域的影响，研究中也涉及疫情对于自治系统贸易演变的影响分析。

本篇论文主体部分将采用如下结构:

1. 对数据集的介绍（包括所述机构，收集方法，时间段，以及自治系统推演算法）
2. 实验环境（机器，编程语言，画图语言）
3. 数据总览
4. 宏观网络演变分析
5. 自治系统层级的实例分析
6. 以国家为单位的聚合分析
7. 新冠疫情的影响分析及启示

第二章 背景

互联网由成千上万个互联网服务提供商(ISP)组成。这些运营商拥有和控制自己名下的自治系统,构成 BGP 中的基础单位。ISP 间存在沟通交流,以便协同和普遍地路由互联网流量,而这种沟通交流最终通常以商业协议的形式展现在网络中。这些商业协议的出发点可能来自利益链、政治因素等等,落实在协议内容中则成为网络中人为引入的路由规则。正因这类政策的存在,自治系统间的网络路由才变得复杂,不定,和难以预测。

研究自治系统间的关系,对 BGP 协议,对互联网的发展,都有重要意义。通过对它的研究,我们可以找到现有路由机制的弊端从而优化网络路径;我们可以通过评估自治系统关系对网络基础设施的影响,来构建更有扩展性的互联网增长模型架构;也可以通过它对现有同行和上游服务商的表现进行评估,从而帮助服务商进行选择。

然而,出于隐私,商业机密等等考虑,直接研究自治系统被证明是非常困难的,因为迄今为止并没有服务商或第三方机构提供公开透明的自治系统协议数据。这些数据只有参与协议运营商知晓。因此,在无法直接获得自治系统数据的情况下,研究者转而选择使用可以公开收集到的 BGP 路由数据,用算法处理后得到计算出的自治系统贸易数据。

关于这类推导算法的研究最早由 L.Gao 在 [4] 中阐述。在她的研究中,将 ISP 间(也即自治系统间)的商业关系分为三类:顾客-提供商型 c2p

(customer-to-provider)，同行型 p2p (peer-to-peer) 和同辈型 s2s (sibling-to-sibling)。这种分类的依据在于自治系统间需要通过购买或商业协定来借助其他自治系统进行网络包传输。在顾客-提供商型中，这种关系的箭头方向标志着资金流动的方向：处于低位的顾客自治系统向处于高位的提供商自治系统付费，以换取提供商自治系统使用其资源帮助自己传输数据包。同行型协议中的双方达成协议，互相帮助对方传输对方本身及其顾客的数据包，以节省服务购买的开销。同辈型关系中的自治系统往往隶属于同一管辖域（如同个 ISP），这类关系通常象征着自治系统团体的合并、收购。Gao 设计算法的根据思想在于，对于一条有效的自治系统贸易链条，每一跳中的中转节点（即帮助其他自治系统进行数据传输的自治系统）都应在本条链条中存在相邻节点对其服务付费。而若遵守这一思想，则所有合理的链条路径必然遵守如下模式：零到多个顾客-提供商型边，紧接着零到多个同行型边，紧接着零到多个提供商-顾客型边。同辈型边可以在任意位置出现，因为这种边相当于内部路径，不代表贸易关系。如果将顾客看作低位节点，而提供商看作高位节点，则这样的模式路径将是一条爬坡后下降的峰值曲线，爬升只有一次，一旦下降后不能再次爬升。而根据自治系统网络的经验规律，在这条峰值曲线顶端的节点，往往是本条链路中度数(degree)最大的节点。因此，对于一条 BGP 路由链路，我们就可以通过先剥取顶峰节点，来逐个推断出其余节点的贸易关系。

自 Gao 的研究后，后续涌现出多位相似研究。Subramanian[5] 等人继承了上述有效路径判定方法，但将 AS 关系推断问题制定为组合优化问题：给定从一组 BGP 路径 P 导出的无向图 G，为 G 中的每条边分配边类型（c2p 或 p2p），使 P 中有效路径的总数最大化。该问题被称为关系类型（ToR）问题，

且后被 Di Battista[6] 等人和 Erlebach[7] 等人独立地证明为 NP 完全问题。在本篇论文中，使用的数据集使用的推演算法类似于 “AS Relationships, Customer Cones, and Validation”[8] 中的研究，它亦是对 Gao 工作的继承，使用一组事先设定的条件进行边关系的推演。

每个自治系统有一个唯一的数字编号，称为自治系统编号(ASN)。自治系统编号由 The Internet Assigned Numbers Authority (IANA)统一管理。当 ISP 申请新自治系统编号时，将经由 IANA 获批得到一个数字。值得注意的是，该数字编号存在被重复利用的可能性。若一个 ISP 决定抛弃自己旗下的自治系统，那么 IANA 将回收这一自治系统编号，并可能在新的 ISP 申请时重新发放这个数字。因此，当进行自治系统编号与自治系统信息匹配时，需要充分考虑编号替换的可能性。

第三章 实验

3.1 数据集

贡献本文所使用的数据的机构为 Center for Applied Internet Data Analysis (CAIDA)。CAIDA 成立于 1997 年，是加州大学圣地亚哥分校超级计算机中心所下属的计算机网络数据科研中心，主持分享许多与计算机网络协议有关的大数据集。这些数据集均为公开。其中，本篇论文所用数据集为 “The CAIDA AS Relationships Dataset” [9]，它是一个由 BGP 路由数据推理出的自治系统关系数据集，关于推理算法的介绍见“第二章 背景”部分。CAIDA 所共享的数据自 1998 年开始，每月一份，最新的数据收集于 2023 年 5 月，即本文写成时。数据名开头为年份月份，结尾均为“01”，如“20230501”等。以下均用上述编号来表示对应的数据。每份数据文件格式一致，各行为以自治系统编号表示的顾客-提供商型边和同行型边，最小的有 5772 行（19980101），最大的有 380544 行（20230101）。本文使用的数据主要为 2000 至 2023 年每年 1 月的数据。

除了上述主要数据集外，匹配自治系统编号和自治系统信息的数据来自于 [10]，它将每个自治系统编号对应至当前所属 ISP，国家（或地区）。本文所使用的匹配信息拉取于 2023 年 2 月 28 号。

如前所述，自治系统编号存在重复利用的可能性。匹配数据仅提供最新数

据，而 CAIDA 的数据则截录了几十年来的历史数据。因此，我们目前将自治系统编号所匹配上的自治系统信息未必是正确的。为了避免受该问题影响，本文涉及自治系统国别和运营商信息的分析中使用的自治系统均只选择在研究期限内持续存在的编号。若某编号在每次数据中不存在，则将其排除不进行进一步分析。但在统计数据总量等与自治系统运营信息无关的研究时，编号仍然会被考虑。

3.2 实验环境

整个实验在两台机器上完成。主机器运行 Archlinux，副机运行 Ubuntu，大部分计算量大的图分析程序在副机上运行。由于本实验中不涉及算法性能、速度评估等，因此对机器的运算资源不作描述。

实验所用程序主要使用 Python 语言，调用 Python-Easygraph 和 NetworkX 图分析库进行图解析。数据解压处理设计少量 Bash 脚本。绘图使用 R 语言。IDE 环境为 vscode。

3.3 数据总览

3.3.1 总量

从 2000 至 2023 年，逐年自治系统(AS)和自治系统协议(Autonomous System Relationship, 缩写 ASR)的总数波动情况见图 1。其中自治系统总量为数据文件中出现的不同自治系统编号个数，而自治系统协议总量即为数据文

件的总有效行数。在图理论中,可将自治系统看作节点,将自治系统协议看作边。由于协议边分为两类:提供商-顾客型边(注意!文中数据处理后顾客与提供商关系展现为 p2c,而非前面出现的 c2p,其实只是方向调换,并无其他区别)和同行型边,在构图时作如下处理:对于提供商-顾客型边,画一条单向边。对于同行型边,分别画两条互相指向的单向边。

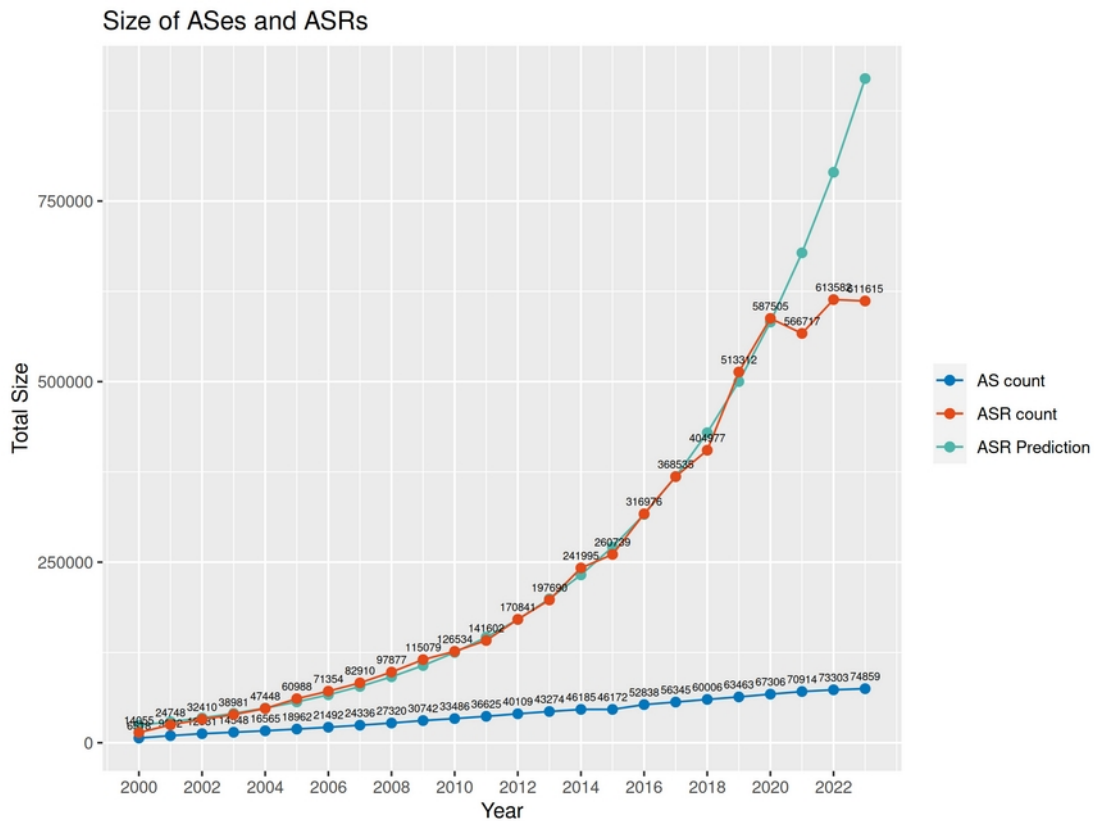


图 1 AS 与 ASR 总量变化 2000-2023

图中自治系统总量(AS count)似乎是一条线性增长线。若将 2000 年视作 0, 2023 年视作 23,则自治系统总量(y)与时间 (x) 的关系可被如下公式描绘:

$$y = 3056.05x + 4346.29$$

以上公式的 R-square 值为 0.9944879055571465。

而另一方面,图中自治系统协议总量(ASR count)在 2021 年前则似乎符合

指数增长。若使用同样的年份转换方法，并以指数函数去匹配数据，则自治系统协议总量(y)与时间(x)的关系可被如下公式描绘:

$$y = 28413.58*(0.15^x) - 4049.77$$

以上公式的 R-square 值为 0.86。注意上述公式计算时去除了 2021-2023 年的数据。图中绿色的"ASR Prediction"曲线即为根据上述公式拟合后得到的自治系统协议总量值。

2021 至 2023 年，虽然自治系统总量的增长趋势似乎未受影响，自治系统协议总量则产生了较大的颠簸。不难联想这一波动的主要原因即是新冠疫情影响。对于疫情影响的分析见“3.7 新冠疫情的影响及启示”，此处不多赘述。

3.3.2 连通分量

在 2000 至 2023 年，除 2005,2007,2008,2010,2011,2020 外，其余年份的连通分量均为 1 个。而在上述非连通年份中，除 2005 年外，其余年份的连通分量均为 2。这其中断裂的，不与主要连通分量相连接的，除却一些较小规模，可能为私人机构使用的小自治系统外（连通分量大小为 2），另有两个较大规模的连通分量，分别为自治系统 1 和 8 所属的连通分量。自治系统 1 是 Level 3 Communications, LLC (原名 LVLT-1)机构下属的自治系统，是互联网中最老，知名度最高的自治系统之一，也曾作为主干网的成分，影响极大一部分网络。但在 2003 年后，其进出度（此处进出度即指与其有协议的自治系统个数）急剧下降，此后出现数次其所在连通分支不与主连通分支相连的情况。而自治系统 8 所属于美国莱斯大学(Rice University)，是该校校园网的主

干节点，主要与其他学术研究机构相连，在 2005 与 2007 年也出现了断连的情况。

相比于主连通分量，这些断裂的小连通分量的规模几乎可以忽略。但它们的存在反映了在自治系统层面的网络，依然存在非主流使用的案例。

3.3.3 度分布

已有多项前人的研究指出，网络级别的图，其节点度数（各类型协议数）分布符合长尾效应。而在自治系统网络中，这一效应同样明显。图 2、3 分别是 2000 年和 2023 年 1 月的自治系统节点度数的 CDF 分布。图 3 中可以看出 90% 的自治系统度数小于 20，而图 1 几乎所有节点度数均小于 20。事实上，因为疫情影响，2023 年超大型自治系统的规模已经小了许多，因此曲线的陡度才略小于图 2。在这期间的 20 年间，图 2 所展现的才是常态。

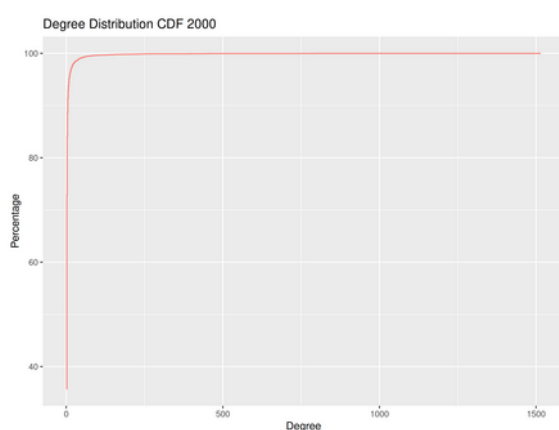


图 2 2000 年 AS 度数分布 CDF

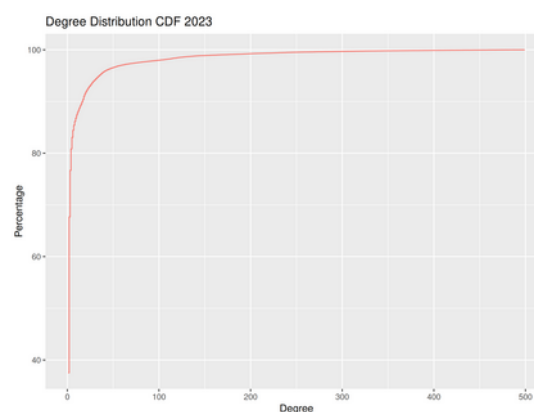


图 3 2023 年 AS 度数分布 CDF

我们也计算了每年的平均节点度数，它从 1998 年的 4.90，一路升至 2022 年的 16.74，在 2023 年首次相较有所下降（16.34）。这一度数的提升主要为

超大型自治系统膨胀的结果。

3.4 宏观网络演变分析

3.4.1 节点量值

此处的节点（node）即自治系统。自治系统总数的逐年增减变化图见图4。图中AS即为“自治系统”的缩写。

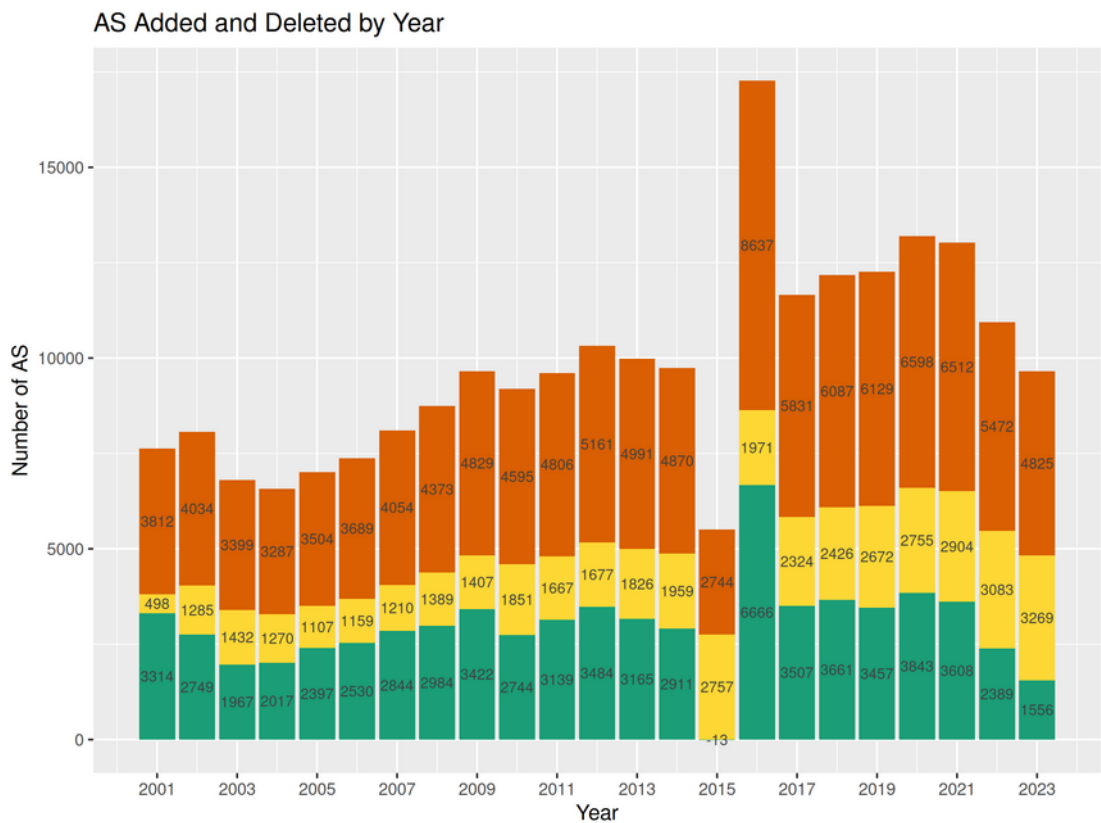


图4 AS 逐年增减量 2001-2023

图中，最上方的橙色色块为增值，中间的黄色色块为减值，下方的绿色色块为净增值。每个色块的具体数值显示在中央，色块长度与数值大小成比例。对于净增值，也即每年净新增自治系统的数量，排除极端年份后稳定在区间

[1500,3900]内，而网络的节点总数自 2002 年以达到了万级，因此其数值波动相较于总量并不明显。这一观察与 3.3.1 中自治系统的线性增长曲线是相符的。

相对稳定的净增值也意味着增值和减值并不稳定。图中可以看出橙色和黄色色块均有逐年增长的趋势。而鉴于二者的插值较为平均，我们可以推论增值的总增长幅度是相似的。这表明自治系统的增减并非全然是市场兴衰的指标，在网络贸易兴隆的年间，增减幅度都会相应上升。事实上，运营商选择撤去一个自治系统除了打算丢弃相关业务外，也完全可能是出于架构设计而进行的合并。这样能够更好地解释自治系统增减值之间的正相关性。

2015 年是图中唯一一个净增值为负的年份。由于图中增减值为 1 月的数据，且是相对前一年的变化值，所以 2015 年的数值实际上是由 2014 年后期的变化形成的。

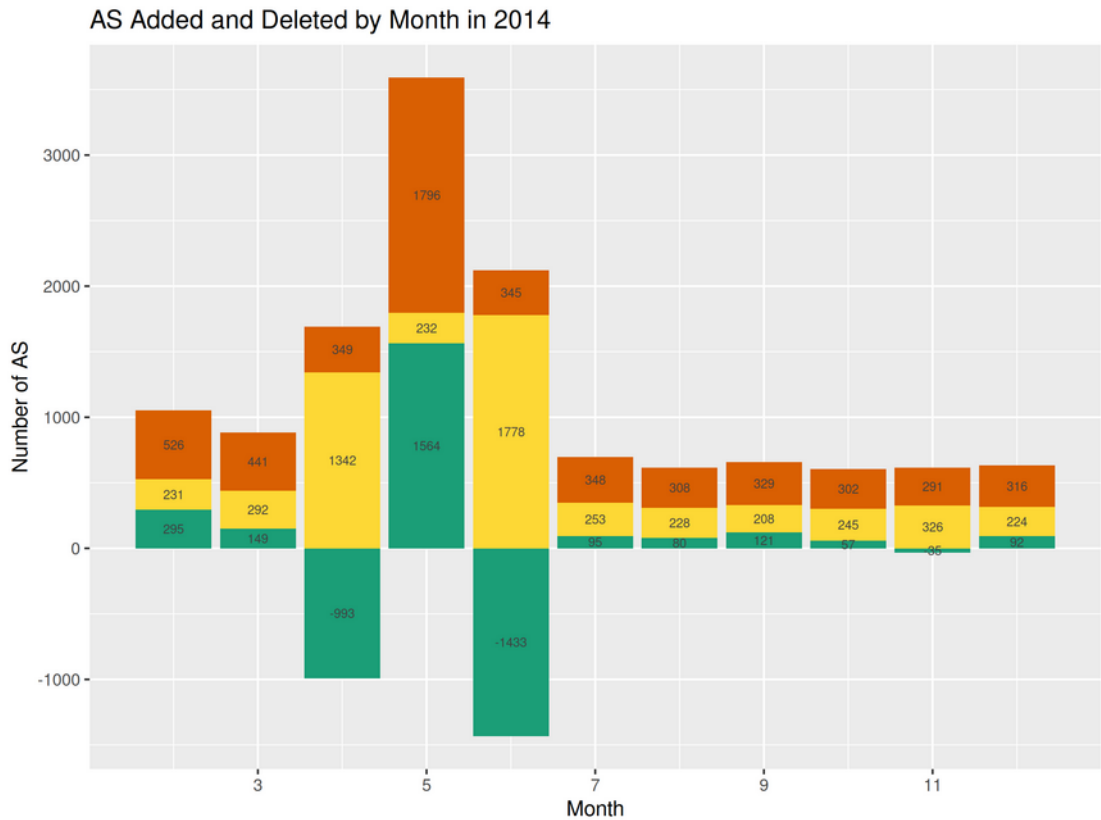


图 5 2014 AS 逐月增减值

图 5 为 2014 年各月份自治系统的增减值情况，绘制方法与图 4 一致。图中 4,5,6 月份数据产生了极大颠簸。4 月减值是正常值的 4 倍左右，5 月增值是正常值的 4-5 倍，6 月减值是正常值的 5-6 倍。这样的数据波动有两种可能。一是事实情况下确实发生了自治系统的大规模删减重组，二是由于统计操作导致的数据波动。提出第二种可能性的根据在于，图中数值均为相对于前一个月的变化值。若某个月的数据在收集过程中遗漏丢失，而下个月恢复正常，那么图中的报复性反弹就可以解释（尤其考虑到数据原始形态为 BGP 路由）。但是图中的情况无法完全被数据收集漏差解释：若 4 月数据存在遗漏，5 月数据收集正常，因此 4、5 月份间产生落差，那么 6 月的峰值只能再次用数据漏差解释。但在随后的 7 到 12 月份中，反弹并未出现，并且其数值大小持续偏小。

从图 4 可以看到，网络每年的数值变化量（包括增减值和净增值）是逐年增大的。2013 年的逐月自治系统增值平均值为 515，即使早在 2010 年这一均值也达到了 478。而 2014 年 7 到 12 月的增值数值在区间[291,348]间浮动，属于偏小的状态。如果继续看 2015 年的数据图（图 5-1），我们会发现在 2015 年年初，增值数据猛增，而在随后的月份，各项数值均恢复到正常水平。

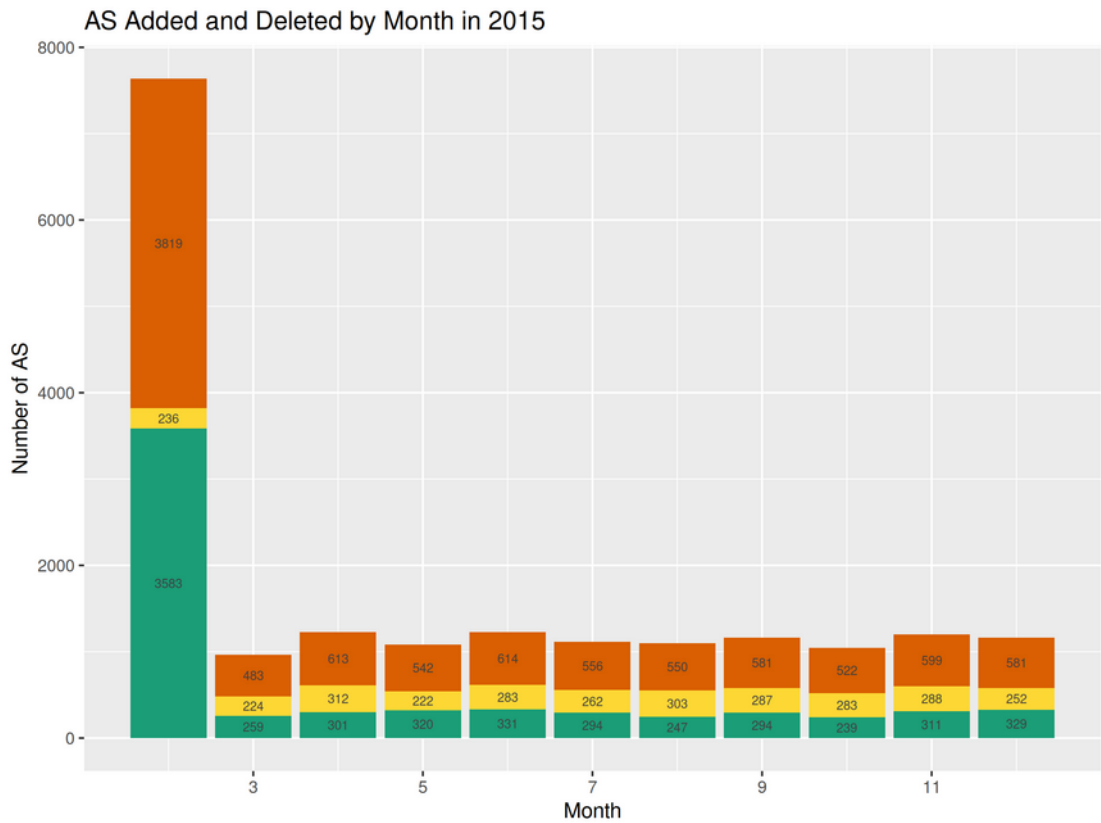


图 5-1 2015 AS 逐月增减值

鉴于上述分析，2014 年 4 至 6 月的颠簸并不能被数据收集误差解释，而是事实情况。且结合 2015 年数据，我们发现对于自治系统数量，存在波动和弹性。一次异常峰值往往紧伴着一次异常的谷值，以将平均数值拉平。

那么，在颠簸的月份中究竟发生了什么？通过统计月份间节点重合度我们发现，4 月删去的节点（1342 个）中，有 1120 个在 5 月重新出现，而 5 月份增加的众多节点（1796 个）中，有 1555 个在 6 月消失。6 月删去的节点

(1778 个) 中, 有 1466 个在 2015 年 2 月重新出现。一个可能的猜想是, 这是运营商的“维护”时段。可能出于架构升级或重组的目的, 运营商将其旗下的自治系统暂时回收, 待维护完后重新放出。至于为何全球的运营商均集中在 2014 年进行维护, 我们不能肯定。也许与互联网协议变更更有关联。

自治系统数量的增减似乎与时间并无显著关联。我们尝试探索了其与月份的联系, 猜测也许在某些月份会有规律性浮动, 或按照一年周期浮动, 但并未观察到类似情况。可以说自治系统增减与时间的关联性并不强。

3.4.2 边量值

此处的边 (edge) 指自治系统间所签订的协议。若将自治系统关系网看作一张有向图, 那么自治系统间签订的协议则构成节点间的边。本数据集中的协议包括提供商-顾客型和同行型。对于提供商-顾客型边, 看作由提供商自治系统指向顾客自治系统的一条单向边。对于同行型边, 视作作为同行双方的节点互相指向的两条单向边。所有边没有权重, 或可看做权重均为一。

本实验中建模使用有向边, 是出于研究贸易关系考虑。贸易双方节点处于不同的位置, 这种差异使用无向边无法表示。而另一方面, 在社群分析等涉及连通性的算法中, 则需要淡化边的方向, 此时我们把有向图转为无向图。有向边 (提供商-顾客) 的方向可以理解为服务流动的方向。

2001 至 2023 年的协议签订总量相较前一年同期的变化图见于图 6。图中“ASR”即为“自治系统协议”的缩写。

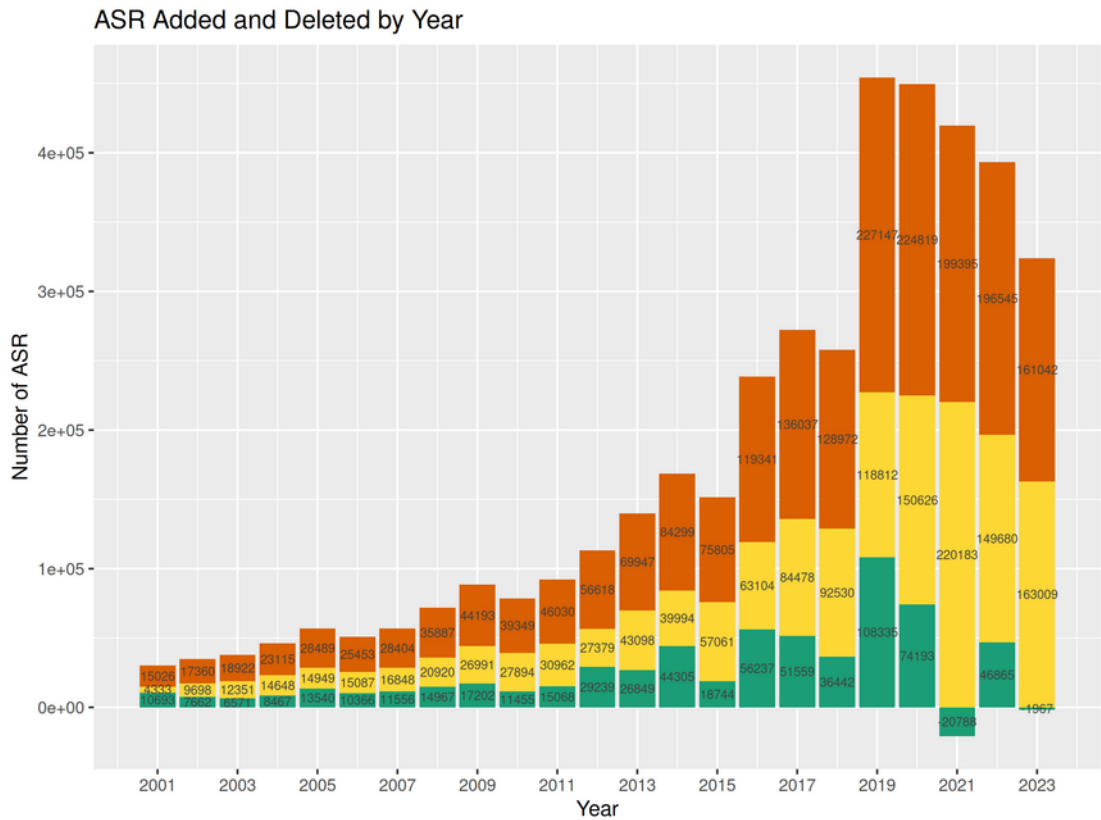


图 6 ASR 逐年增减值 2001-2023

图中，最上方的橙色色块为增值，中间的黄色色块为减值，下方的绿色色块为净增值。每个色块的具体数值显示在中央，色块长度与数值大小成比例。相较 3.4.1 中节点的波动图图 4,边也即协议量的增长趋势更为明显。与节点基本稳定的净增值不同，绿色色块代表的净增值有持续增加的趋势。在 3.3.1 中我们通过计算印证了自治系统协议的增长基本符合指数增长，也与这里的现象符合。

在节点数减小的 2015 年，边数仍然保持了增加，但增势稍有放缓。而在疫情影响的 2021-2023 三年，边数在 20 年来首次减少（2021 与 2023 年净减）。疫情对自治系统协议量的影响远大于对自治系统量的影响。

按年为单位看，边数总体是稳定上涨的。但若以月为单位，我们发现边数的变化比节点变化频繁许多。

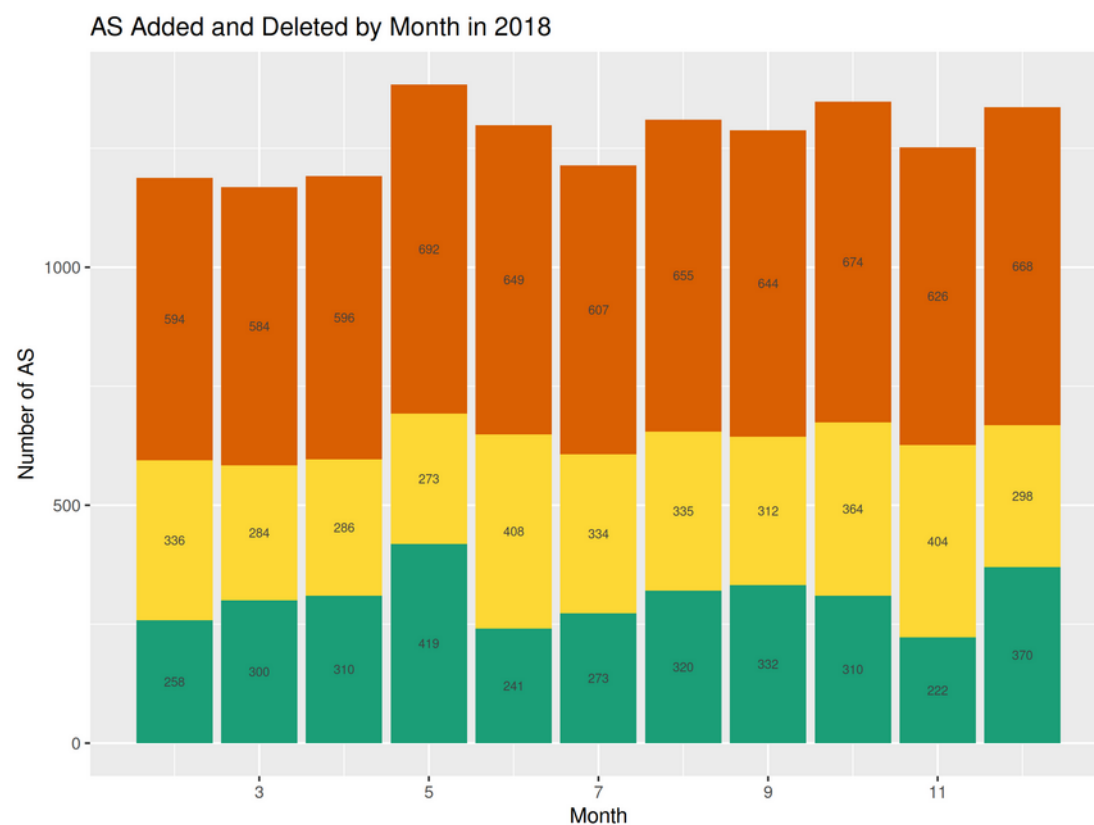


图 7 2018 AS 逐月增减值

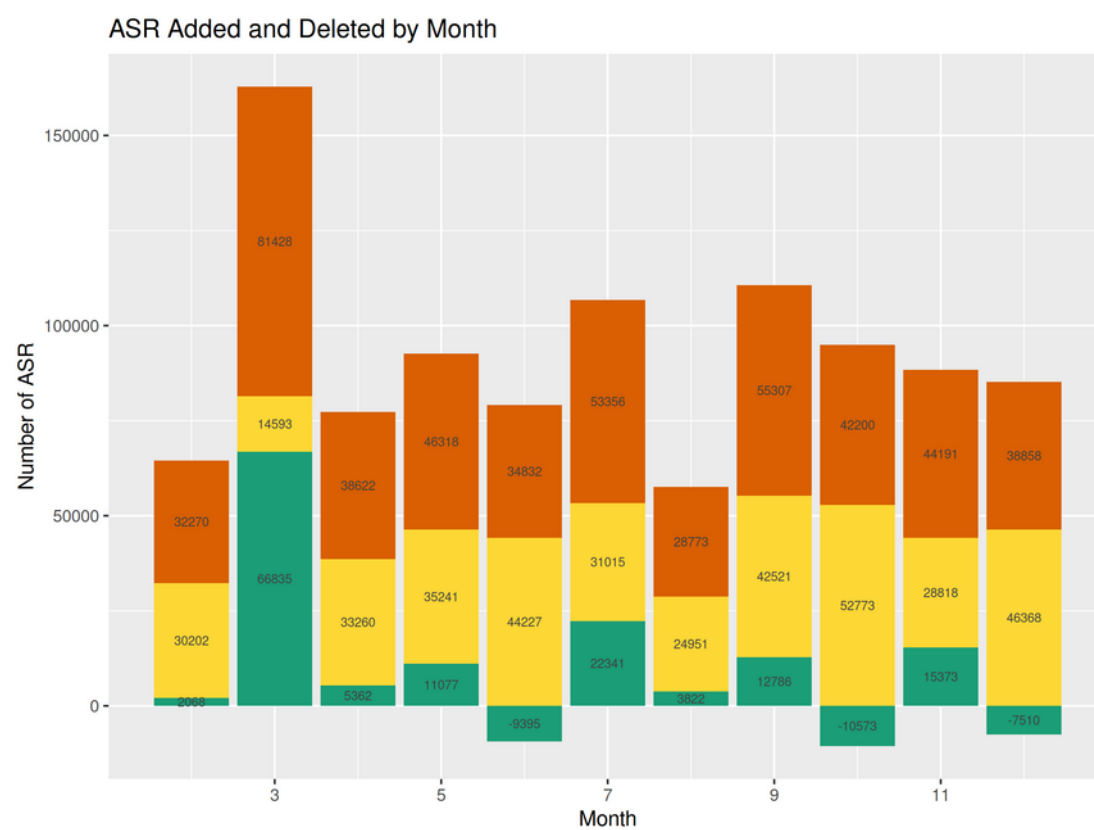


图8 2018 ASR 逐月增减值

图7、8分别为2018年节点数的按月变化，和边数的按月变化图。其余年份与上述二图相差不大。通过对比我们可以看出：虽然节点数稳定增加，但边数依然频繁出现突增突减；且随着节点数的增多，边数的增量并未随之增加（图8橙色色块）。相较于节点，边的波动更为频繁，幅度也更大。这一观察与现实情况是相符的。启动和取缔一个自治系统所消耗的资源、精力要大于签订协议耗费的精力，运营商需要申请编号，配置环境，加入路由等。而自治系统间协议本身就是灵活的。运营商在不想维系原先的合约时可以重新选择或通过谈判得到满意的服务，与潜在的利润相比，更换的代价相对较小。运营商在面临结构调整的情况下，必然会优先考虑调整自治系统间的协议，而尽量避免自治系统级的更替。

3.4.3 量值增减比

上述关于节点和边的量值趋势图中，我们可以大致看出二者总量都在稳定上升。在这一节中，我们来看一下这20年间增长趋势的变化。

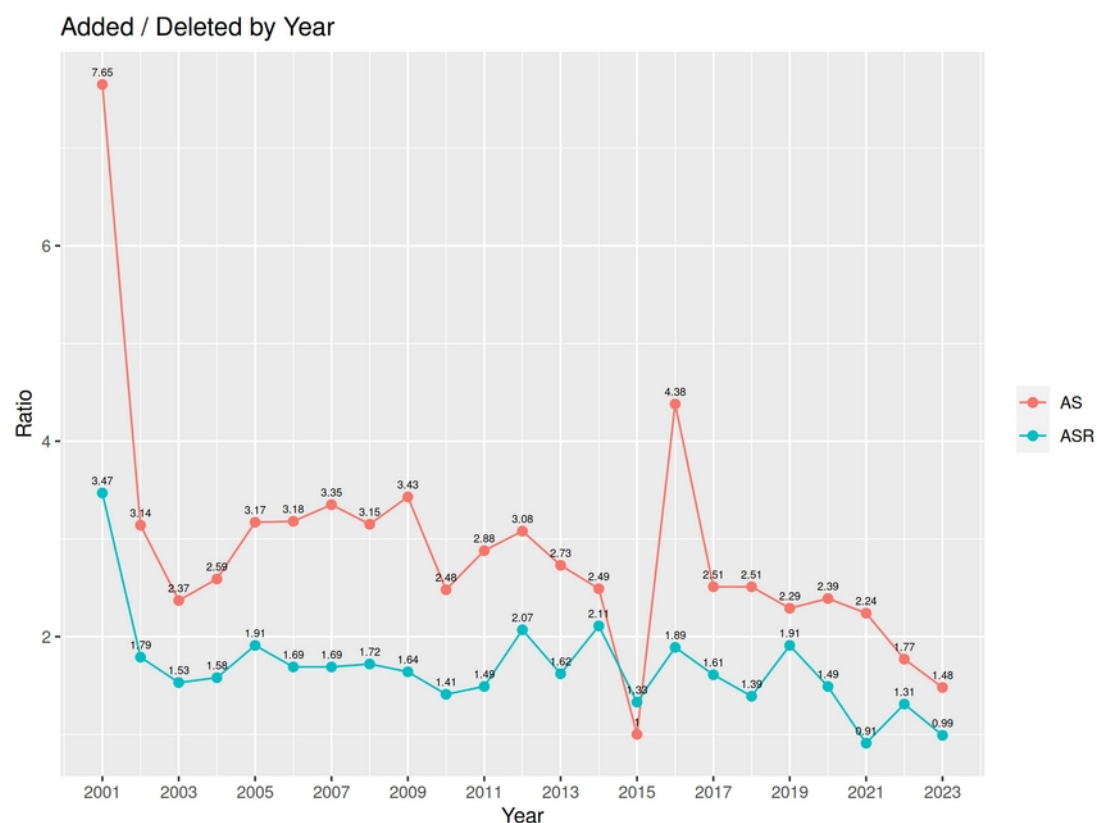


图9 AS 与 ASR 逐年增减量比值 2001-2023

图9计算了2001至2023年每年节点（边）的增量与减量的比值。红色为节点（AS）的变化曲线，蓝色为边（ASR）的变化曲线。比值越大，增长趋势越大。当比值为1时，总量维持不变。

总体来说，两条曲线多年间均在缓慢下降。这说明随着互联网发展成熟，开发新地区，拓展新网络的势头渐渐殆尽。这一点在后面章节绘制的世界自治系统社群图中也可以看出。而代表边的蓝线多数都低于代表节点的红线，表明调整（取缔）协议的操作要远多于调整节点的操作（我们已经知道新增协议数要远高于新增节点数，因此比值的落差来自删减协议数）。在2021年后，协议增减比值首次破1，象征着节点间关系的调整成为运营商的新重心。

3.4.4 新增（减）边与点的关系

当一个自治系统被删去，其所有相关的协议也将一并消失。同样，当新增一个自治系统，它也需要通过其新减的协议关系被路由表所知晓。那么对于新增或新减的协议，它们与同期新增减的自治系统关联度如何？运营商是更倾向于与新增节点签订协议，而是选择原先节点？

为了计算这个问题，我们首先选取一个基准时间(pivot time)，记为 P。随后我们选择一个在 P 之后的进行比对的时间，记为 N。P 时间所对应的自治系统集合，也即节点集合，记为 Node_P。对应的自治系统协议集合，也即边集，记为 Edge_P。同样，N 时间对应的节点集记为 Node_N,边集记为 Edge_N。通过计算集合差，我们可以得到自时间 P 至 N 的新增节点集和新增边集，即

$$\text{added_Node} = \text{Node_N} - \text{Node_P}$$

$$\text{added_Edge} = \text{Edge_N} - \text{Edge_P}$$

接下来我们需要统计 added_Edge 中，有多大比例的端点(end point,即边所连接的节点)处于 added_Node 中。为此，我们分别统计两端点均在 added_Node 中(ASR with both ends in added AS)的边数 N_both，及只有一端节点处于 added_Node 中(ASR with one end in added AS)的边数 N_one, 并计算它与新增边集 added_Edge 长度的比值：

$$r_both = N_both / \text{length}(\text{added_Edge})$$

$$r_one = N_one / \text{length}(\text{added_Edge})$$

得到端点位于新增节点中的新加入边的比例 r 后，我们还需要计算随机状态下

新边位于新加入点中的概率，也即计算基准值(pivot)。首先我们计算新加入点在总节点集中的比例 r_new_node :

$$r_new_node = \text{length}(\text{added_Node}) / \text{length}(\text{Node_N})$$

若要构造一条随机新边，那么其一端节点落于新增节点的概率应等于 r_new_node 。另一端节点也在新增节点内的概率相同。因此，对于两端点均在 added_Node 中的基准值，即为

$$\text{pivot_for_both} = r_new_node^2$$

而对于一端处在新增节点，另一端否的情况，可以类似推理算出为

$$\text{pivot_for_one} = 2*(r_new_node)*(1-r_new_node)$$

以上四个公式即为判断新增边与新增点关联度的曲线函数。对于新减，计算方法与上完全相同，只是将新增集合替换为新减集合。

注意上述过程有一个非平凡的影响量，即基准时间 P 的选取。由于公式计算的是累计值，因此基准值应尽量取早。本实验出于绘图清晰度考虑，选择了 20 年的观测区间，绘制了 2002-2022 和 2000-2020 的图，此处展示前者，但反映的规律一致。

图 10 是新增边位于新增节点中比例变化曲线。时间基准为 2002 年。

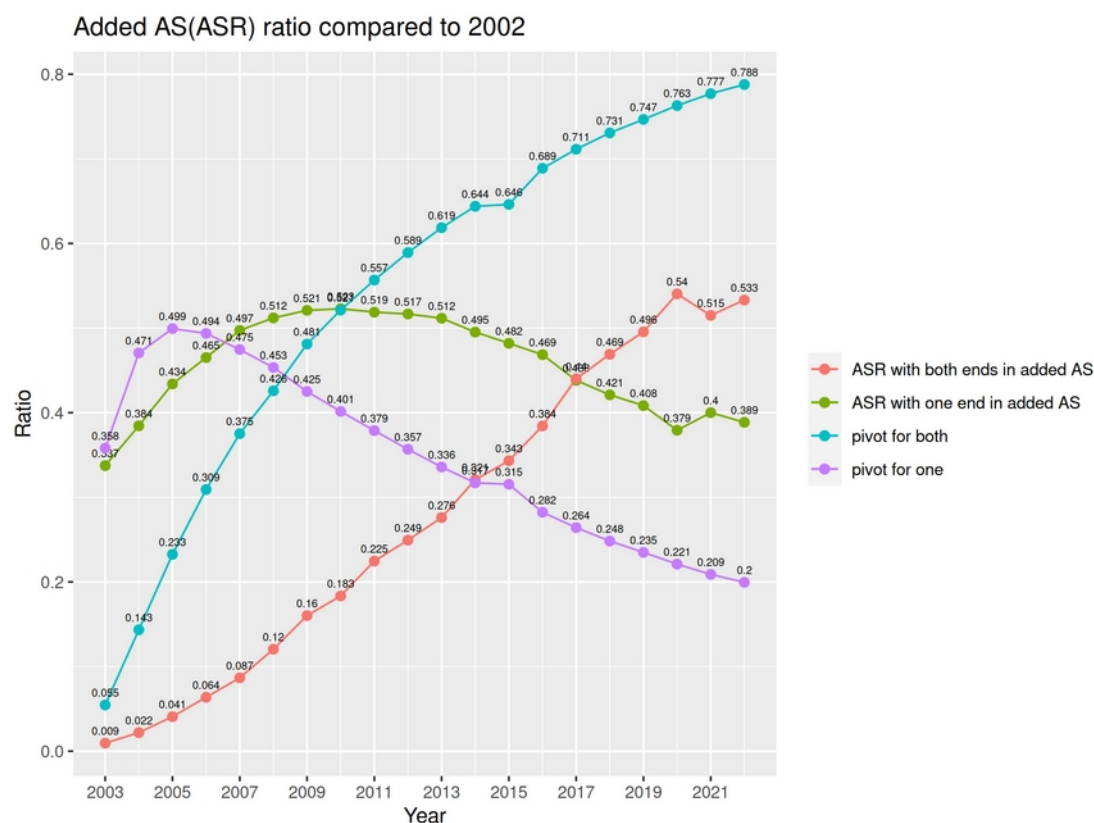


图 10 新增 ASR 位于新增 AS 中的比例

先看两端端点均在新增点中的情况，即红色的"ASR with both ends in added AS"，基准曲线为蓝色。基准值曲线在初始时值非常小(0.055)，但至 2021 年已到达 0.788，实际上反映了新增点占总节点比例的提升。在 2002 年时，节点总量仅有 32k,至 2021 年时已有 566k。因此到后期，2002 年的旧节点仅占 5.6%左右。而在 20 年间，实际的两端端点均在新加点中的新边比例曲线始终大幅低于基准值。由于基准值是随机情况下的概率，因此这一差异表明新增边并不“愿意”将两端点均置于新增点中。那么，即使在新增边占绝大多数的 2021 年，有很大一部分新增边仍然在与原先那只占 5.6%的老节点签订协议。

我们再看仅有一端节点位于新增点中的情况，即绿色的"ASR with one

end in added AS", 基准曲线为紫色。这两条线更有戏剧性：在 2007 年前，实际值低于基准值，而在 2007 年后，实际值高于基准值，且将差值渐渐拉大。实际值低于基准曲线，意味着在 2002 后的最初几年，多数新增边并不符合“一边在新增点内”的模型（或是说符合这一模型的比例偏小）。结合红色曲线，我们知道在这几年新增边也同样不符合“两边均在新增点内”的模型。那么在 2003 至 2006 这几年，多数协议仍然在原先的旧节点间建立。而在 2007 年后，实际曲线逐渐高于基准值，说明符合这一模型的新边比例逐渐升高，近乎是基准值的 2 倍。这表明在后面几年中，新增节点与原有节点建立协议的倾向性是很高的。这也帮助解释了前述两端点均于新增点中边比例偏低的现象。

此外，图中在 2021 年，两条实际数值线产生了波动。这是由于 2021 年新增节点比例 ($r_{\text{new_node}}$) 小于前面年份导致。图中红线的基准计算公式为 $r_{\text{new_node}}^2$ ，与 $r_{\text{new_node}}$ 正向关，因此 2021 年突然降低；而绿线的基准计算公式为 $2*(r_{\text{new_node}})*(1-r_{\text{new_node}})$ ，当 $r_{\text{new_node}}$ 等于 0.5 时取到最大值， $r_{\text{new_node}}$ 大于 0.5 后单调递减。图中绿线的峰值显然已过，且 2021 年时新增节点所占比例已大于 0.5，故 $r_{\text{new_node}}$ 的突然降低使得绿线在 2021 年短暂升高。

上述分析说明两件事情：一，老节点的影响力大而深远。即便经过了 20 年，2002 年的老节点只占 5% 的情况下，仍有许多新的协议倾向于和这部分节点签订。二，新加入节点从出现到发挥影响力，需要几年时间。从 2003 至 2006 这几年的情况来看，虽然新增节点所占比例不低，协议仍主要在原先节点间建立。事实上，参照 3.3.3 中的度数分布，新节点 50% 以上均为度小于 3 的边缘

节点。这类节点可能会与中心的强连通团（mesh）建立关系，也可能维持边缘节点(tree node)的状态多年。这样来看，老节点持续的影响力也可以解释。

接着我们再来看新减的情况。图 11 绘制的是新减边与新减节点的关联曲线。时间基准也为 2002 年。

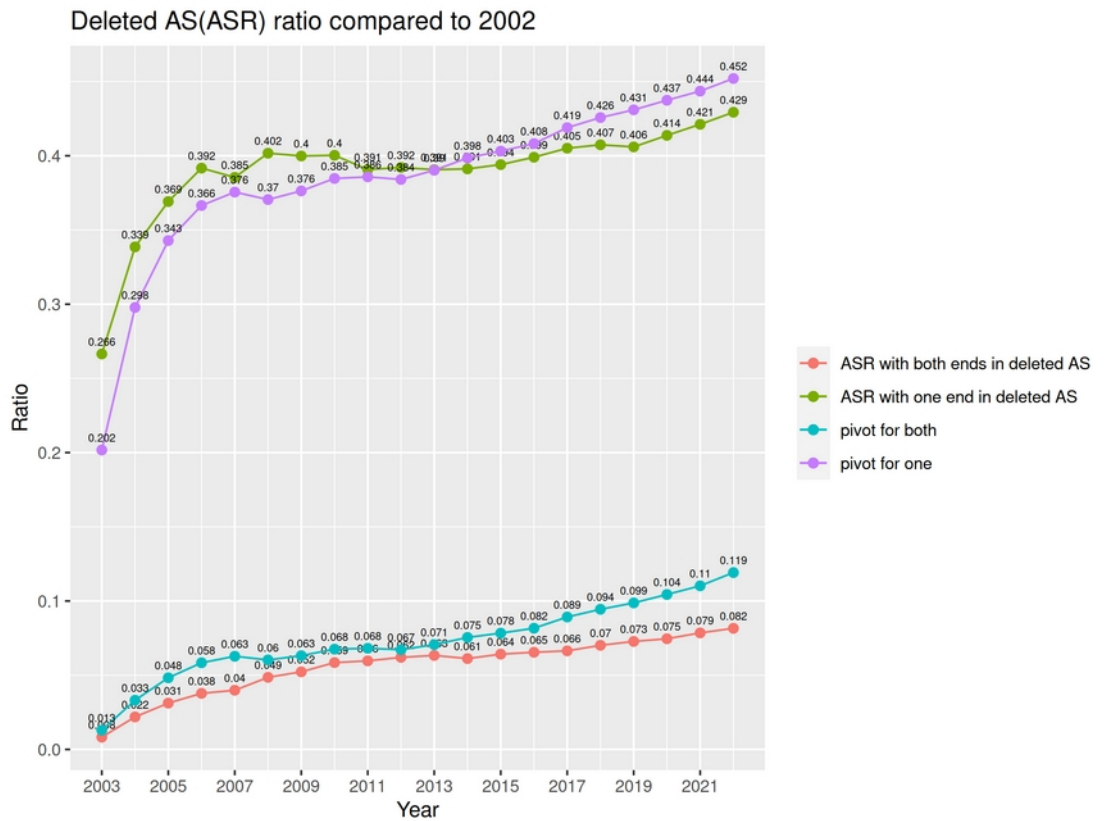


图 11 新减 ASR 位于新减 AS 中的比例

与新增图非常不同的是，新减节点的实际比例曲线与基准曲线相差不大（最大差值为 0.64），几乎相合。

虽然相差不大，但两端均处于新减节点的边比例，即红色的 "ASR with both ends in deleted AS"，始终低于蓝色的基准线，且在 2013 年后有差距拉大的趋势。这说明删减的协议，随着时间推移，越来越多的出自原有节点。仅一端处于新减节点的边比例，即绿色的 "ASR with one end in deleted

AS", 先略高于紫色的基准值（2003-2013），后低于基准值，且差距逐渐扩大。这说明在 2003 年后的开始几年，较多的删减协议为留存节点和新减节点间签订。而至后期，删减协议则倾向于来自保留下来的老节点间（2013 年后，红绿两者模式的实际比例均低于基准，因此占更多比重的边只能是剩下那种）。

通过上述分析可知，总得来说，新减边与新减节点的关联性并不强，但后期删除的协议更倾向于来自留存的节点。

3.4.5 增减节点（边）的连接性

在这一节中，我们探究新增减的自治系统的连接属性。也即它们在网络中的影响力。

对于节点（自治系统），我们计算了多项指标，包括 effective size, degree 等。得到的数据中，这些指标的代表性相差不大。为了简化表示，我们使用自治系统的协议数，也即边数（degree）来代表其连通性。

表 1 新增节点度数分布

	20020101	20120101	20220101
≤ 1	1525	2775	3422
(1, 2]	2035	1889	1278
(2, 5]	381	323	400
(5, 10]	75	77	94
(10, 100]	18(max: 75)	96	248
> 100	0	1(max: 246)	30(max: 3743)

表 1 为 2002 年，2012 年，2022 年同期（一月）新增节点的度数分布。在三个年份中，均有超过半数的新增节点度数小于等于 2,说明新增节点的主要来源是全新开发的自治系统此外，三年中新增的较大度数节点度数持续增大，

2022 年最大节点度数为 3743,在当年所有节点中的度数排名为 29。表明各年均稳定存在一些对于区间节点的维护。

表 2 新减节点度数分布

	20020101	20120101	20220101
≤ 1	642	1056	1890
(1, 2]	416	465	770
(2, 5]	163	105	268
(5, 10]	39	23	47
(10, 100]	25	27	95
> 100	0	1	13

表 2 为 2002 年，2012 年，2022 年同期（一月）新减节点的度数分布。相较于新增节点，新减的高度数节点比例相对更小。2022 年新减的最高度数节点为 1132。说明运营商在进行节点删除操作时，对高度数节点更加谨慎。对于边（自治系统协议），可用于判定其影响力的指标包括边上的介数中心性（betweenness centrality），可用于检查结构洞的效应值（effective size）等。我们本想结合结构洞理论，猜测新增边可能会与潜在的结构洞相联系（structural hole spanner）。

结构洞（structural hole）指的是网络中连接两个或多个不同社群的节点或关键人物。在社交网络中，结构洞是指一个人同时具备着两个或多个不同社交群体的联系，是社交网络中的重要节点。这些节点连接着不同社群，有着更广泛的信息流动和资源获取的机会。在组织或企业内部，拥有结构洞的人员通常可以获得更多的资源和信息，并且更容易在组织层级中升职和获得更多的影响力。因此，结构洞的概念在社会网络分析和组织行为学等领域有着重要的意义。在图中，结构洞可以通过网络分析和社区检测等方法进行识别和分析，有

助于了解网络中节点之间的联系和影响力，从而更好地理解网络的结构和功能。

理论上说，对于一个以获取流量为目的的网络，结构洞，根据其定义，正是可以帮助运营商赚取更多利益的枢纽位置。但经过计算新边的结构洞指标，我们并未发现新边成为 structural hole spanner 的倾向。可能的解释有很多。一是选取的指标不适合当前图。由于结构洞存在多种算法，比较精密的算法运算成本较高，因此本实验中转而选取了一种替代指标，即通过节点上的效应值和约束(constraint)来计算结构洞。但根据上述替代指标计算的结构洞，经过比对与直接根据度数大小排名得到的节点相似度极高。这说明替代算法并不十分理想，或者当前的自治系统网络连通性已非常高，因而结构洞的存在几乎等同于高度数节点。第二种解释是，运营商新建协议的重心不在于占取枢纽我们从表 1 已经看到，绝大多数新增点度数极小，位于网络边缘，属于树形结构，仅服务小区域用户。对于其管理的其他节点，碍于运营的成本，商业谈判约束，行情和周遭网络结构频繁变化等等，都决定了运营商会选择尽量优化已有节点，即使知晓一个可能结构洞的存在，也未必会冒险改变现有框架去实现它。

虽然结构洞理论的探索并不成功，但结构洞算法所代表的，用网络结构中关键位置去定义边、点影响力的思路，仍是可取的。因此我们换了一种指标，使用自治系统协议（边）两端所属社群（community）的分布情况去衡量边的影响力。

社群检测(community detection)是指在复杂网络中识别具有内在紧密联系的节点群体的过程。其算法方法包括模块度优化、谱聚类、标签传播等。通过对网络进行社区检测，可以帮助我们了解网络结构的特征和模式，揭示网络

中节点之间的相互作用关系，发现潜在的子群体和社区结构，并为预测节点行为和网络演化提供基础。在本文中，子群体由自治系统构成。自治系统子社群的划分，标志着自治系统网络的聚合结构。这些自治系统子社群通常代表着在自治系统网络中，拥有类似连接模式、服务类型和流量传输模式等特征的自治系统的集合。

我们使用 NetworkX 中提供的 community detection louvain 方法，对 2001 至 2023 年 1 月的自治系统数据均进行了社群检测。对于每份数据，运行该算法后得到一组自治系统编号集，每个集合即为一个社群。社群的大小分布不均，断层较大。以 2020 年的社群结果为例，共有 137 个社群，最小的社群含 2 个自治系统，最大的含 18011 个自治系统。137 个社群中比例最高的社群为大小为 2 的社群，有 52 个。由于除大小小于 20 的小规模社群外，其余社群的大小都是独特的(unique)，因此我们通过社群大小给所有每个时间点网络中的社群排名。

图 12、13 分别为 2003，2023 年新增（减）边两端节点所属社群排名的 CDF 曲线。"Added"开头曲线为新增，"Deleted"开头曲线为新减，"Both"表示两端点均符合，"One"表示仅有一端点符合。如"Added Both"曲线上点(x, y)代表“共有比例为 y 的新增边两端节点所属社群的排名均不大于 x”。另外，在同一个社群内的边（即两端点属于同一社群）所占比例也在图中标出。自 2000-2023，这一比例在区间 [0.65,0.8] 间浮动，没有明显的时间规律。

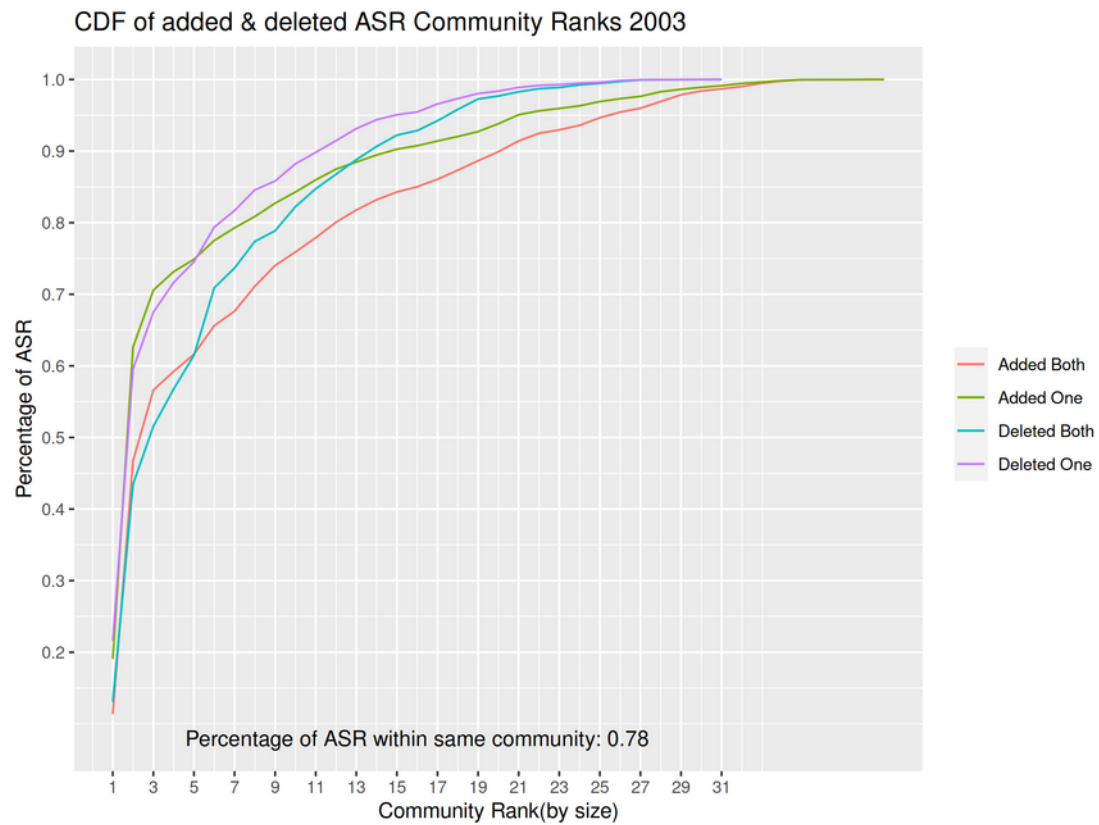


图 12 2003 新增减 ASR 社群排名 CDF

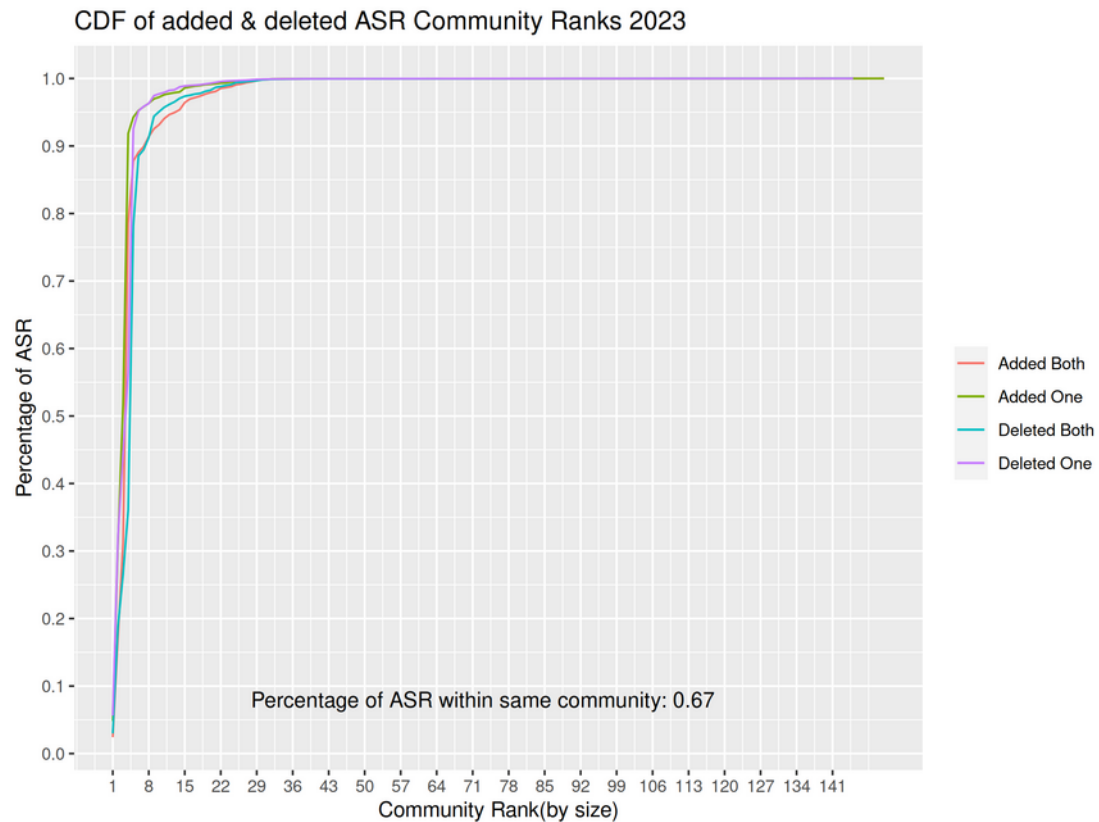


图 13 2023 新增减 ASR 社群排名 CDF

由于两端点的约束条件比一端点严格，因此图中"Both"标注的橙、蓝两条线分别位于标注"One"的曲线下方。从图 12 可以看到，有至少 50% 的新增（减）边来自排名前三的社群，以及有至少 90% 的新增（减）边来自排名前二十的社群。图 13 中，向排名靠前社群的聚集效应更加明显：90% 的新增（减）边属于排名前八的社群。

在两图中，四条曲线的相对位置是近乎一致的。在社群排名 5 往前，新增边对应的曲线收敛速度略快于新减边。而在社群排名 5 之前，新减边的曲线持续高于新增边。这一现象表明新增边相比新减边略更集中于排名前 4 的大社群，而在排名 5 往后，新减边相对更集中于排名靠前的大社群。

3.5 自治系统层级的实例分析

在前面章节中介绍了自治系统边与节点变化的宏观规律。在这一章中，我们仍然聚焦自治系统的波动，但将从更底层的案例出发，选择部分自治系统，观察其规律。由于这一章为案例分析，将从绘制的图表、数据表中提取规律，但结论主要来自主观推理，因此受主观因素影响。

3.5.1 各层级自治系统的度数变化

由于运营商的政策更迭，自治系统往往会经历拆分、合并、重组、重定协议等操作。在这期间，经历维护的自治系统将表现出波动。其中最明显的波动特征即为协议数变化，即该自治系统（节点）的度数变化。在观察具体自治系

统度数逐年变化规律的时候，我们发现当某些自治系统度数经过多年积累到达一个足够大的值时，这个自治系统的度数会在很短的时间内急剧下跌，甚至消失。这样的自治系统量级非常大，突然消失而不产生任何后果看起来不符合业务逻辑。我们的猜想是，这个消失的自治系统由于种种原因被运营商拆分成了多个替补自治系统，这些替补继承了原先大自治系统的大部分业务协议。

为了证实这个猜测，我们进行了如下探索：对于一个突然消失的大自治系统 AS1，在消失同年，计算其邻居（直接相邻节点）的相邻节点与 AS1 原先邻居的交集大小及变化，并对这些邻居的数据和背景信息进行观察。我们确实找到了一些类似于拆分替补的案例。如编号 25462 这个节点。它在 2008 年后消失，消失时仍有 646 个邻居（也即度数）。而在 2009 年，有同一个运营商"RETN-UA-AS"旗下的另一个自治系统，编号 9002 突然出现（20080101 数据中不存在），且在 20090101 的数据中，9002 节点与 570 个 25462 原先的邻居相连，占原先协议数的 88%。但这类明显的例子并不多，大多数情况下我们确实能够找到新旧节点共享邻居数超过 60% 的新节点，但我们没有其他诸如“同属一个运营商”这样强有力的证据说明这两个节点的重合度不是巧合。因此，想直接证明这一点是比较困难的。

由于直接分析单个节点的拆分和合并规律未能成功，我们退而探索是否能从更长的时间跨度和更多的节点数上归纳出规律。

图 14、15、16、17 分别为 20140101 数据中按度数排名前 10 自治系统，20100101 数据中排名 200 左右的 10 个自治系统，20100101 数据中排名 1000 左右的 10 个自治系统，及 20200101 数据中排名 5000 左右的 10 个自治系统在 2000 至 2023 年区间内的逐年度数情况。此处节点已进行筛选，保

证节点在中途并未出现更换更名。选择不同时段自治系统观测，一是为了随机性，二是也可以观察时段特征。

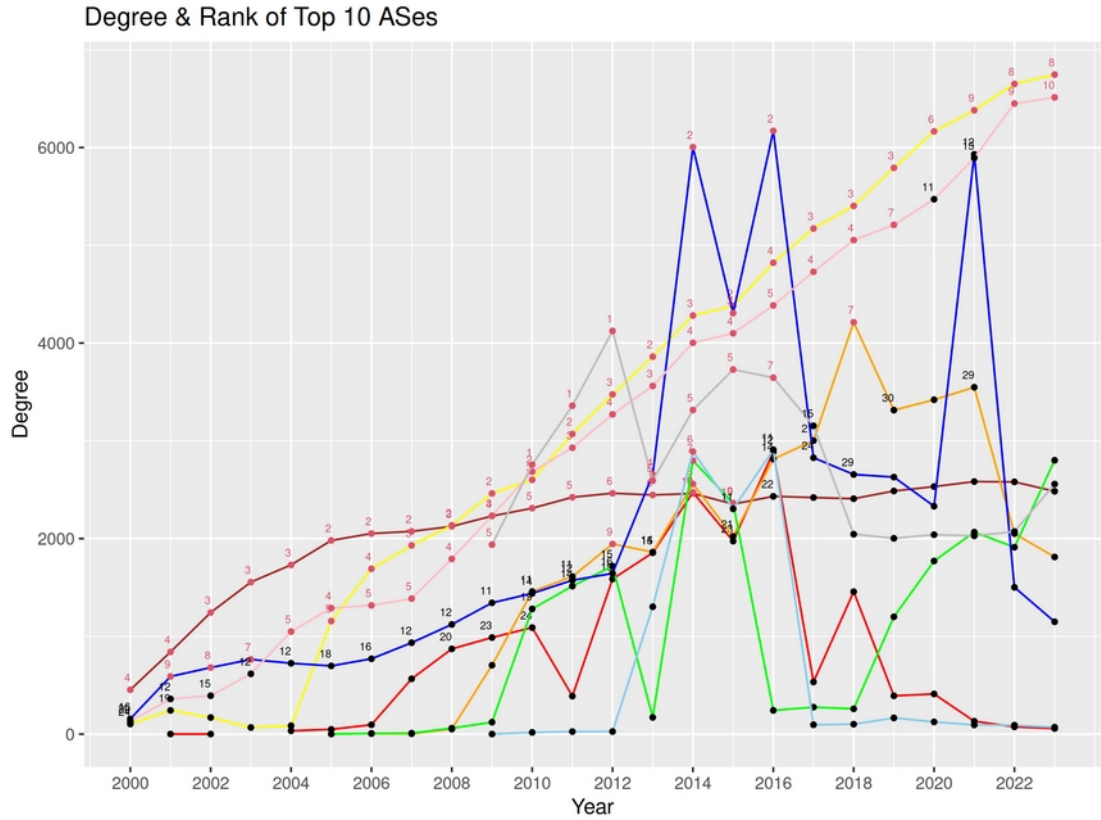


图 14 2014 度数前 10 名 AS 的度数曲线 2000-2023

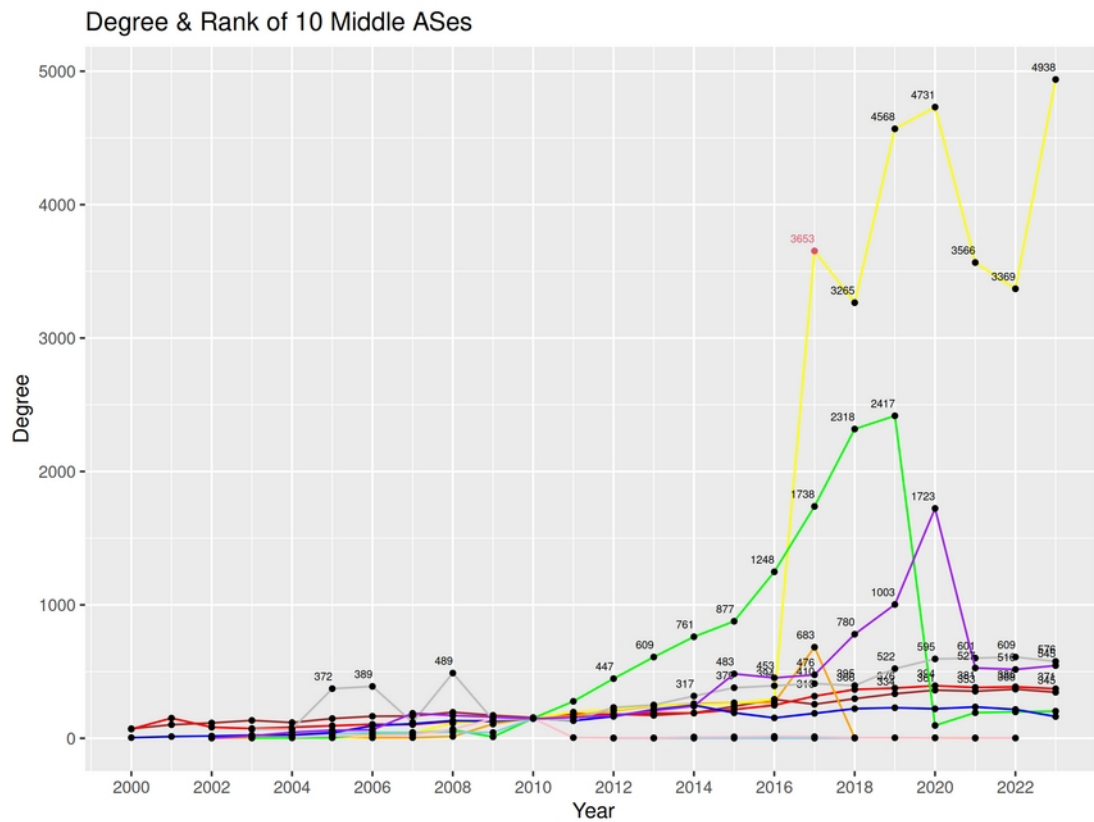


图 15 2010 度数中间 10 名 AS 的度数曲线 2000-2023

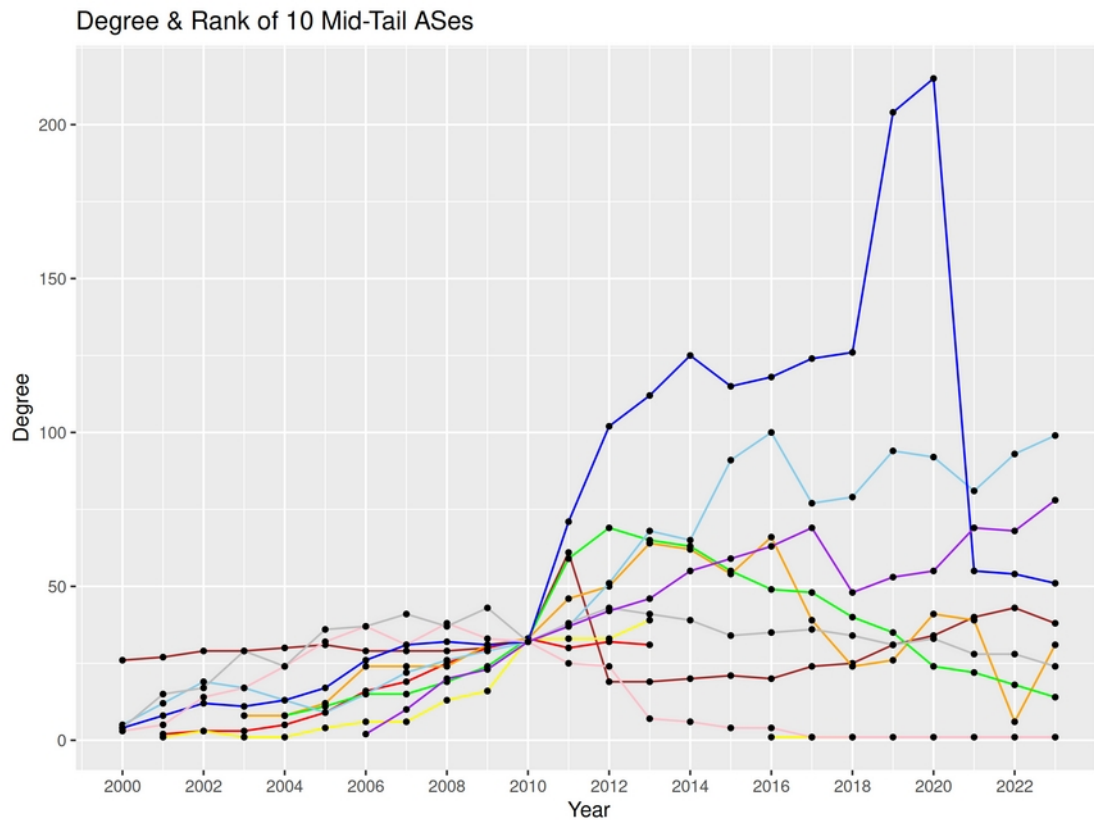


图 16 2010 度数中后 10 名 AS 的度数曲线 2000-2023

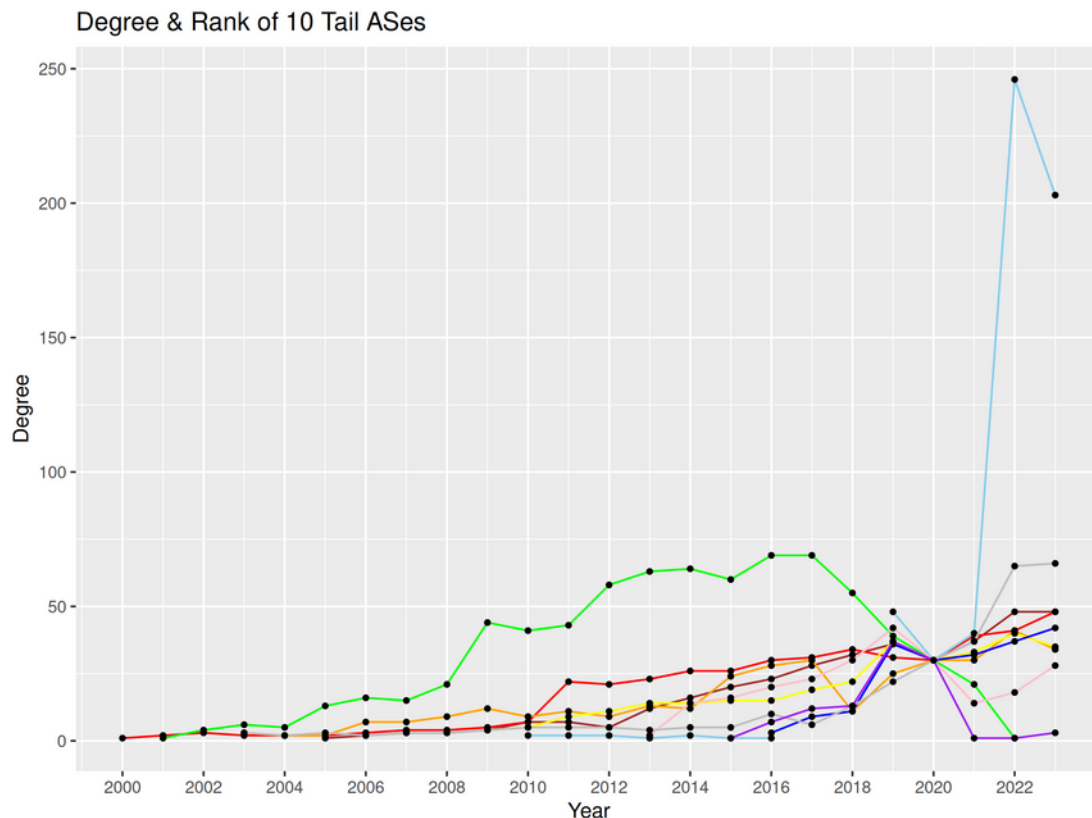


图 17 2020 度数末 10 名 AS 的度数曲线 2000-2023

上述图中，排名进入前十的节点颜色标红。后三张图（图 15,16,17）中节点度数大于 300 会将度数标出。

图 14 中，2010 年度数前十的自治系统在两年后仅剩 6 个维持在前十，而在 2023 年仅剩 2 个。2010-2012 年是图中 3 条曲线仅有的进入前十的时段。对于其余大部分曲线，也在某些年段表现出了“上涨-下跌”的模式。在这 23 年区间内，共有 60 个自治系统进入过排名前十。但观测其余年份前十节点的度数趋势图会发现，除了极少几条曲线保持涨势到最后，其余自治系统仅在自己进入前十的 2-3 年内排名高，随后度数下跌，并不再出现前面的涨势。对于保持住涨势的节点，亦或其出现较晚，亦或其的增长幅度从未过高，但是几乎稳定（如图 14 中黄色、粉红两条线）。在这里可以得到一个模糊的结论，即

节点度数上涨到一定值后，一旦开始回落，后续就几乎不会再出现同样的涨势。

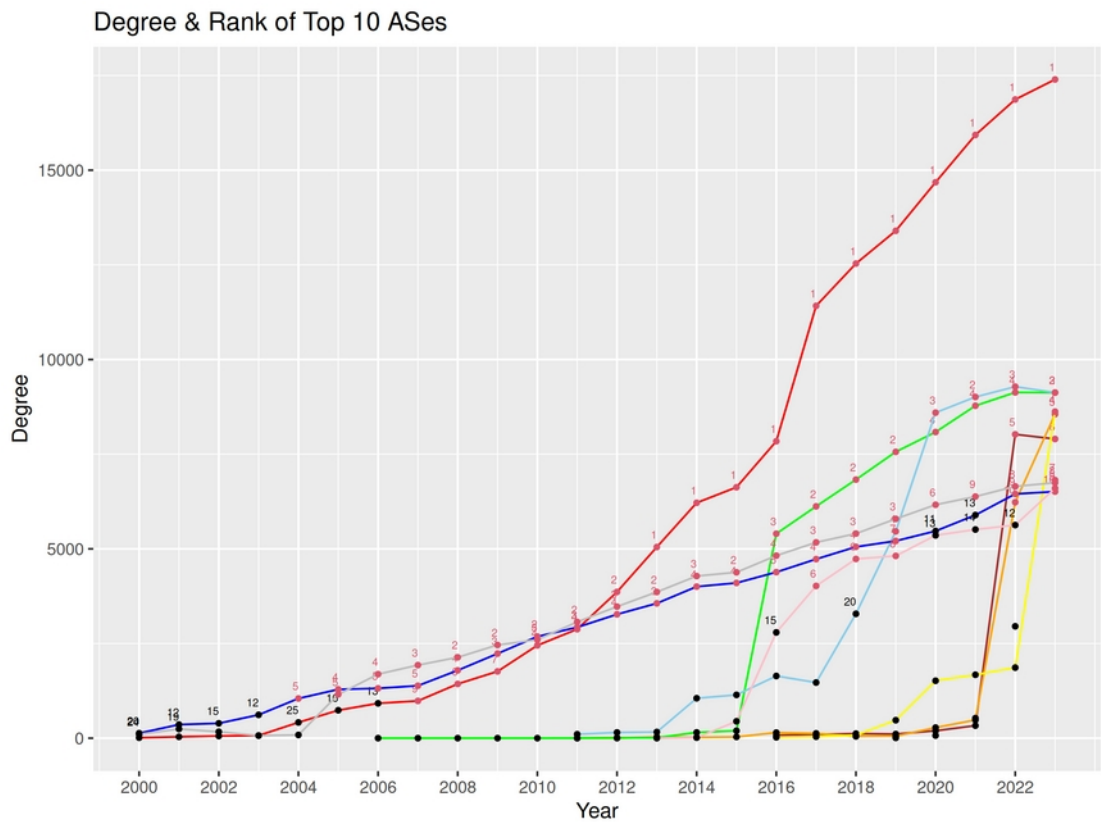


图 18 2023 度数前 10 名 AS 的度数曲线 2000-2023

图 18 是 2023 年 1 月排名前十节点的度数趋势图。注意，此处显示出了所有自治系统编号，并未筛选曾在某年间不存在的节点。可见这些自治系统在 2000-2023 区间内均为上升趋势。可以预测，在后续 1-2 年里这其中将有多 个自治系统度数骤减。

其余图表中的自治系统随机选自对应数据文件处于该度数排名区间的自治系统中。由于自治系统度数分布的长尾效应，这部分节点的体量很大，因而随机性更强。图 15、16、17 中均可看出“X”型分布：在聚合的时间点前后，各条曲线的走势均发散，表明曲线度数随年波动。而在聚合时间后的发散程度要高于聚合时间前（体现为聚合时间后曲线间的方差更大），说明随着年限增大

网络的活跃度更高，熵增加。也可能是由于随着这些采样自治系统前些年份的资源积累，后期的影响力、活力更大。后一种解释的支撑在于，在这三张图中图 15（其中节点排名处于中位，1000 名左右）的自治系统曲线在聚合时间点后的活跃度要显著大于其他图。这张图中的自治系统相比于其他图低度数的自治系统拥有更多资源（相邻节点，积累的经验等），因此后期的活跃度可能也与此有关。

3.5.2 单个自治系统的自我中心网络

为了更精细地抓取自治系统与其周边节点的关系，我们选取了一些节点，观测其各时间段的一跳自我中心网络(ego network)。自我中心网络是指以某个节点为中心，包含该节点和与其直接相连的节点所构成的子图。通过分析自我中心网络，可以揭示节点在局部网络中的重要性、影响力和作用，发现潜在的社群结构和节点聚集性，预测节点的行为和演化等。

在“3.3.2 连通分量”一章中，我们曾提到 LVL1-1 机构下属的编号为 1 的自治系统。图 19 是自治系统 1 在 2000,2004,2005,2020 年的自我中心网络可视化图。图中红色节点为中心节点(ego node)，其余为与其有协议关系的邻居节点（此处协议为提供商-顾客型协议或同行型，即只标出中心节点的顾客节点和同行节点，单向边指向为顾客节点，双向边为同行型，后面所有自我中心网络规则与此相同）。节点内部第一行为节点的自治系统编号，第二行为其总度数（包括所有协议），节点大小与其度数正向关。

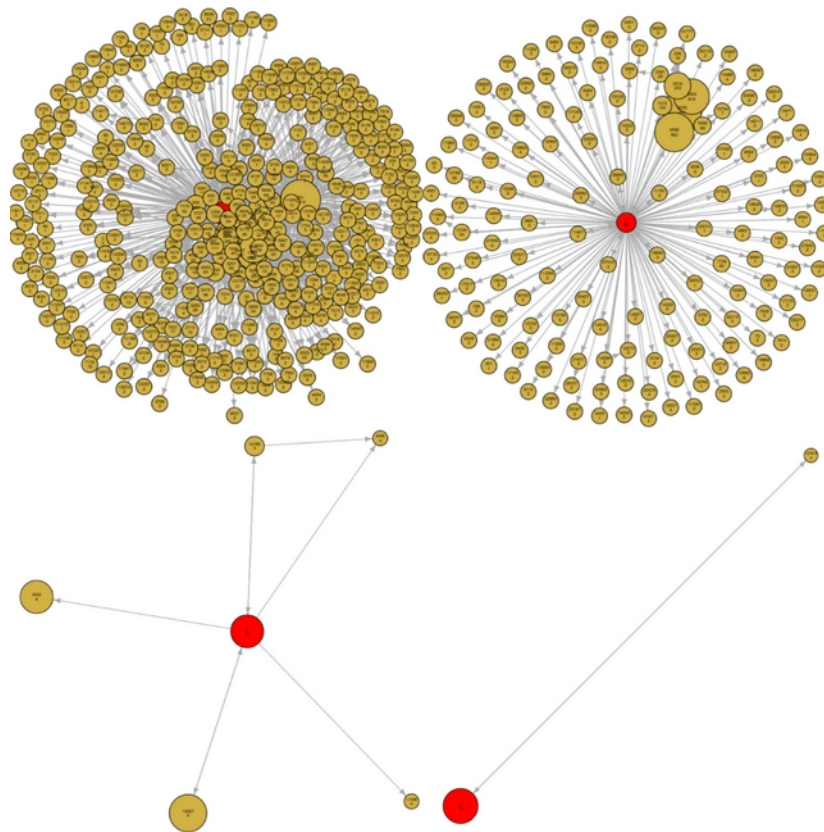


图 19 AS 1 ego network (2000,2004,2005,2020)

前文中曾提到，自治系统 1 曾经是骨干网中的一员（2000）。2004 年开始，其边数大量流失，在 2020 等年份与主连通集团断裂。图 19 截取了这几个时间点。在自治系统 1 鼎盛时期，其结构其实相对是比较简单的发散性结构，其自我中心网络大部分都是直接相连的顾客节点，成团的情况很少。

图 20 是自治系统 6295 在 2001, 2006, 2013, 2018, 2020, 2021, 2022, 2023 年的自我中心网络可视化图。自治系统 6295 是运营商 LUNAVI-WA（美国）旗下的用于连接其数据中心和客户网络的自治系统。通过这个自治系统，我们可以清楚地看到一个新生于 21 世纪初自治系统的发展历程。

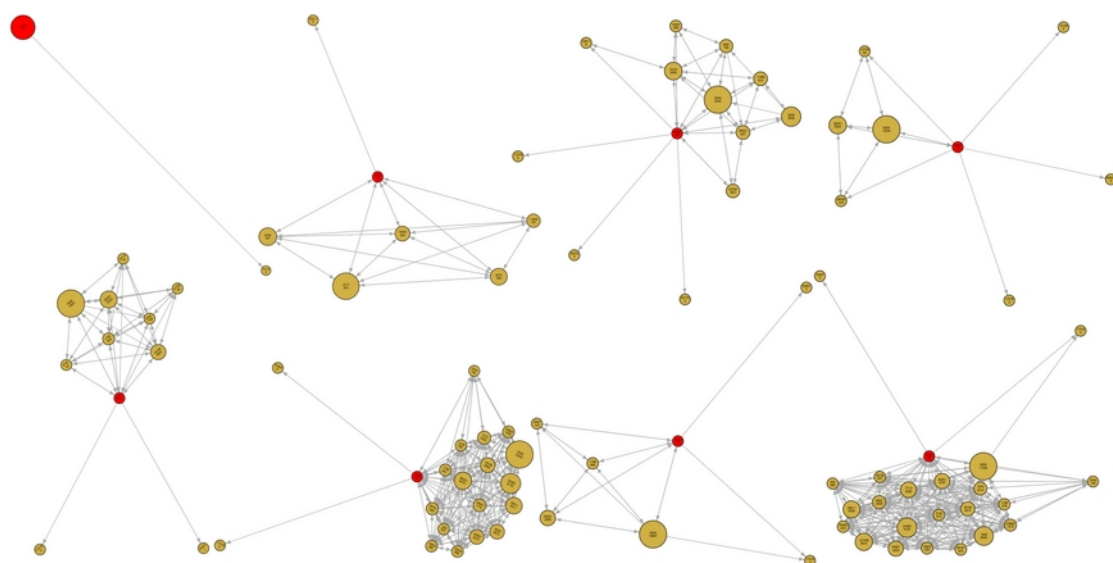


图 20 AS 6295 ego network (2001,2006,2013,2018,2020,2021,2022,2023)

6295 初期只获取服务（大多数新自治系统初期均是如此）。2001 年，它获得了第一个用户，自治系统 12175，属于美国的 YELMNET 运营商。2006 年时，它被第一次接纳为一个团（cluster）的成员，这个团内部节点互相均达成了同行合作协议（p2p）。2013 年时它解除了先前的合作协议，成为了另一个影响力非常大的自治系统 6939 所属团的一部分。（6939 是由全球互联网服务提供商 Hurricane Electric 所拥有和运营的。Hurricane Electric 总部位于美国，是世界上最大的 IPv6 网络之一，并为客户提供 IP 传输服务、托管服务和网页主机服务。Hurricane Electric 的自治系统 6939 与全球各地的其他 ASN 广泛连接，为客户提供全球互联网连接。该公司还维护着庞大的同行网络（peering network），使其能够直接与其他网络交换流量，从而提高网络性能并降低延迟）。此后自治系统 6295 一直保持着和 6939 的合作关系。在 2018 年和 2022 年，6295 的协议量分别有了大幅减退，但随后又很快恢复。

相比于 1980 年成立的自治系统 1 的自我中心网络，6295 发展后期的自我

中心网络集聚系数大得多 (clustering coefficient)。事实上, 加入和建立同行网络是大部分自治系统的策略, 也是它们最好的策略。同行关系让协议网络中的节点互相提供服务而无需另外购买服务, 节省了资金。成立较早的自治系统 1 所处的自我中心网络几乎没有团结构, 这也许也促成了其二十世纪初的瓦解。

3.6 以国家为单位的聚合分析

在前人的研究中, 也曾根据自治系统运营商所属国家将数据进行聚合 [2]。由于本文着力于自治系统的贸易性质, 国家因素的影响自然相当重要。

我们根据自治系统编号匹配数据将属于同一国家 (或地区) 的自治系统聚合, 新的节点成为国家 (或地区) 代号, 对应的自治系统协议数则转化为边上的权重。

3.6.1 国内外协议总量

我们希望对各国协议总量发展情况做大致的认识, 同时比比内外贸易的比重。下列图绘制了若干国家 (或地区) 2021 至 2023 1 月的总内外协议数。标题中两位大写字母即为该国家 (或地区) 缩写, 如 "CN" 代表中国, "US" 代表美国。图中 "domestic count" 即国内签订的协议 (两端点均属于国内运营商), "foreign count" 指和别国运营商签订, 而 "top foreign country count" 另外汇出了与该国与其贸易关系最多国家的协议曲线, 以作为对比。

下面图 21-26 分别为加拿大, 中国, 英国, 巴西, 韩国和美国的逐年内外

协议数图。

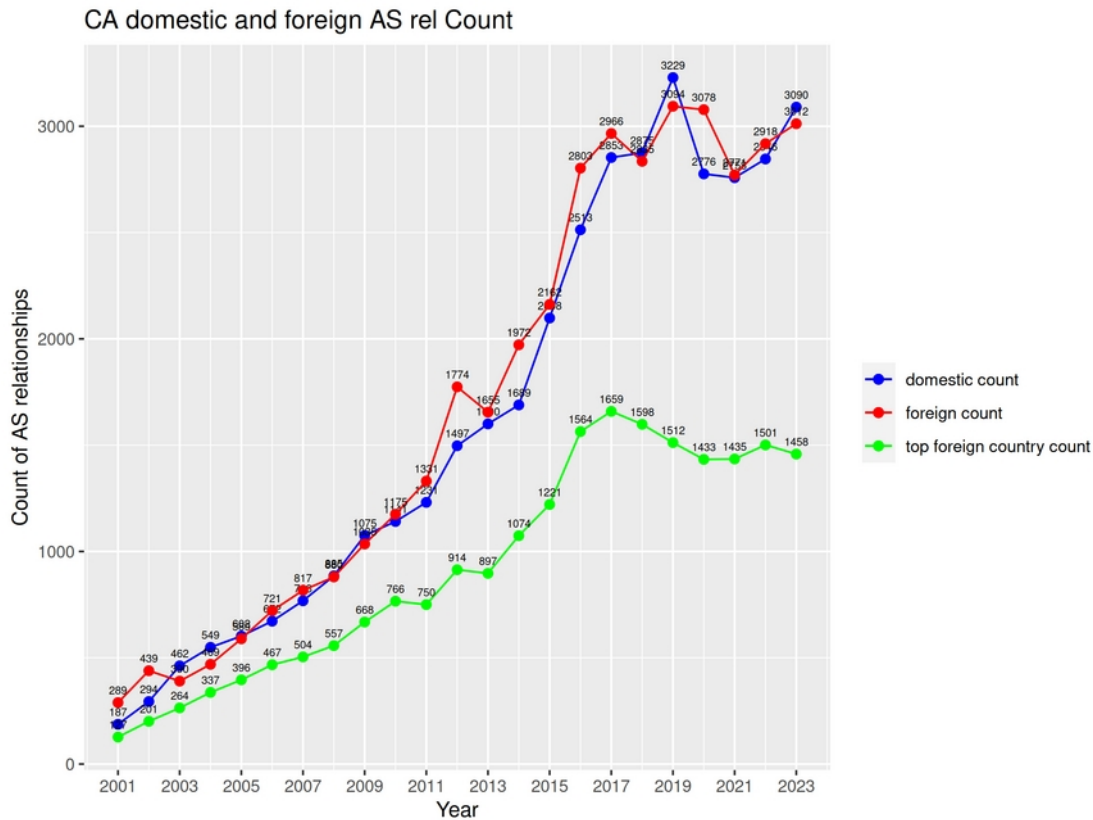


图 21 加拿大国内外 ASR 总量 2001-2023

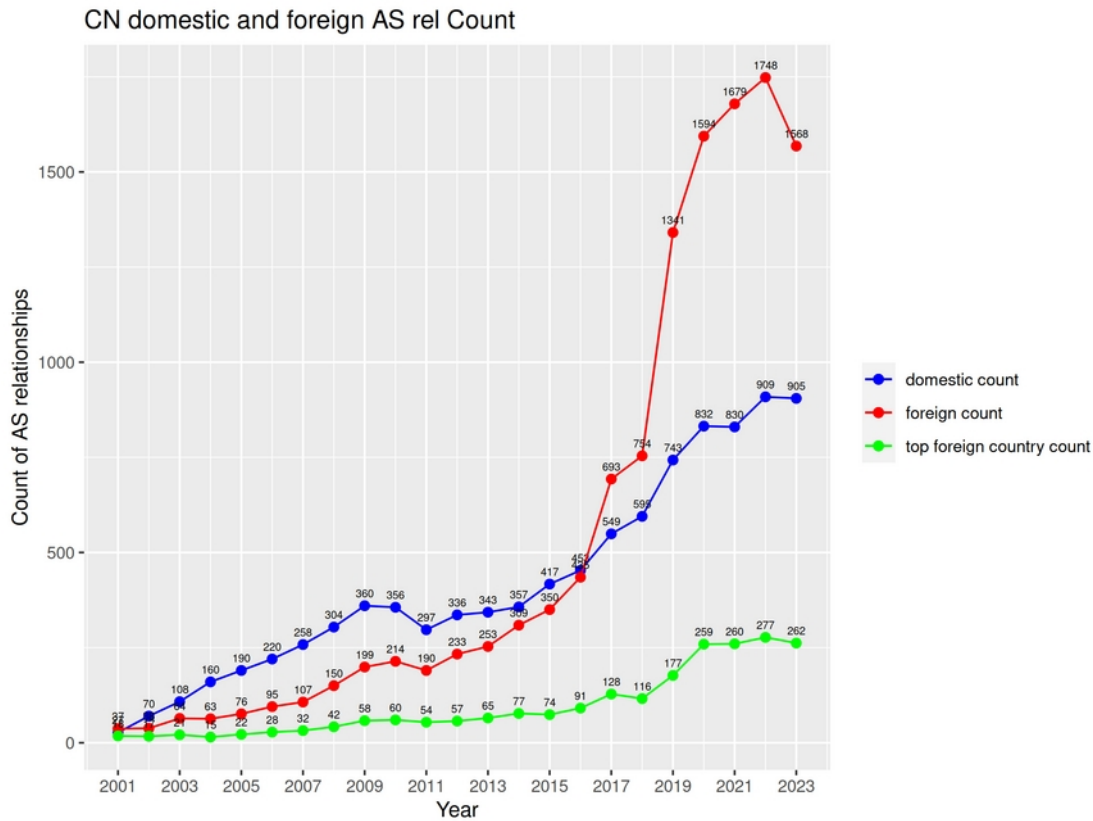


图 22 中国国内外 ASR 总量 2001-2023

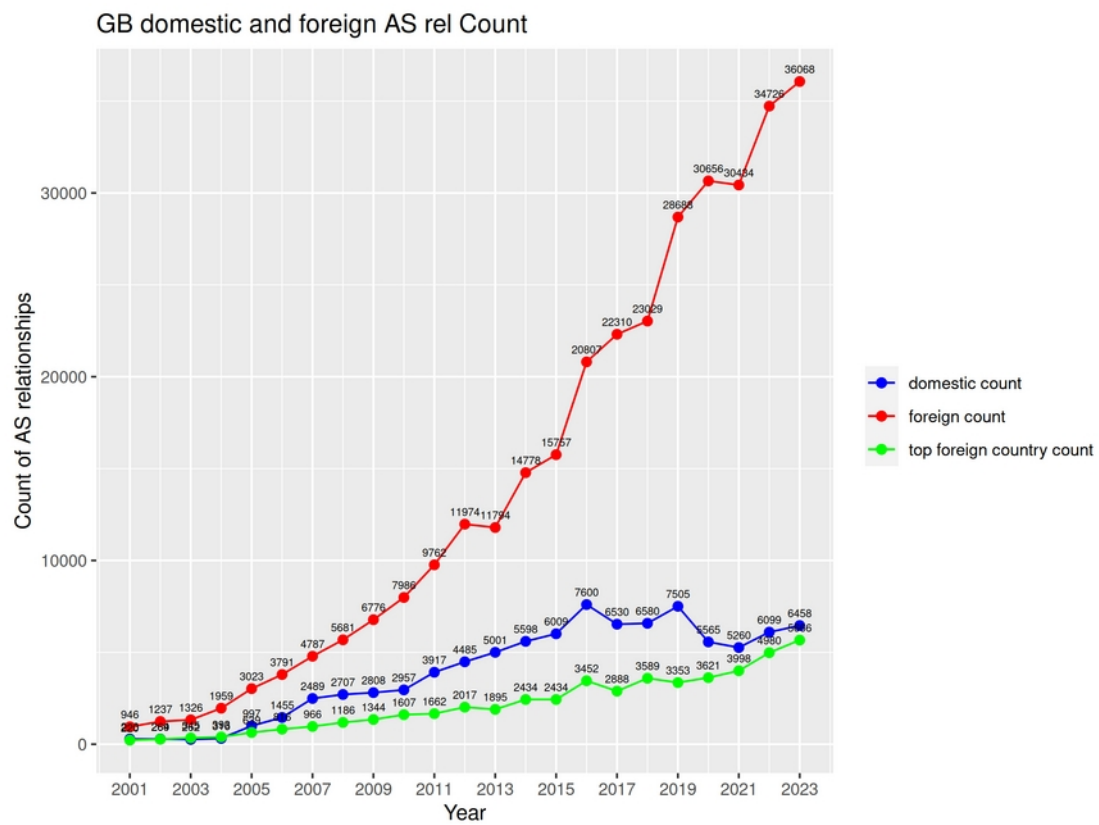


图 23 英国国内外 ASR 总量 2001-2023

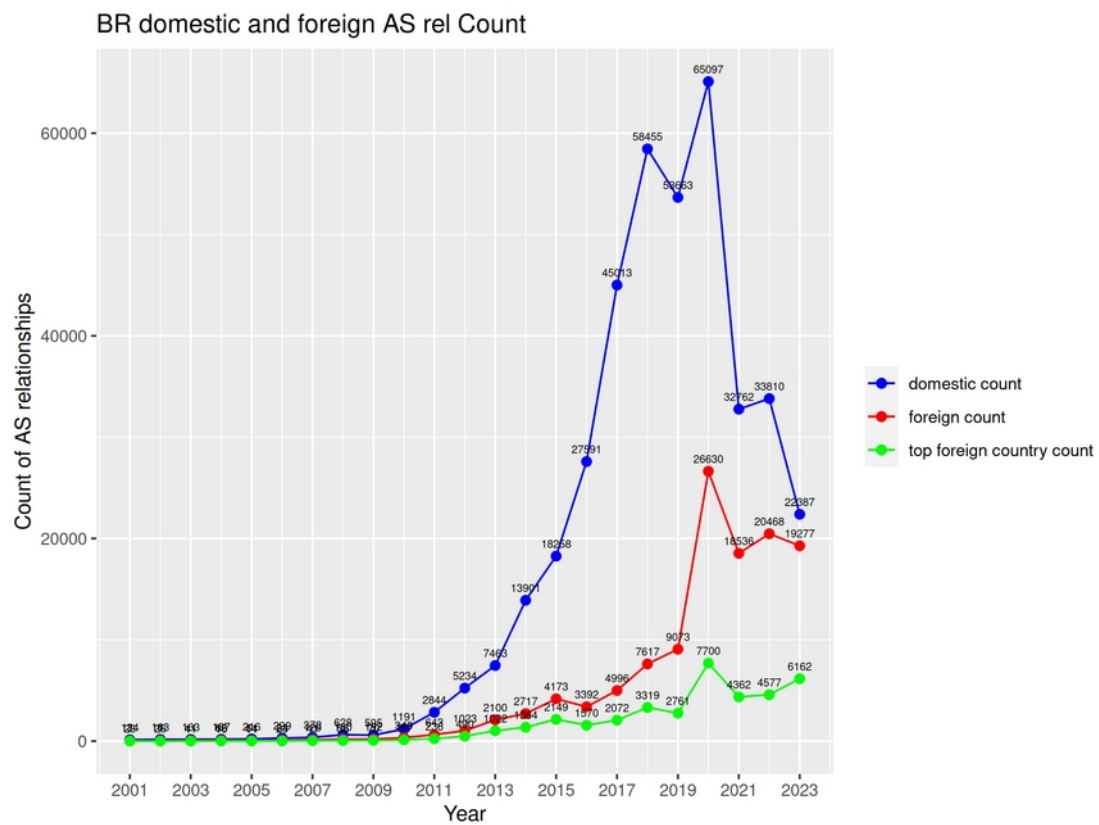


图 24 巴西国内外 ASR 总量 2001-2023

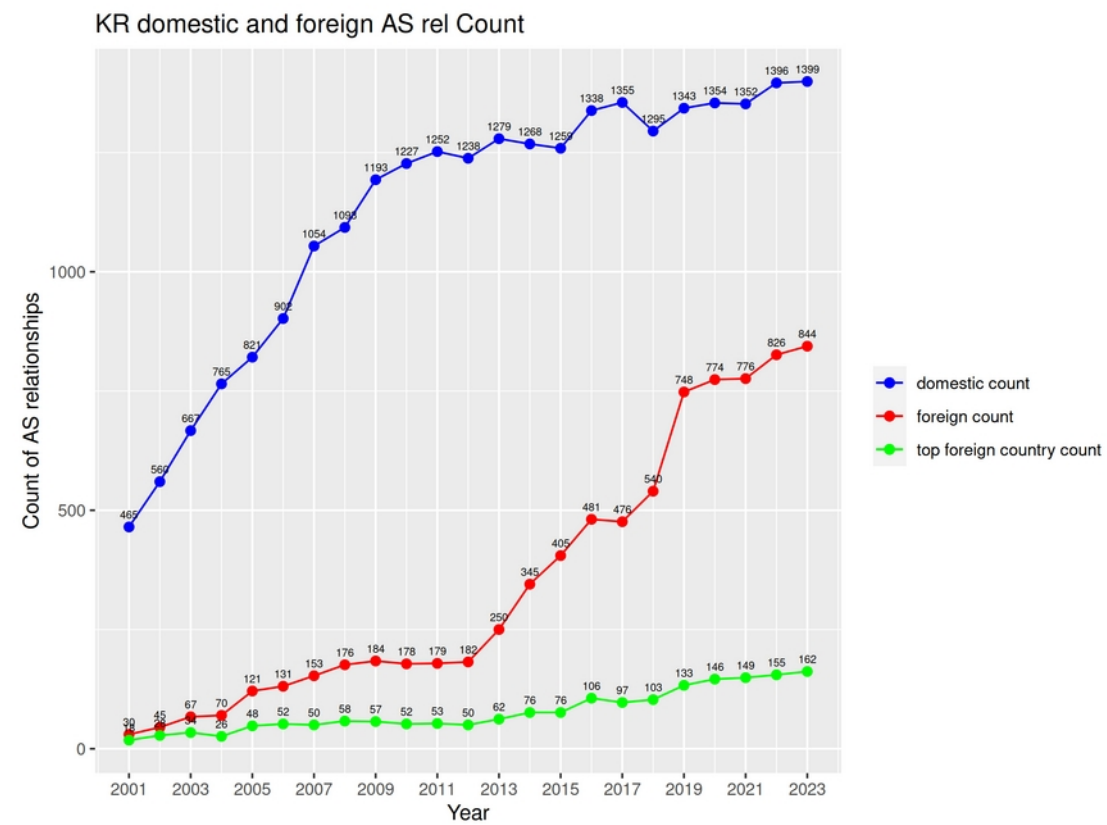


图 25 韩国国内外 ASR 总量 2001-2023

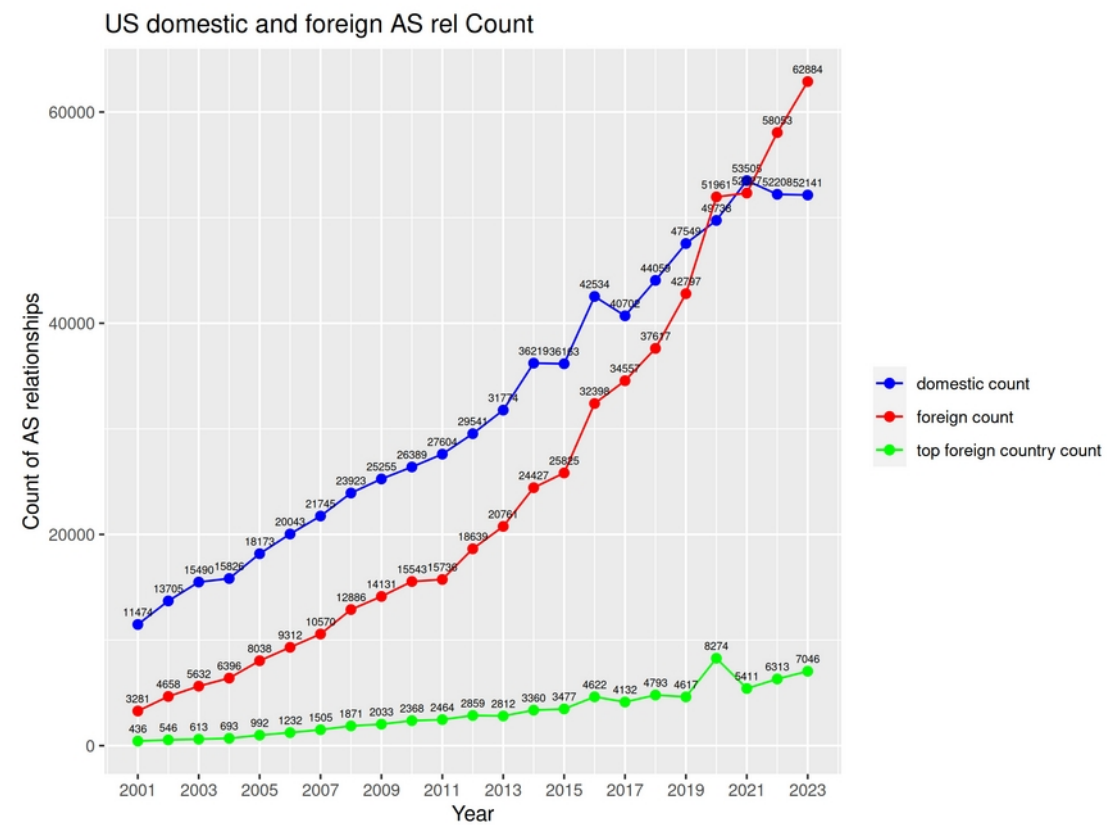


图 26 美国国内外 ASR 总量 2001-2023

加拿大（图 21）的协议量曲线是非常“标准”的曲线。此处标准，指符合指数增长规律。在“3.3.1 总量”一章中，我们已经看到全世界自治系统协议总量的增长曲线符合指数函数。而加拿大的曲线在 2020 前，也较高程度地符合指数增长。此外，加拿大的国内外贸易量几乎持平。

中国（图 22）的曲线就不符合指数规律，国外贸易总量在 2016 年方赶超国内总量，许多在前期借助别国大自治系统的国家都有这个标志（如日本，菲律宾，泰国）。

英国（图 23）作为老牌互联网贸易出口国家，是世界上自治系统总量最多的国家之一，其自治系统构成的团在欧洲密度很大。而同属于互联网贸易大国（仅从数量而言）的巴西曲线（图 24）跟英国相比跌宕许多，受疫情波动更大，且国内贸易量占优。

韩国（图 25）作为世界上网络设施最好的国家，其自治系统量级并不算大，国内贸易占主，且出口贸易相对较少。

美国（图 26）是世界上自治系统占有量最大的国家，曲线走势相对平稳。上述国家只是一些典型代表。其余国家各自的图示也各有不同。聚合过的数据相对来说噪音更小，更能反映出所属集团的趋势走向。比对国家聚合数据和原数据体现出的规律差异，也是相当有意思的课题。碍于本实验时间限制，这些工作留待将来探索。

3.6.2 国家社群

对于国家聚合数据，我们同样运算了社群检测(community detection louvain)，并将每个国家（或地区）的数据点绘制在其首都的经纬度位置，形成了2000至2023年1月国别社群的可视化世界图。同个社群的国家（或地区）颜色相同。图中坐标系为经纬度形成的矩形世界坐标，此外为了易于观测将没有点的区域删去，留下经度 $[-100, 180]$ 和纬度 $[-50, 70]$ 。

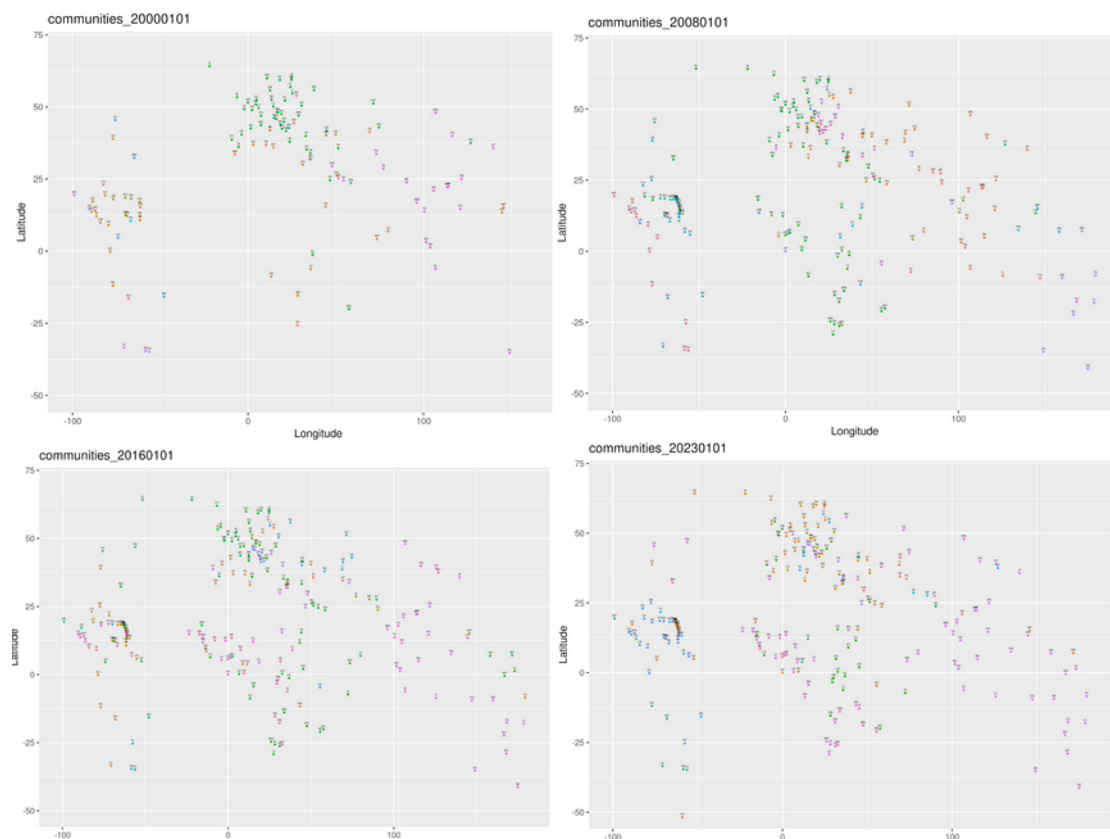


图 27 2000, 2008, 2016, 2023 国家社群

图 27 为 2000, 2008, 2016, 2023 年 1 月份国家为单位的社群可视化世界图。由于社群的演化，我们并未设定标准判断跨年间的两社群是否由同一社群演变而来，因此同一地区的社群在其他年份颜色可能发生变化。

图中可以清晰看到互联网在世界上的普及过程。2000 年时，互联网普及最完善的地区在欧洲（绿色社群），亚洲也拥有相对统一的社群。2008 年，南美加勒比海沿岸国家和巴西形成了统一社群；非洲大陆、中南亚国家的社群也开始出现。2016 年，非洲中部社群基本形成，亚洲南北连接澳洲形成跨度较长社群。2023 年与 2016 年变化不大，边缘新出现少量国家，说明世界网络社群形态已基本稳定。

3.7 新冠疫情的影响分析及启示

在前面各章分析中，我们多处遇到了 2019 年底开始的新冠疫情对自治系统网络发展造成的影响。本章中，将从三个维度，从大到小地分析疫情波动对网络总量、国家、运营商造成的影响，以及得到的启示。

3.7.1 总量

在 3.3.1 的图 1 中，我们已经看到虽然自治系统总量的增长趋势相对平稳，而自治系统协议总量则在 2021 与 2023 年产生了较大的颠簸。由于图中数据为 1 月数据，因此实际反映的波动应在 2020 与 2022 年。这两年分别是新冠疫情对世界经济贸易影响较大的两年（2021 相对稳定）。

如果不受疫情影响，根据我们计算出的指数增长曲线（图 1 中标"ASR Prediction"），图中 2021 至 2023 年自治系统协议总量的预测值分别是 678318,789833,919572。而 2023 年实际的数据是 611615, 相差 307957，这个数值与 2016 年时自治系统协议总量相似。

从 2023 年后的数据，我们很难推测疫情后的自治系统协议长期涨势。若按照 2021 年的涨势估计，自治系统协议总量要恢复到 2023 年的预估水平需要 5-6 年。但是 2021 年仍处于疫情中，世界经济并不算完全恢复，因此乐观估计，今后的长期发展速度应是高于 2021 年水平的。实际情况可能并没有上述那样悲观。

3.7.2 国家

从 3.6.1 中我们可以发现，不同国家受疫情影响也不尽相同。有些国家保持了增长，只是增长速度放缓（加拿大，法国），有些几乎不受影响（欧盟，奥地利，日本，韩国，泰国，澳大利亚，新加坡），有些国内贸易减缓，但国外贸易反而增加（以色列，法国，阿拉伯联合酋长国，新西兰，菲律宾，美国），有些是国内贸易增加，国外放缓（阿根廷，印度，意大利，葡萄牙，土耳其），有些则国内外贸易均受较大影响（中国，巴西，瑞士，马来西亚，俄罗斯）。虽然国家这一政治性概念对于网络来说不全然严谨，但以国家为单位展现出的这些规律，还是从宏观上说明了政治、经济与网络间的关系。

以韩国、巴西为例，两国政府分别在 1990 年初期和末期发动了互联网普及工程。时至今日，韩国已成为世界公认的网速、覆盖率第一的国家。而巴西拥有世界前列的自治系统储量，其自治系统出口贸易也联合了许多周边国家（见图 27 中巴西的社群）。然而巴西也公认面临许多问题，包括覆盖率和网络稳定性等。2021 年，巴西的国内自治系统协议总量几乎跌至 2020 年同期的一半（65097 至 32762），是受疫情影响协议数波动最大的国家。与此同

时，韩国的自治系统协议体量最高在 1399,且自 2009 年以来便未出现明显波动。虽然韩国国土面积较小，但相较于其他许多相同体量的国家而言，其自治系统总量仍然偏小（日本，意大利的体量均为其多倍）。因此自治系统体量可以说明该地区网络发展的活跃程度，但未必能够体现网络的质量和先进程度。

其次，我们发现影响一个地区在网络波动到来时抗干扰能力，与重心在国内还是国外贸易并无绝对关联。自治系统级别协议量的波动，主要原因在于顾客方运营商出于节约成本或产业升级考虑，主动撤销协议。因此，疫情期间所见的协议体量波动，最大的影响因素是经济环境。这一结论的支撑来源于，我们发现对于所有自治系统网络受疫情影响的国家，其波动时间段与这一国家自身疫情进展、政策管制最严重的时期是吻合的。且大部分国家最大自治系统外贸国是美国，而美国的自治系统网络并未受到严重波折。因此我们可以推断出这些波动是顾客方主动操作形成的。

3.7.3 运营商

在疫情冲击下，运营商选择了通过取消自治系统协议（猜测是以取消或不主动续签合约的方式）来节省开支。事实上，除了新冠外，在运营商遇到资金短缺，或进行设施架构升级时，取缔部分自治系统协议并在日后重新建立连接是非常常规的维护手段。在“3.5.2 单个自治系统的自我中心网络”中，我们看到对于单个自治系统，经常出现度数减少，随后再次增加成团的情况。但新冠疫情期间，是影响最大，最集中的一次（见图 6 中 2021 及 2023 数据）。

但新冠带来的影响或许不至于此。疫情期间，随着各个地区的情况依次恶

化，社会秩序受到影响，人们大范围转而使用线上办公。世界总体互联网流量上升了 30-50% [11]。这一变化幅度之大，速度之快，令许多区域运营商措手不及，也随之产生了某些基础设施不完善地区的网络堵塞、缓慢、故障频繁。为应对这一局面，互联网企业、厂家、网络提供商采用的方法包括：降低流媒体画面质量 [12]，取消带宽上限 [13]，扩建基础设施，全面拥抱 5G 等等。可以说，新冠冲击试探了原有网络的抗压能力，并带动了一波产业升级。

此外，一些在疫情冲击下破产的运营商也可能面临收购、重组的局面，这可能使得疫情后的聚集效应更强。

第四章 结论

本文通过对 CAIDA 自治系统关系网络构成的图分析，深入探讨自治系统网络技术和经济方面的相关性，发现了若干规律。从体量上看，自治系统（图中的节点）总量符合线性增长函数，自治系统协议（图中的边）总量符合指数增长函数，全图多数时间为一到两个高度聚集的连通分量，节点度数分布极度不均，长尾效应明显。在网络演变方面，协议量的波动大于节点量的波动，波动存在报复性反弹现象，总体趋于平稳。自治系统（及协议）的增量减量比值自 2001 年至 2023 年缓慢下降，2021 年协议比值首次跌破 1，目前两比值均在 1 左右，说明自治系统网络的增长趋势逐渐放缓。新增边模式倾向于连接新节点与网络初期的老节点（2000 年左右），表明老节点的影响力持续至今。从社群属性上说，2023 年大小排名前 8 的社群占领了 90% 的新增减协议，表明网络活跃度集中于大社群。此外，我们也对部分典型自治系统进行了案例剖析并绘制了以国家为单位的社群演变世界图。最后，新冠疫情对网络的影响启示我们，当今互联网已进入存量发展阶段，运营商可以通过取缔协议节缩开支，但最终改善基础设施和架构升级才能增强网络的健壮性以应对后疫情时代的网络需求。

参考文献

- [1] Damien Magoni, Jean-Jacques Pansiot. Analysis of the Autonomous System Network Topology. Computer Communication Review, 2001, 31 (3), pp.26-37. 10.1145/505659.505663. hal-0034447
- [2] Yue Zhang, Guozheng Yang, Zhihao Luo. Research on Topology Evolution of Autonomous System Network[J]. The 11th International Conference on Communication and Network Security, 2021: 66-79.
- [3] Liu Xiao, Wang Jinfa, Jing Wei, de Jong Menno, Tummers Jeroen S, Zhao Hai. Evolution of the Internet AS-level topology: From nodes and edges to components. Chinese Physics B, 2018, 27(12): 120501
- [4] L. Gao, On Inferring Autonomous System Relationships in the Internet, IEEE/ACM Transactions on Networking, December 2001.
- [5] L. Subramanian, S. Agarwal, J. Rexford, and R. H. Katz, Characterizing the Internet Hierarchy from Multiple Vantage Points, IEEE INFOCOM, 2002.
- [6] G. Di Battista, M. Patrignani, and M. Pizzonia, Computing the Types of the Relationships between Autonomous Systems, IEEE INFOCOM, 2003.
- [7] T. Erlebach, A. Hall, and T. Schank, Classifying Customer-Provider Relationships in the Internet, Proceedings of the IASTED International Conference on Communications and Computer Networks (CCN), 2002.
- [8] M. Luckie, B. Huffaker, k. claffy, A. Dhamdhere, and V. Giotsas, AS Relationships, Customer Cones, and Validation , in Internet Measurement Conference (IMC), Oct 2013, pp. 243--256.
- [9] The CAIDA AS Relationships Dataset, <19980101-20230101> [ps://www.caida.org/catalog/datasets/as-relationships/](https://www.caida.org/catalog/datasets/as-relationships/)
- [10] <https://bgp.potaroo.net/cidr/autnums.html>
- [11] <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>
- [12] <https://techcrunch.com/2020/03/20/amazon-follows-netflixs-lead-reducing-streaming-quality-in-europe/>
- [13] <https://techcrunch.com/2020/03/12/coronavirus-could-force->

isps-to-abandon-data-caps-forever/

致谢

本文的是在导师的指导下完成的，从开题到构思、包括最后提炼，都得到了老师的细心指点和提携。老师对写作和科研的认真与执着给我留下了十分深刻的印象，同时也给予了我莫大的帮助，仅在此献上我最真挚的敬佩与感激。

同时也要在此感谢帮助和支持我的学长学姐们，他们在自己也很忙碌的情况下帮助解决了我的很多问题，many thanks~