



HAL
open science

Mesure de l'influence d'une pandémie sur l'écosystème de peering Internet

Justin Loye, Sandrine Mouysset, Katia Jaffrès-Runser

► To cite this version:

Justin Loye, Sandrine Mouysset, Katia Jaffrès-Runser. Mesure de l'influence d'une pandémie sur l'écosystème de peering Internet. 23èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications (AlgoTel 2021), Université de La Rochelle, May 2021, La Rochelle, France. hal-03211336

HAL Id: hal-03211336

<https://hal.science/hal-03211336v1>

Submitted on 28 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mesure de l'influence d'une pandémie sur l'écosystème de peering Internet.

Justin Loye¹, Sandrine Mouysset² et Katia Jaffrès-Runser³

¹IRIT & LPT, Université de Toulouse, UPS, Toulouse, France - Email : prenom.nom@irit.fr

²IRIT, Université de Toulouse, UPS, Toulouse, France - Email : prenom.nom@irit.fr

³IRIT, Université de Toulouse, Toulouse INP, Toulouse, France - Email : prenom.nom@irit.fr

Des changements majeurs sont observés dans le jeu de données PeeringDB entre janvier 2020 et juin 2020 : une nette augmentation de la capacité provisionnée dans le réseau Internet coïncide avec le confinement soudain d'une grande partie de la population en Europe, dans les Amériques et en Asie. PeeringDB présente l'interconnexion globale des réseaux au niveau des points d'échange Internet (IXP), des centres de données et des systèmes autonomes (AS). Cet article examine quels acteurs de l'écosystème de peering Internet bénéficient de ces changements et où ils se situent dans le monde. Pour cela, nous modélisons un instantané de PeeringDB sous la forme d'un graphe bipartite dirigé pondéré qui capture la taille de port que possèdent les AS au niveau des IXP en utilisant les méta-données disponibles. Ce nouveau modèle d'Internet est analysé pour illustrer les caractéristiques pertinentes d'un réseau complexe qui regroupe les 10268 AS et les 734 IXP dans des zones géographiques d'influence. Une analyse de la matrice de Google réduite révèle l'augmentation de la capacité déployée par les grands réseaux de diffusion de contenu tels que Netflix ou Facebook vers les fournisseurs d'accès Internet d'Espagne, du Brésil, d'Inde, de Pologne ou d'Italie.

Mots-clefs : réseaux complexes, covid-19, peering, matrice de Google

1 Introduction

Internet est structuré autour de systèmes autonomes (AS), administrés par différents acteurs tels que des fournisseurs d'accès Internet (FAI), des fournisseurs de contenu ou des universités entre autres. Ils sont répartis tout autour du globe et s'interconnectent en différents points de présence soit directement, via des connections privées, soit au travers d'un accord de *peering* réalisé au sein d'un *Internet eXchange Point* (IXP). Un IXP est un organisme, généralement à but non lucratif, qui administre une infrastructure d'interconnexion entre AS. Typiquement, une AS qui participe à un IXP est en mesure d'échanger des données avec les autres AS présentes à cet IXP selon un accord de peering. Au début des années 2000, la majeure partie des AS se connectent en privé, selon une relation client/fournisseur, induisant une topologie hiérarchisée d'Internet. La dernière décennie a vu un déploiement d'envergure des IXP, induisant un aplatissement de la topologie d'Internet [GILO11]. Pour signaler leur présence à un IXP et susciter de nouveaux accords, les AS s'enregistrent dans la base de données PeeringDB [LLD⁺14] en indiquant, entre autres, leur identifiant (ASN), leur nature (`Network Type`), une estimation du rapport entre trafic entrant et sortant de l'AS (`info_ratio`) et leur taille de port avec l'IXP (`port_size`). Les accords de peering et le trafic réel entre les AS au sein d'un même IXP ne sont pas renseignés car ce sont des données sensibles.

Depuis décembre 2019, le virus Covid-19 a touché tous les continents, se propageant dans le monde entier en quelques mois depuis la Chine vers le reste de l'Asie, l'Europe, les Amériques, l'Afrique et l'Océanie. Une augmentation du trafic a été observée au sein d'un IXP Européen majeur [FGL⁺20]. L'objectif principal de ce papier est de comprendre quand, où et comment cette augmentation du trafic s'est produite. Cette dernière question se résume au fait d'identifier les entités du réseau qui créent et bénéficient de cette augmentation du trafic. Pour atteindre ces objectifs, nous nous appuyons sur deux méthodes d'analyse des données : *i*) une analyse temporelle des `port_capacity` accumulés pour différents ensembles régionaux de IXP; *ii*) une analyse complexe du réseau PeeringDB, qui s'appuie sur notre nouveau modèle de réseau bipartite dirigé et pondéré, le *PeeringDB capacity graph* (ou *pDB c-graph*). La construction de la matrice de

Google réduite [FJRS16] de ce graphe permet de mettre en lumière l’investissement de grands fournisseurs de contenu pour renforcer leurs liens avec certains réseaux d’accès régionaux.

2 Impact de la pandémie sur la capacité

La capacité d’un IXP (d’un AS), mesurée par la somme du `port_size` de tous les ports déclarés par ses AS (avec ses IXP), représente une borne supérieure sur le trafic échangé. Les IXP renseignent leur pays d’installation dans PeeringDB. Ainsi, à partir des instantanés de peeringDB, on peut construire des séries temporelles de capacité par continent ou pays. On applique à ces séries temporelles un filtre de Canny [Can86], outil du traitement du signal, pour détecter les fortes variations de capacité. La Figure 1 présente la réponse du filtre de Canny aux séries temporelles (février 2019 - août 2020) de 4 pays répartis dans 4 régions du globe.

La réponse du filtre est maximale en Avril 2020, période de la première vague de Covid-19 et des confinements associés. Ce pic est aussi visible dans les séries agglomérées par continent où l’Amérique du Sud (0.157) et l’Amérique du Nord (0.126) connaissent les plus fortes variations.

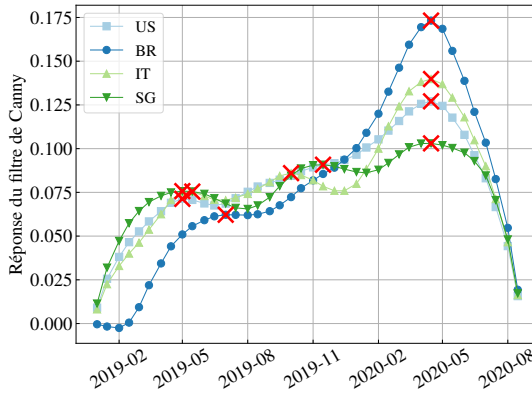


FIGURE 1: Réponse du filtre de Canny à la capacité des IXP agglomérés par pays.

	PageRank graphe pDB	PageRank graphe de [NSD17]	PageRank Inversé graphe pDB
1	IX.br São Paulo	IX.br São Paulo	IX.br São Paulo [+0]
2	DE-CIX Frankfurt	DE-CIX Frankfurt	DE-CIX Frankfurt [+0]
3	AMS-IX	AMS-IX	Facebook [+0]
4	LINX LON1	LINX LON1	AMS-IX [+0]
5	Amazon	EPIX.Katowice	Akamai [+0]
6	Equinix Singapore	NAPAfrica IX	Netflix [+2]
7	NAPAfrica IX	IX.br Rio	LINX LON1 [+0]
8	SIX Seattle	SIX Seattle	Google [-2]
9	IX.br Rio	France-IX Paris	Microsoft [+1]
10	NL-ix	NL-ix	Apple [-1]

TABLE 1: Meilleur classement PageRank et PageRank inversé issu de *pDB c-graph* (colonne 1 et 3), et du graphe non dirigé et non pondéré de [NSD17] (colonne 2), construits pour l’instantané 2020/06/01. Le gain en classement par rapport à 2020/01/01 est précisé pour le PageRank inversé. Les AS sont identifiés en gras.

3 PeeringDB vu comme un réseau complexe

our pouvoir observer plus finement la nature des changements qui ont eu lieu dans PeeringDB entre janvier 2020 et juin 2020, nous avons choisi de modéliser les relations entre AS et IXP à l’aide d’un graphe bipartite capable de mettre en exergue les variations de capacité intervenues dans le réseau. Parmi les quelques travaux mesurant la topologie d’Internet par des données IXP publiques, le modèle de Nomikos et al. [NSD17] est le plus proche du modèle *pDB c-graph* décrit ici.

Modèle Le *pDB c-graph* est construit comme un graphe bipartite, où l’ensemble des sommets \mathcal{V} est réparti entre les ensembles \mathcal{A} et \mathcal{X} représentant les AS et IXP d’un instantané PeeringDB. Les liens non dirigés $e_{i \leftrightarrow j} \in \mathcal{E}$ sont pondérés avec la taille de port $ps(e_{i \leftrightarrow j})$. Si un AS a plusieurs ports à un IXP, la somme des ports pondère le lien. Comparativement à [NSD17], nous utilisons en plus la méta-donnée `info_ratio` caractérisant la proportion entre trafic entrant et sortant d’un AS par les étiquettes “Heavy Outbound”, “Mostly Outbound”, “Balanced”, “Mostly Inbound”, “Heavy Inbound” ou “Not Disclosed”[†]. Nous associons aux liens une direction et une proportion de trafic β : “Heavy *” et “Mostly *” sont associés respectivement à $\beta_H = 0.95$ et $\beta_M = 0.75$. Les AS “Balanced” et “Not Disclosed” sont associés à une proportion 0.5. Le graphe est ainsi décrit par la matrice d’adjacence pondérée W où un AS i étiqueté :

- “* Outbound” : $W_{ij} = \beta \cdot ps(e_{i \leftrightarrow j})$ et $W_{ji} = (1 - \beta) \cdot ps(e_{i \leftrightarrow j})$,
- “* Inbound” : $W_{ij} = (1 - \beta) \cdot ps(e_{i \leftrightarrow j})$ et $W_{ji} = \beta \cdot ps(e_{i \leftrightarrow j})$.

[†]. Environ 82% des AS ont une étiquette valide. Ces AS représentent environ 94 % de la capacité totale du réseau.

Centralité et Diffusivité Le PageRank (PR) du réseau construit à partir du modèle non dirigé et non pondéré de [NSD17] permet d'identifier les sommets centraux du graphe. Le Tableau 1 liste le top 10 des noeuds où l'on observe l'hégémonie des gros IXP. L'un des principaux intérêts de *pDB c-graph* est de distinguer la capacité du trafic entrant du trafic sortant. Ainsi, PR identifie les noeuds les plus susceptibles de recevoir du trafic. On observe dans le Tableau 1 que le top 10 s'en voit modifié, et notamment que l'AS *Amazon* y fait son apparition. Il est possible, en inversant le sens des liens d'un graphe dirigé, de construire la métrique du "PageRank inversé" (PRI). Dans notre cas, on identifie alors les noeuds les plus susceptibles de diffuser du trafic. Le Tableau 1 identifie les géants du contenu dans le top 10 des sommets diffuseurs, corroborant leur rôle incontournable dans l'Internet d'aujourd'hui [LLD⁺14]. De plus, le classement PRI nous permet d'observer des modifications d'envergure dans le graphe avant et après la première vague de Covid-19, avec notamment une augmentation de la diffusivité de Netflix. Aussi, une étude de la sensibilité du graphe aux valeurs de β choisies nous montre que les classements des top-20 en PR et PRI sont stables.

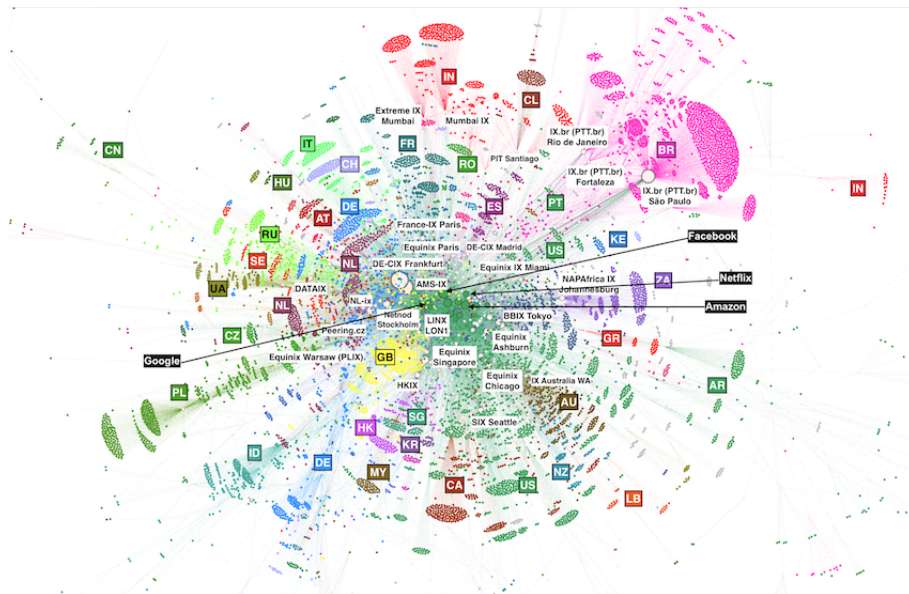


FIGURE 2: Visualisation du *pDB c-graph*. Le code couleur identifie les pays des AS ; les IXP sont en blancs. La taille d'un noeud est proportionnelle à la somme des capacités entrantes et sortantes. Tracé "Force direct layout" (Gephi).

Structure géographique Le graphe comprend 10267 AS reliés par 56808 liens à 734 IXP au 1er juin 2020. Une visualisation est proposée dans la Figure 2. Le pays d'appartenance d'un AS n'est pas présent dans PeeringDB, mais déduit du pays des IXP où l'AS est le plus présent. On retrouve la proximité géographique des AS, illustrant la mission principale de point d'échange local des IXP. On retrouve les géants du contenu à l'épicentre du réseau, liés au label *US* en vert. Cette position souligne une politique de diffusion privilégiée vers les régions du monde par le réseau de peering.

4 Augmentation du trafic régional et fournisseurs de contenu

L'épidémie de Covid-19 a provoqué une demande accrue de contenu de la part des clients domestiques, forçant les "hypergéants" [LLD⁺14] comme Facebook, Google ou Netflix à adapter leur stratégie de peering en augmentant leur capacité aux IXP, comme nous l'avons vu dans la section 2. Les clients accèdent généralement au contenu par un fournisseurs d'accès internet (FAI). Comme les accords de peering sous-jacents ne sont pas répertoriés, nous proposons dans la suite d'identifier les liens du graphe entre ces hypergéants et les principaux FAI régionaux, et d'observer leur variation entre deux instantanés de PeeringDB avant (01-01-2020) et après (01-06-2020) la première vague.

Méthode Notre méthode comprend deux étapes. La première consiste à identifier les AS. Les hypergéants sont identifiés par le top 20 des AS dans le classement PageRank Inversé. Les FAI régionaux importants,

concentrateurs de trafic, sont choisis par pays avec le top 4 des FAI au sens du PageRank. La seconde étape consiste à déterminer les changements de peering entre ces deux types d’AS. Le pDB c -graph étant bipartite, de tels liens ne peuvent pas être observés. Pour les faire apparaître, on utilise l’algorithme de la matrice réduite de Google de Frahm et al. [FJRS16]. Pour un graphe donné et sa matrice de Google G , on calcule le complément stochastique de G , noté G_R , pour un réseau réduit r composé des AS des hypergéants et des FAI sélectionnés. On s’intéresse à la décomposition spectrale de cette matrice de Google réduite qui distingue les liens directs ($r \rightarrow r$) des liens indirects ($r \rightarrow s \rightarrow r$), s étant le réseau complémentaire à r . Ces liens indirects correspondent dans notre cas à l’association d’AS entre elles par leur présence commune à un ou plusieurs IXP. La décomposition du complément stochastique s’écrit $G_R = G_{rr} + G_{pr} + G_{qr}$, avec G_{pr} et G_{qr} représentant les liens indirects. G_{pr} est de rang 1 et ne contient pas d’information car ses colonnes sont colinéaires au vecteur de PageRank des N_r noeuds. Ainsi, on ne s’intéresse dans la suite qu’à G_{qr} .

Résultats On ne retrouve des liens forts dans G_{qr} que pour la diffusion des hypergéants vers les FAI. Il est possible de soustraire G_{qr} du 01-06-2020 à G_{qr} du 01-01-2020 car les colonnes sont normalisées. Ainsi, les éléments positifs ou négatifs (i, j) de cette nouvelle matrice G_{diff} montrent un renforcement ou un affaiblissement des liens de peering entre les AS (i, j) comme illustré dans la Figure 3. Les connections de Netflix ont le plus augmenté, surtout au Brésil, en Espagne, en Pologne et à Hong Kong. Un lien particulièrement fort entre Post Group Luxembourg et Netflix apparaît. Plusieurs hypergéants diffusent vers l’Italie.

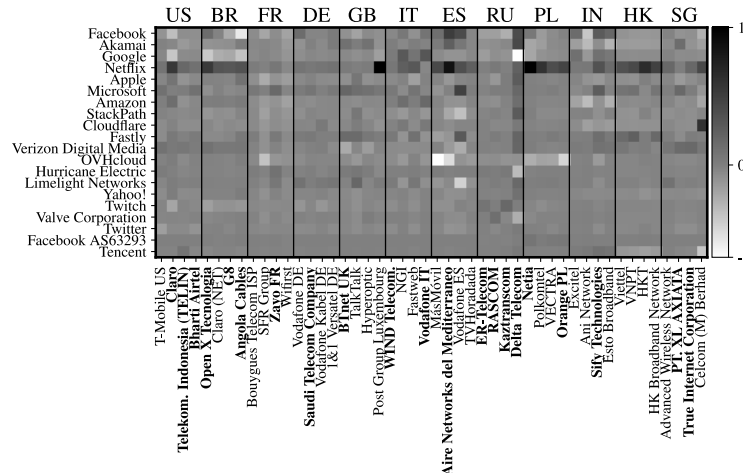


FIGURE 3: Illustration de G_{diff} : variation de la diffusion des 20 hypergéants vers le top 4 des FAI régionaux.

5 Conclusion

Cet article résume des travaux qui proposent un nouveau modèle de réseau Internet pondéré et orienté issu de PeeringDB. Ce pDB c -graph est un outil intéressant pour identifier les relations modifiées par la première vague du Covid-19 au printemps 2020.

Références

- [Can86] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6) :679–698, 1986.
- [FGL⁺20] A. Feldmann, O. Gasser, F. Lichtblau, E. Pujol, I. Poese, C. Dietzel, D. Wagner, M. Wichthuber, J. Tapiador, N. Vallina-Rodriguez, O. Hohlfeld, and G. Smaragdakis. The lockdown effect : Implications of the covid-19 pandemic on internet traffic. In *Proceedings of the ACM IMC*, page 1–18, New York, NY, USA, 2020.
- [FJRS16] K. M Frahm, K. Jaffrès-Runser, and D. L Shepelyansky. Wikipedia mining of hidden links between political leaders. *The European Physical Journal B*, 89(12) :1–21, 2016.
- [GILO11] E. Gregori, A. Improta, L. Lenzini, and C. Orsini. The impact of IXPs on the AS-level topology structure of the Internet. *Computer Communications*, 34(1) :68 – 82, 2011.
- [LLD⁺14] A. Lodhi, N. Larson, A. Dhamdhare, C. Dovrolis, and Kc Claffy. Using peeringDB to understand the peering ecosystem. *ACM SIGCOMM CCR*, 44(2) :20–27, 2014.
- [NSD17] G. Nomikos, P. Sermpezis, and X. Dimitropoulos. Re-mapping the internet : Bring the IXPs into play. In *IEEE INFOCOM WKSHPS*, pages 910–915, Atlanta, GA, USA, 2017.