

Historic Graphs : Cartographie dynamique des liens sur Wikipédia

Antonin Segault

▶ **To cite this version:**

Antonin Segault. Historic Graphs : Cartographie dynamique des liens sur Wikipédia. H2PTM'19 : De l'hypertexte aux humanités numériques, Oct 2019, Montbéliard, France. pp.147-161. hal-02332066

HAL Id: hal-02332066
https://hal.science/hal-02332066v1

Submitted on 28 Jul 2025

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.

Historic Graphs

Cartographie dynamique des liens sur Wikipédia

Antonin Segault

Laboratoire Dicen-IdF
Université Paris Nanterre
200 avenue de la République, 92 000 Nanterre
antonin.segault@parisnanterre.fr

RÉSUMÉ. *Cet article propose d'étudier l'évolution des liens hypertextes qui interconnectent les articles de l'encyclopédie en ligne Wikipédia. Il s'agit d'exploiter les données stockées dans les historiques des versions des articles pour dresser des cartes dynamiques de la structure des liens au cours du temps. Nous présentons Historic Graphs, un outil dédié, puis décrivons son usage à travers quelques cas d'étude. Nous discutons ensuite l'interprétation de ces cartes dynamiques et leur intégration dans une démarche de recherche complète.*

ABSTRACT. *This paper investigates the evolution of the hypertext links that connect articles on Wikipedia's online encyclopedia. We aim to use the data stored in the articles' revision history to build dynamic maps, displaying the changes in link structures over time. We present Historic Graphs, a dedicated tool and describe it use through a few case studies. We then discuss the interpretation of such maps, and how it integrates in a reseach process.*

MOTS-CLÉS : *wikipédia, historiques des versions, liens hypertextes, graphes, visualisations*

KEYWORDS: *wikipedia, revision history, hypertext links, graphs, data visualisation*

1. Introduction

L'encyclopédie en ligne Wikipédia se classe depuis plusieurs années parmi les sites web les plus visités au monde (Alexa, 2019). Avec plus de 48 millions d'articles, répartis dans 302 langues, il constitue également un système d'organisation des connaissances d'une étendue colossale. Visant à « présenter une synthèse raisonnée de l'ensemble du savoir humain établi. » (2019b), Wikipédia s'inscrit dans la tradition encyclopédique du Dictionnaire raisonné des sciences, des arts et des métiers de Diderot et d'Alembert. Son adossement aux technologies de l'information modernes et aux usages du Web collaboratif, surclassant les fiches cartonnées du Mundaneum de Otlet et la Fontaine, lui ont permis d'atteindre une échelle sans précédent dans l'histoire.

À ce titre, Wikipédia représente un objet unique pour l'étude de l'agrégation des connaissances, de l'édition collaborative et de l'organisation des communautés distribuées. Les données que renferme son système d'information permettent une analyse à large échelle de ces processus, selon des approches pouvant relever des Humanités Numériques. Dans cet article, nous présenterons ainsi une méthode pour construire des cartographies dynamiques des liens hypertextes entre les articles de l'encyclopédie, afin de rendre visible l'évolution de leur structure au cours du temps. Nous reviendrons tout d'abord sur les fondements techniques et humains qui sous-tendent la production des articles, puis proposerons des outils et méthodes pour en collecter les traces et les cartographier. Nous décrirons ensuite quelques études de cas visant à discuter les promesses et les limites de cette approche.

2. Processus d'édition collaborative

Le succès de Wikipédia repose sur l'articulation d'une large communauté et d'un outil d'écriture collaborative de type wiki. MédiaWiki, le logiciel utilisé, permet à tout internaute de créer un nouvel article ou de modifier un article existant, sans même nécessiter une inscription. Néanmoins, la fiabilité des articles ainsi produits s'avère étonnement bonne, tout juste inférieure aux encyclopédies traditionnelles, pour un nombre de sujets traités largement supérieur (Rogers, Sendijarevic et others, 2012). Ce résultat, illustration du concept de « sagesse des foules » (Surowiecki et Riot, 2008), découle d'une « vigilance participative » de toute la communauté (Cardon et Levrel, 2009).

De fait, pour assurer la qualité des contenus et le bon fonctionnement du projet Wikipédia, le processus d'écriture se double d'activités de discussion, de vérification et de régulation : « les wikipédiens ne se contentent pas d'écrire l'encyclopédie. Ils discutent, se disputent, négocient et s'accordent sur les pages Discussion. Ils se modifient, se corrigent et se révoquent sur les pages Historique. Ils se testent, s'amusent et apprennent dans les pages Bac à sable et Bistrot. Ils montent la garde, veillent, vérifient, corrigent, protègent et bloquent à l'aide de la Liste de suivi et des pages de Modifications récentes. » (Cardon et Levrel, 2009). Ces activités reposent

sur un ensemble de règles et de bonne pratiques codifiées, elles aussi développées par la communauté au fil des années (Langlais, 2015). Ainsi, Wikipédia n'est pas seulement une encyclopédie hypermédia, mais également un véritable « lieu de production et de transmission de connaissances » (Beaude, 2015).

3. Étude de Wikipédia

Les caractéristiques spécifiques de Wikipédia et sa place particulière dans le Web actuel en font un objet de recherche unique. De nombreux travaux ont porté sur sa communauté, son processus d'édition collectif et les règles complexes qui le régissent (Sahut, 2016 ; Langlais, 2015). Mais cette encyclopédie peut également constituer un terrain de recherche pour les sciences humaines et sociales, à travers lequel étudier l'évolution des sujets qui y sont décris (Gorgeon et Swanson, 2009). Il s'agit alors d'exploiter les données relatives aux articles et à leur développement pour analyser la société qu'ils dépeignent, la société qui s'y dépeint elle-même.

L'étude des articles de Wikipédia comme sources de données nécessite de « suivre le médium » (Rogers, 2009), c'est-à-dire de prendre en compte l'ensemble des règles, pratiques et affordances qui ont affecté leur rédaction : « La Source Nativement Numérique doit donc être pensée dans le contexte du système d'information où elle a été produite » (Paloque-Berges, 2016). Là où Diderot et d'Alembert ont usé des renvois entre articles pour permettre une navigation transversale à travers l'Encyclopédie (Blanchard et Olsen, 2002), le système wiki s'appuie sur des liens hypertextes. Les contributeurs de Wikipédia sont invités à utiliser les liens pour matérialiser les « connexions pertinentes au sujet d'un autre article, qui aidera le lecteur à mieux comprendre l'article » (2019a). Ces liens, pointant vers des « sujets d'intérêt » (*matters of concern*) (Latour et others, 2005 ; Weltevrede et Borra, 2016), représentent par conséquent d'importantes clefs pour l'analyse des articles.

L'interconnexion des pages de Wikipédia peut être décrite comme un graphe, dans lequel des arcs orientés décrivent les liens hypertextes entre des nœuds représentant les articles. Cette structure se prête bien à des représentations cartographiques, à l'image de l'outil Seealsology (DensityDesign et Médialab Science Po, 2019), qui génère des visualisations interactives à partir du réseau de liens inclus dans les sections « Articles connexes » (*see also* sur la version anglophone de l'encyclopédie). À notre connaissance, de telles représentations n'ont pas encore été exploitées pour analyser les données de l'historique des articles. Plusieurs outils ont été créés pour visualiser le développement d'un article au cours du temps : Historic Flow montre les portions de texte conservées et modifiées entre chaque version (Viégas, Wattenberg et Dave, 2004), tandis que Contropedia analyse les ajouts et suppressions des liens pour identifier les sujets de controverses dans les articles (Borra et al., 2015). Mais aucun ne permet à ce jour de produire des cartographies dynamiques de Wikipédia, montrant l'évolution temporelle de la structure de liens entre un ensemble d'articles.

4. Cartographie dynamique de Wikipédia

La construction de cartes dynamique de Wikipédia nécessite l'analyse des différentes versions stockées dans les historiques des articles. En raison du nombre d'articles de l'encyclopédie, la carte ne peut en présenter qu'une infime portion, limitée à quelques centaines, pour des questions de lisibilité mais aussi de faisabilité technique. De même, si l'historique d'un seul article peut contenir des centaines de versions, la collecte de données doit se limiter à un nombre restreint, ne dépassant pas quelques dizaines de points temporels. En dépit de ces restrictions, la production manuelle d'une telle carte représenterait un travail considérable, impliquant la consultation et l'analyse de plusieurs milliers de versions de pages, pouvant chacune contenir plusieurs centaines de liens. Nous proposons par conséquent l'automatisation du processus de collecte et d'extraction de la structure des liens.

À l'aide du langage Python, nous avons développé Historic Graphs, un dispositif de fouille des historiques des articles Wikipédia. Cet outil s'appuie sur les interfaces de programmation (API – *Application Programing Interface*) de MédiaWiki, le programme sur lequel est bâti l'encyclopédie. Ces interfaces permettent d'accéder une grande variété de données sur des articles. Historic Graphs prend pour entrée les noms d'un ou plusieurs articles, ainsi qu'une série de dates. À l'aide des APIs, le programme identifie les versions des articles en vigueur à ces dates, et collecte les listes des liens présents dans ces versions. Il peut ensuite reproduire ce processus pour l'ensemble des liens trouvés, de manière récursive, jusqu'au niveau de profondeur souhaité. Tous ces liens, caractérisés par un article source, un article cible et un point temporel, sont combinés dans une même structure, constituant un graphe de « zone » (Mercklé, 2011), centré sur le (ou les) articles de départ. Ce graphe est ensuite exporté sous la forme d'une série de fichiers CSV listant les liens existant pour chaque date.

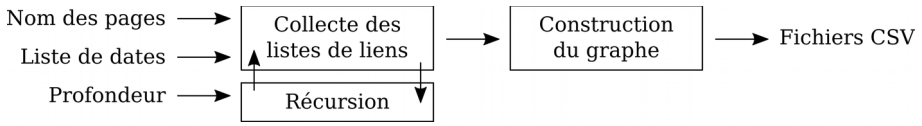


Figure 1. *Schéma de principe du processus de collecte de données*

La comparaison de ces fichiers permet notamment d'identifier les liens étant apparus ou ayant été retirés entre deux versions. Cependant, au vu de la masse de données, des techniques de visualisation des données sont plus adaptées pour l'analyse des corpus ainsi collectés. La structure des fichiers CSV produits par Historic Graphs est compatible avec le logiciel Gephi. Cet outil de traitement des graphes offre de nombreuses fonctionnalités pour leur représentation graphique, mais aussi pour le calcul de métriques issues de l'analyse structurale des réseaux

(Bastian, Heymann, Jacomy et others, 2009). Il dispose en outre d'un système de chronologie permettant la manipulation de graphes dynamiques, dans lesquels des nœuds ou des liens sont associés à des intervalles ou à des points temporels. Une frise temporelle permet de filtrer l'affichage d'une période donnée.

La seule importation des fichiers de données dans Gephi n'est cependant pas suffisante pour la création de cartes lisibles et pertinentes. La mise en forme du graphe nécessite de nombreux paramétrages, afin de mettre en lumière les évolutions intéressantes de la structure des liens. Il n'existe pas de « recette miracle » : selon les caractéristiques du corpus analysé et les objectifs de l'étude, différentes approches peuvent être envisagées, nécessitant généralement un travail d'ajustements successifs avant d'aboutir à une représentation satisfaisante. La carte résultante est issue des données, mais aussi et surtout de l'outil de visualisation et des choix des chercheurs qui le manipulent.

5. Études de cas

Afin de présenter les possibilités d'un tel outil, les analyses qui peuvent s'en nourrir, et les questionnements méthodologiques qu'elles suscitent, nous allons décrire la réalisation de quelques cartes, portant sur un article Wikipédia intitulé « Attentats du 13 novembre 2015 en France ». Décrivant les attaques terroristes survenues au Bataclan, dans les rues des onzièmes et douzièmes arrondissements parisiens ainsi qu'aux abords du Stade de France, cet article a aujourd'hui atteint une taille et un niveau de détail conséquents. Surtout, sa rédaction a commencé dès les premières heures des événements, rassemblant de nombreux contributeurs pour compiler et croiser des informations multiples et parfois contradictoires. Le développement de cet article constitue ainsi un cas particulièrement intéressant pour l'étude des dynamiques de l'encyclopédie.

5.1. *Premières cartes et observations*

Nous avons dans un premier temps utilisé Historic Graphs pour cartographier les liens dans deux versions issues des premières heures de l'article sur le Wikipédia Français : la première, datant du 13 novembre à 23h UTC (minuit heure française), et une autre datant du 14 novembre à 4h UTC[1] (5h heure française). La collecte, limitée à un seul niveau de profondeur (seulement les articles directement liés à « Attentats du 13 novembre 2015 en France »), a duré environ 25 minutes, aboutissant à deux fichiers CSV correspondant aux deux versions temporelles.

Une fois importés dans Gephi, les nœuds ont été positionnés avec l'algorithme de spatialisation Force Atlas 2. Pour générer ces cartes, nous avons choisi de spatialiser avec l'ensemble de nœuds affichés, avant d'utiliser la frise chronologique pour les

[1] Les heures UTC seront utilisées dans cet article, car il s'agit du fuseau horaire de l'API, quand bien même la collecte de données porte sur le wiki français.

filtrer selon l'heure (Figure 2). Cette approche permet d'assurer que chaque nœud gardera la même position sur l'ensemble des intervalles temporels. Mais on peut également relancer la spatialisation à l'affichage de chaque intervalle, ou même manipuler la frise chronologie pendant que le programme de spatialisation est lancé, afin de visualiser dynamiquement le rapprochement ou l'éloignement différents articles ou groupes d'articles au cours du temps.

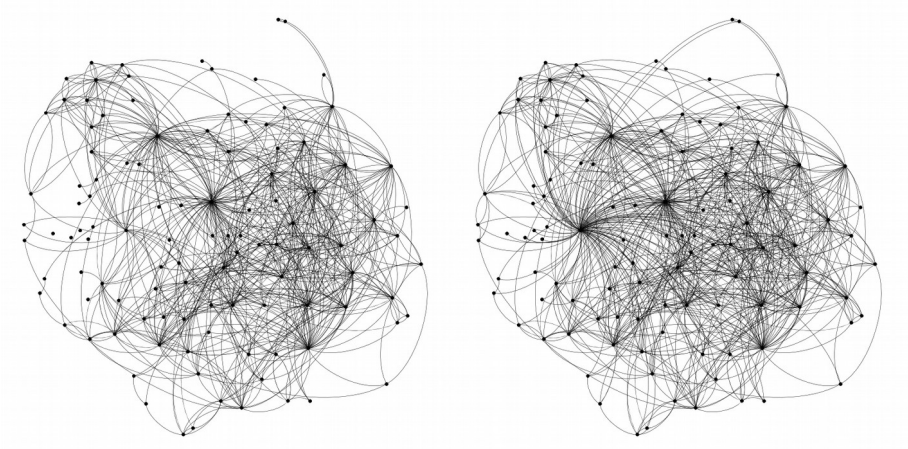


Figure 2. *Cartes des liens de l'article « Attentats du 13 novembre 2015 en France » le 13 novembre à 23h (gauche) et le 14 novembre 2015 à 4h UTC (droite)*

On constate que les cartes présentent de très nombreuses connections, même pour le premier repère temporel. En effet, le programme capture l'ensemble des liens existants entre les articles à la date donnée. Si à 23h le 13 novembre 2015 l'article « Attentats du 13 novembre 2015 en France » n'en est qu'à ses balbutiements, ne contenant que 25 liens, les autres articles – mentionnés dans cette version ou dans la suivante – forment déjà un réseau dense. Entre les deux versions, la principale différence est le développement rapide de l'article « Attentats du 13 novembre 2015 en France », dont les liens forment une étoile très visible à la gauche de la seconde carte. Au vu des statistiques fournies par Gephi, il s'avère que ce développement est principalement incrémental : seul deux des liens présents dans la première version ne le sont pas dans la seconde, tandis que 84 autres ont été ajoutés.

Ces premières observations montrent que la seule lecture des cartes n'est que rarement suffisante pour comprendre le processus d'édition des articles. Les graphes peuvent permettre l'identification de motifs ou de phénomènes particuliers, mais pas leur explication. L'analyse nécessite généralement un mouvement va-et-vient entre la carte, les données brutes issues de Historic Graphs, des métriques statistiques (fournies par Gephi ou d'autres outils) et, parfois, la consultation des articles Wikipédia dans un navigateur web.

5.2. *Densité et groupes d'articles*

Si la densité des réseaux de liens entre les articles complique la lecture des cartes, plusieurs méthodes peuvent être employées pour tenter d'y remédier. Tout d'abord, il apparaît que parmi les nœuds ayant le plus haut degré de centralité – c'est-à-dire les plus connectés aux autres – certains présentent peu d'intérêt pour l'analyse. Il s'agit souvent d'articles décrivant des pays ou des dates, qui sont mentionnés dans des articles n'ayant aucun autre élément en commun. Ces nœuds occupent ainsi des positions de *bridging*, reliant des parties du graphe qui autrement seraient déconnectées. La suppression de quelques-uns de ces nœuds très centraux peut, dans certains cas, rendre la carte plus lisible, favorisant le repérage des groupes d'articles fortement interconnectés (ou *clusters*). Nous avons repris la seconde des cartes précédemment décrite, qui présentait les articles liés à « Attentats du 13 novembre 2015 en France » le 14 novembre 2015 à 4h. À partir de ce graphe, six nœuds ont été supprimés : « Paris », « 2015 », « 13 novembre », « 14 novembre », « 2008 », et « Novembre 2018 ». On remarque une plus grande lisibilité de la carte, en bas de laquelle apparaît un *cluster* assez isolé, comprenant des articles liés à la géographie de Paris : rues, places, arrondissements… (Figure 3).

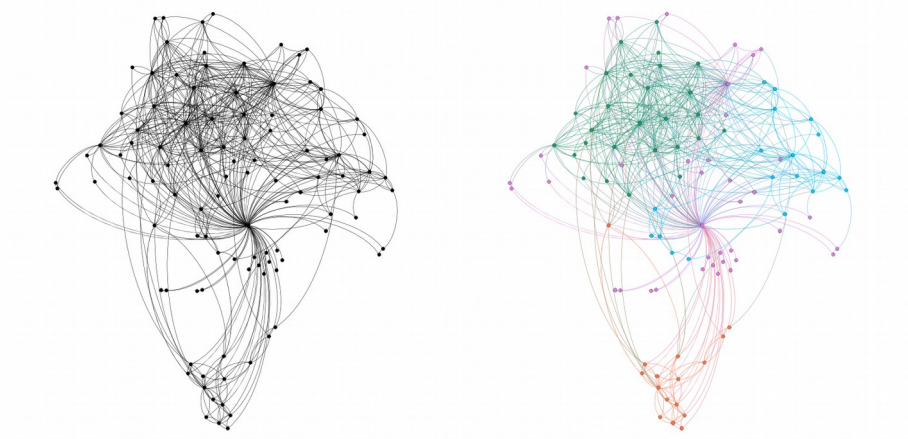


Figure 3. *Cartes des liens de l'article « Attentats du 13 novembre 2015 en France » le 14 novembre 2015 à 4h, après suppression de nœuds centraux (gauche) et après application d'un algorithme de détection des communautés (droite)*

L'identification des *clusters* peut également être facilitée par les outils d'analyse intégrés à Gephi. L'algorithme de détection de communauté (intitulé Modularité, dans l'interface du logiciel) permet en effet de décomposer le graphe en plusieurs sous-ensembles, sur la base des liens existant entre les nœuds. Plusieurs paramètres peuvent être modifiés, affectant le temps de calcul mais aussi le nombre de communautés susceptibles d'être identifiées. Les « classes de modularité » ainsi

détectées peuvent ensuite être intégrées à la visualisation, notamment en colorant les nœuds correspondants. Ici, par exemple, la configuration de l'algorithme a permis la constitution de quatre grands ensembles dans la carte (Figure 3). Un premier groupe, coloré en rose, rassemble des articles relatifs au déroulement de l'attentat. Le cluster de la géographie parisienne a bien été identifié, en orange, mais quelques nœuds sans rapport – Eagles of Death Metal et Wikinews – y ont été ajoutés en raison de leur proximité dans le graphe. Un troisième groupe, en bleu, rassemble des articles relatifs aux institutions françaises – l'Élysée, François Hollande, Laurent Fabius – mais aussi quelques dirigeants étrangers – Barack Obama, Angela Merkel, David Cameron, Benjamin Netanyahou – et, plus étonnant, les articles relatifs à Twitter et aux hashtags. Une quatrième communauté, en vert, regroupe la plupart des pays et personnalités politiques dont les condoléances ont été rapportées dans l'article sur les attentats.

Il apparaît que les résultats de l'algorithme de détection de communauté ne correspondent pas toujours à des groupes d'articles pouvant être rattachés à une thématique commune. Son usage nécessite souvent une démarche d'essai-erreur, au cours de laquelle les paramètres sont modifiés jusqu'à obtenir un nombre et une distribution des groupes qui soit ensuite interprétable avec plus ou moins de pertinence. Si l'algorithme de détection de communautés peut faciliter l'identification de groupes d'articles, ses résultats sont souvent liés à des attentes ou des idées préconçues du chercheur quant aux groupes qu'il pense pouvoir ou devoir apparaître dans la carte.

5.3. *Développement progressif d'un article*

Différents paramètres visuels de la carte peuvent également être ajustés afin de mieux rendre compte de l'évolution d'un article au cours du temps. Si les algorithmes de spatialisation basés sur des forces – tels que Force Atlas 2 (Jacomy, Venturini, Heymann et Bastian, 2014) utilisé plus tôt – facilitent l'identification de groupes ou clusters de nœuds interconnectés, l'article utilisé comme point de départ de la collecte de données (et qui est par conséquent connecté à tous les autres, au moins dans le cas d'une collecte à un niveau de profondeur) peut se retrouver isolé au milieu des groupes, dans une position très centrale, difficilement interprétable. D'autres algorithmes de spatialisation tels que Fruchterman Reingold – également basé sur des forces, mais produisant des graphes plus réguliers – sont parfois plus adaptés (Figure 4). Ici, par exemple, les liens de l'article « Attentats du 13 novembre 2015 en France » ont été cartographiés heure par heure (du 13 novembre à 22h – moment exact de la création de l'article – au 14 novembre à 3h). Pour faciliter la lecture, les liens issus de cet article central sont colorés en noir tandis que les autres restent grisé.

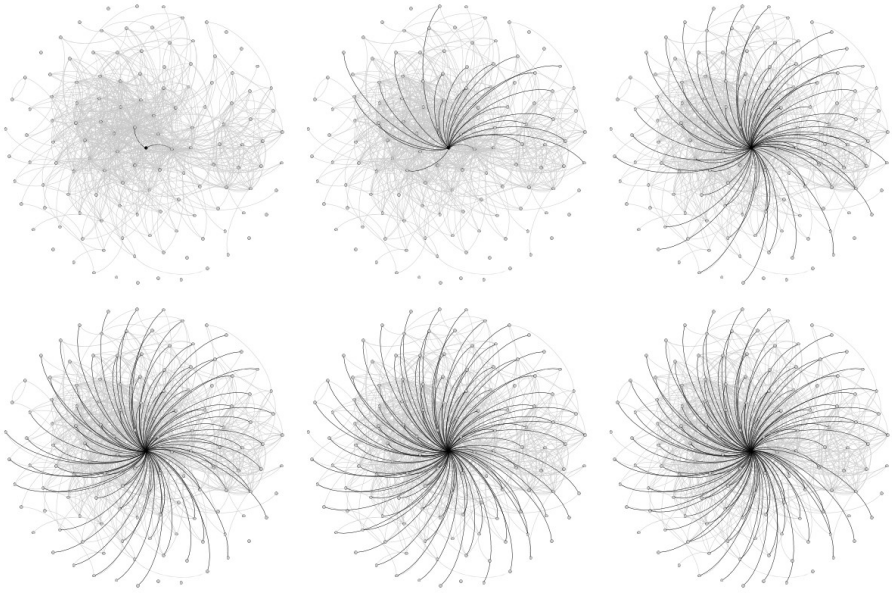


Figure 4. *Cartes des liens de l'article « Attentats du 13 novembre 2015 en France » le 13 novembre à 22h, 23h et minuit, puis le 14 novembre 2015 à 1h, 2h et 3h.*

Ces cartes semblent confirmer le caractère incrémental du processus de développement de l'article. En effet, le nombre de liens issus de l'article augmente visiblement à chaque intervalle temporel, tandis que très peu de liens en sont supprimés entre chaque version. En revanche, la lecture des cartes ne renseigne guère sur la nature des articles mentionnés au cours des différentes versions. Sont simultanément ajoutés des liens relatifs au déroulement des attaques (lieux, armes, etc.), aux opérations de polices y répondant (acteurs, plans d'action) et aux hommages (pays et dignitaires). Le déséquilibre vers la partie haute de la carte à 23h ne semble pas pouvoir être rattaché à un quelconque *cluster* d'articles similaires par leur thématique ; il s'agirait plutôt d'un artefact de l'algorithme de spatialisation.

La comparaison de plusieurs cartes disposées côte-à-côte, comme ici, n'est pas toujours aisée. Il est difficile de déterminer quels liens ont été ajoutés ou supprimés. Cela s'avère plus simple lorsque l'on fait défiler les cartes superposées, à la manière d'un diaporama. Ce type de comparaison peut être réalisée directement dans Gephi, en utilisant les options de lecture de la frise chronologique, mais aussi en assemblant des séquences d'images exportées sous forme de courtes vidéos[2].

[2] Les six cartes précédemment présentées ont ainsi été assemblées sous forme d'un GIF animé, non intégré dans cet article pour des raisons évidentes, mais visible sur https://framagit.org/retrodev/historic-graphs

5.4. *Évolution des liens périphériques*

Plutôt qu'à l'évolution de l'article lui-même, on peut aussi s'intéresser à la manière dont les événements conduisant à la création ou la modification d'un article affectent les autres articles qui l'entourent. Il s'agit alors de déterminer dans quelle mesure le processus d'édition est concentré sur un seul article ou, au contraire, se déploie également vers les articles liés.

Pour éclairer ces questions, deux méthodes ont été combinées. Tout d'abord, nous avons collecté les liens issus de l'article « Attentats du 13 novembre 2015 en France » le 13 novembre à 21h – c'est-à-dire avant sa création – et le 14 novembre à 12h. Si l'idée d'une cartographie des liens avant la création de l'article peut paraître contre-intuitive, elle reste cohérente avec le fonctionnement de Historic Graphs : l'outil dresse une liste de tous les articles mentionnés dans les deux versions, puis dresse des cartes correspondant à leur structure aux deux moments. Sur la carte du 13 novembre, l'article « Attentats du 13 novembre 2015 en France » apparaît comme un nœud isolé – car, n'existant pas, il ne contient aucun lien – mais les autres articles sont bien présents et interconnectés.

La principale différence entre les deux cartes serait évidemment l'apparition de nombreux liens vers et depuis l'article « Attentats du 13 novembre 2015 en France ». Pour mieux visualiser les changements extérieurs à cet article, nous avons donc supprimé le nœud le concernant. Les cartes ne représentent alors plus que l'évolution de la structure des liens périphériques, entre ces articles connexes. Pour éviter l'écueil que constitue la comparaison de cartes côte-à-côté, nous avons cette fois exploité l'apparence des liens, en sélectionnant manuellement dans Gephi puis en colorant ceux apparus entre les deux versions de la carte (Figure 5).

Parmi les 34 nouveaux liens périphériques ainsi mis en évidence, plusieurs distinctions peuvent être opérées. De nombreux liens vers les articles de dates (« 13 novembre », « Novembre 2015 », etc.) semblent indiquer l'ajout de mention des événements dans des articles relatifs aux lieux (« Bataclan », « Stade de France », etc.) et aux acteurs (« Eagles of Death Metal ») impliqués dans les événements, de même que des liens entre ces articles. D'autres ajouts de liens, concernant les dispositifs de sécurité (« Plan Rouge », « État d'urgence en France », etc.) témoignent plutôt de modifications ou d'activation de ces mesures en réponse aux attentats. Enfin, il est difficile de déterminer si certains ajouts sont liés aux événements ou fortuits. Par exemple, la création de liens entre « Bombardements aériens » en « Syrie » ne peut être directement associée au développement de l'article « Attentats du 13 novembre 2015 en France ». Il se pourrait néanmoins qu'elle résulte du regain d'attention pour la lutte contre l'organisation État Islamique consécutif aux attentats.



Figure 5. *Carte des liens périphériques de l'article « Attentats du 13 novembre 2015 en France » montrant les ajouts entre le 13 novembre à 21h et le 14 novembre 2015 à 12h*

L'analyse des liens périphériques nécessite de démêler les relations complexes qui traversent les processus d'édition d'articles associés à un même événement. La carte à elle seule ne rend compte que de la simultanéité – ou de l'antériorité/postérité – de plusieurs modifications des liens entre articles, et ne permet pas d'établir une éventuelle relation de causalité. L'examen des différentes versions de l'article, des métadonnées accompagnant la modification (pseudonymes ou adresses IP, commentaires, etc.) et des pages de discussion s'avère souvent nécessaire. Des représentations cartographiques associant les articles et leurs contributeurs – toujours de manière dynamique – pourraient également être envisagées.

6. Discussions

À la lumière de ces analyses, plusieurs éléments de discussions peuvent être relevés, quant au processus de rédaction des articles, à son étude par le biais de cartes dynamiques, mais aussi à la fabrication de ces cartes.

Tout d'abord, nous avons vu que les articles de Wikipédia forment un graphe dont la densité importante complique l'analyse. Certains nœuds peu spécifiques et

donc très centraux – tels que les dates ou les grands espaces géographiques – peuvent être supprimés pour faciliter la lecture des cartes, sans que cela n'en réduise substantiellement le contenu informationnel. Néanmoins, le nombre de connexions réellement informatives reste toujours élevé, même entre des articles relatifs à des sujets éloignés. Il apparaît ainsi que la structure hypertextuelle de l'encyclopédie ne constitue pas un réseau multipolaire, formé d'ensembles thématiques lâchement connecté entre eux, mais un seul réseau très dense. Les nouveaux nœuds et les nouveaux liens qui sont créés, y compris en réponse à des événements d'actualité exceptionnels, s'insèrent alors dans cette structure existante, sans y amener de restructuration majeure.

L'analyse des cartes dynamiques de liens nécessite par ailleurs de prêter attention aux différentes temporalités qui se mêlent – et se confondent parfois – dans le développement des articles. Les temps des faits (ici, l'attentat) et de leur récit par des sources primaires ou secondaires (notamment les médias) ne correspondent pas toujours au temps du développement de l'article. En dépit de la visée encyclopédique de Wikipédia, la rédaction d'un article peut débuter très vite, avant que les faits ne soient tous connus, ou même avant qu'un événement ne se soit achevé. Dans d'autres cas, les modifications peuvent être retardées ou plus lentes, interrompu puis repris, au gré des disponibilités et des priorités des contributeurs. En outre, le développement de l'encyclopédie dans son ensemble est sujet à des évolutions de ses affordances, de ses normes, et de la composition de sa communauté au cours du temps. Cette lente dérive ou *drift* de la plateforme (Salganik, 2018) est susceptible de modifier le processus de production des articles et les structures hypertextuelles résultantes. Si ces changements n'affectent guère des analyses portant – comme dans cet article – sur de courts intervalles temporels, leur impact peut être considérable sur des études à l'échelle de quelques années.

Enfin, il convient de rappeler que Historic Graphs – ainsi que les logiciels d'analyse des graphes qu'il produit, tels que Gephi – n'est qu'un outil s'intégrant dans une démarche scientifique complète. La nature exploratoire de la méthode présentée dans cet article ne saurait signifier que des résultats originaux émergent spontanément de la création de cartes dynamiques. Au contraire, ce sont les cartes qui naissent des hypothèses et des postulats des chercheurs. Chaque paramètre du processus de collecte de données et de construction de la carte est choisi, parfois directement, parfois après de nombreux essais, en fonction du résultat attendu – ou au moins des résultats considérés comme recevables. De même, le processus d'interprétation de la carte est fortement contraint par la question de recherche et le cadre théorique préalablement choisi. Si l'étude diachronique de Wikipédia constitue un riche sujet d'étude pour le champ des Humanités Numériques, seuls certains types de questions sont susceptibles de tirer bénéfice de la méthode d'analyse des cartes dynamiques des liens entre articles décrite ici. Et la validité des réponses qui pourront y être apportées ne dépendra pas tant de la qualité des outils utilisés que de la rigueur de la démarche dans laquelle ils s'intègrent.

7. Conclusions

Cet article visait à analyser le processus de production des articles de l'encyclopédie Wikipédia au regard de l'évolution de leurs liens hypertextes au cours du temps. Nous avons proposé une méthode exploitant les historiques des versions des articles pour dresser des cartes dynamiques présentant la structure des liens qui les unissent en différents points temporels. La description de quelques cas d'études a montré que l'usage de telle cartes nécessite un processus itératif, par essai-erreur, afin de produire des cartes susceptibles d'éclairer la problématique choisie. Ainsi, la construction et l'interprétation des cartes dynamiques doivent-elles s'insérer dans une démarche de recherche complète et cohérente.

L'outil de collecte de données Historic Graphs présenté dans cet article est un tout nouveau dispositif, dont le fonctionnement est encore susceptible d'évoluer, et dont les usages pour la recherche sont toujours à explorer. Il est par conséquent mis à disposition selon les termes de la licence libre GNU GPL, afin de favoriser sa circulation, sa réutilisation et son amélioration. Le script, sa documentation et des exemples d'utilisation sont disponibles sur https://framagit.org/retrodev/historic-graphs. Vous y trouverez également les données exploitées pour les cas d'étude décrits plus haut, ainsi que des versions plus complètes et plus haute résolution des cartes accompagnant cet article.

Remerciements

L'idée de cet article est issue de discussions avec Aymeric Bouchereau (ELLIADD) dans le cadre de travaux non publiés à ce jour. L'auteur tient à le remercier pour les échanges riches et bienveillants tout au long des dernières années.

12. Bibliographie

BASTIAN, Mathieu, Sebastien HEYMANN, Mathieu JACOMY et OTHERS. « Gephi : an open source software for exploring and manipulating networks. », dans *ICWSM 2009 : Proceedings of the Third International Conference on Weblogs and Social Media*, 8, 2009, p. 361–362.

BEAUDE, Boris. « De quoi Wikipédia est-elle le lieu? », dans BARBE, Lionel, Louise MERZEAU et Valérie SCHAFER (dir.), *Wikipedia, objet scientifique non identifié*, Presses Universitaires de Paris Ouest (PUPO), 2015.

BLANCHARD, Gilles et Mark OLSEN. « Le Système de renvois dans l'Encyclopédie: Une Cartographie des structures de connaissances au XVIIIe siècle », *Recherches sur Diderot et sur l'Encyclopédie*, n°31-32, 2002, p. 45.

BORRA, Erik et al. « Societal controversies in Wikipedia articles », dans *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, ACM, 2015, p. 193–196.

CARDON, Dominique et Julien LEVREL. « La vigilance participative. Une interprétation de la gouvernance de Wikipédia », *Réseaux*, n°154, 2009, p. 51–89.

DENSITYDESIGN et MÉDIALAB SCIENCE PO. *Seealsology*. En ligne au : <http://tools.medialab.sciences-po.fr/seealsology/>, consulté le 24 avril 2019.

GORGEON, Arnaud et E. Burton SWANSON. « Organizing the Vision for Web 2.0: A Study of the Evolution of the Concept in Wikipedia », dans *Proceedings of the 5th international Symposium on Wikis and Open Collaboration*, ACM, 2009.

JACOMY, Mathieu, Tommaso VENTURINI, Sebastien HEYMANN et Mathieu BASTIAN. « ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software », *PLOS ONE*, vol. 9, n°6, 2014.

LANGLAIS, Pierre-Carl. « {{Référence nécessaire}} L'émergence d'une norme wikipedienne », dans BARBE, Lionel, Louise MERZEAU et Valérie SCHAFER (dir.), *Wikipedia, objet scientifique non identifié*, Presses Universitaires de Paris Ouest (PUPO), 2015.

LATOUR, Bruno et OTHERS. *Reassembling the social: An introduction to actor-network-theory*, Oxford university press, 2005.

MERCKLÉ, Pierre. *Sociologie des réseaux sociaux*, La découverte, 2011.

PALOQUE-BERGES, Camille. « Les sources nativement numériques pour les sciences humaines et sociales », *Histoire@Politique*, vol. 30, n°3, 2016, p. 221-44.

ROGERS, Richard. *The End of the Virtual : Digital Methods*, Amsterdam, Amsterdam University Press, 2009.

ROGERS, Richard, Emina SENDIJAREVIC et OTHERS. « Neutral or National Point of View? A Comparison of Srebrenica articles across Wikipedia's language versions », *Proc. Wikipedia Academy*, 2012.

SAHUT, Gilles. « L'autorité importée dans Wikipédia : la question de la qualité des sources citées », *Quaderni*, n°91, 2016, p. 93-105.

SALGANIK, Matthew. *Bit by Bit, Social reseach in the digital age*, Princeton University Press, 2018.

SUROWIECKI, James et Elen RIOT. *La sagesse des foules*, JC Lattès, 2008.

VIÉGAS, Fernanda B., Martin WATTENBERG et Kushal DAVE. « Studying cooperation and conflict between authors with history flow visualizations », dans *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 2004, p. 575–582.

WELTEVREDE, Esther et Erik BORRA. « Platform affordances and data practices: The value of dispute on Wikipedia », *Big Data & Society*, vol. 3, n°1, 2016.

« Wikipedia:Manual of Style/Linking », dans *Wikipédia*. En ligne au : <https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Linking>, consulté le 24 avril 2019a.

« Wikipédia:Wikipédia est une encyclopédie », dans *Wikipédia*. En ligne au : <https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Wikip%C3%A9dia_est_une_encyclop%C3%A9die>, consulté le 24 avril 2019b.