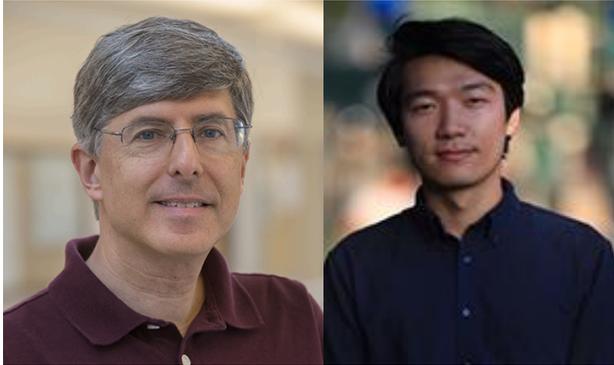


[COVID Information Commons \(CIC\) Research Lightning Talk](#)



Transcript of a Presentation by Steven Skiena and Xingzhi Guo (CUNY Stonybrook), February 2022

Title: Knowledge Graph intégrant l'évolution du COVID-19

Funded by NSF Office of Advanced Cyberinfrastructure, Directorate for Computer & Information Science & Engineering (OAC/CISE) through the Northeast Big Data Innovation Hub Seed Fund Program.

[Youtube Recording with Slides](#)

[February 2022 CIC Webinar Information](#)

Transcript Editor: Saanya Subasinghe

Transcript

Slide 1

Merci de nous avoir invité. Je m'appelle Xingzhi. Je suis étudiant en quatrième année de doctorat et je travaille avec le professeur Steven Skiena à l'université de Stony Brook. Le professeur Skiena est également le directeur de l'Institut d'IA de Stony Brook. Cette présentation porte sur l'évolution de l'intégration des graphes de connaissances au cours de la conférence COVID, à mesure que notre monde évolue. Le contenu principal est basé sur un article récent publié à SIGKDD 2021, en collaboration avec Boajian [Zhou] lorsqu'il était postdoc ici.

Slide 2

D'accord, passons à la suite. Commençons par la première chose. Qu'est-ce que l'intégration de graphe ou, pour d'autres, l'intégration de réseau ou l'intégration de nœuds ? Fondamentalement, il s'agit d'une fonction qui associe un nœud du graphique à un vecteur numérique que nous appelons vecteur d'intégration. Ainsi, comme le montre l'exemple suivant, nous pourrions faire correspondre chaque nœud de ce graphique à cet espace 2D. L'intégration peut donc capturer la signification du nœud dans le graphique original, mais le fait que les nœuds les plus proches dans le graphique peuvent avoir des vecteurs similaires dans l'espace d'intégration. Ainsi, dans cet exemple, vous pouvez voir que les nœuds de même couleur sont également proches les uns des autres sur ce plan 2D. Et surtout, grâce à ce vecteur d'intégration numérique de faible dimension, nous pouvons appliquer les algorithmes

d'apprentissage automatique existants à de nombreuses tâches en aval. Par exemple, la classification des nœuds, le regroupement des nœuds pour la découverte de communautés ou la détection des valeurs aberrantes. Cependant, cet exemple est un graphe statique où il n'y a pas de nouvelle arête, pas de nouveau nœud, tout est fixe, mais dans la vie réelle, notre monde est en constante évolution. Il en va de même pour les graphes et les nœuds du monde réel.

Slide 3

Jetons un coup d'œil au graphe changeant du monde réel - ou nous pouvons l'appeler un graphe dynamique. Dans le graphe de connaissances des liens de Wikipédia, chaque nœud est un article Wiki décrivant généralement des entités du monde réel et chaque arête est un hyperlien reliant deux articles, un peu comme la citation que nous faisons lorsque nous rédigeons un article. Ce graphe est donc à grande échelle. Nous avons des millions de nœuds et des centaines de millions d'arêtes et il ne cesse de s'agrandir comme vous pouvez le voir sur cette petite figure. Il est certain que certaines entités peuvent changer de nombreuses fois et qu'elles ont été capturées par ce graphe de connaissances dynamique. J'aimerais vous montrer un exemple précis, celui de la ville de Wuhan.

Slide 4

Avant COVID, la ville de Wuhan est probablement moins connue des gens du monde entier. À la fin de 2019 et au début de 2020, je pense que la plupart des gens savaient qu'il s'agissait du premier lieu de l'épidémie de COVID. Je pense qu'il s'agit là d'un bon exemple de changement d'entité. Ici, la figure montre un article de Wikipédia sur Wuhan. Je mets en évidence les liens hypertextes qu'il contient. Les deux premiers paragraphes sont des événements géo[logiques] ou historiques liés à Wuhan. Mais en 2019, nous constatons soudainement que de nombreux nouveaux liens liés à COVID ont été créés. Pour en revenir à la perspective de l'intégration des graphes, la question est de savoir comment suivre efficacement ces intégrations de nœuds dans ce graphe massif dynamique afin de pouvoir détecter le mouvement d'intégration du nœud et de le comparer dans le temps pour voir comment il évolue.

Slide 5

D'accord, ensuite, cette question nous motive à concevoir un nouvel algorithme qui peut traiter un problème que nous appelons un sous-ensemble d'intégration de nœuds dans les grands graphes dynamiques. Cet algorithme nous permet de suivre l'intégration de plusieurs nœuds prédéfinis. C'est donc un sous-ensemble de nœuds au lieu des quatre nœuds du graphe qui nous intéresse à mesure que le graphe évolue. L'idée principale est donc d'utiliser le page rank personnalisé, un algorithme très performant utilisé par le moteur de recherche Google. L'un des avantages par rapport à d'autres méthodes est que nous pouvons calculer uniquement ce dont nous avons besoin pour le sous-ensemble de nœuds, alors que la plupart des autres algorithmes doivent calculer tous les embeddings de chaque nœud à chaque fois. Mais ils n'en utilisent finalement qu'une partie, de sorte que le reste est simplement gaspillé. Notre algorithme est donc plus efficace et plus rapide et convient parfaitement à ce problème. Pour plus de détails, veuillez vous référer à notre article, qui se trouve ici [<https://arxiv.org/abs/2106.01570>]. Comme l'illustre cet exemple, notre méthode peut calculer

l'intégration d'un nœud spécifique, en l'occurrence Wuhan, sur plusieurs années. Nous pouvons nous attendre à un énorme mouvement d'intégration entre 2019 et 2020.

Slide 6

D'accord, ensuite - voyons les résultats de l'expérience. Tout d'abord, nous recueillons l'instantané du graphique de Wikipédia en anglais chaque jour en 2020. Comme je l'ai déjà mentionné, il s'agit d'un très grand graphique dynamique. Vous pouvez voir dans ce tableau qu'environ 30 000 nouvelles arêtes sont insérées chaque jour. Nous avons publié les données dans ce dépôt GitHub afin qu'elles soient très faciles d'accès. Nous suivons l'évolution de Wuhan et de certaines d'autres villes chinoises. Cette figure montre les changements que nous avons détectés dans l'espace d'intégration. Comme vous pouvez le voir, la courbe de Wuhan est proéminente. Elle présente un pic important et nous avons mis en évidence plusieurs pics avec des annotations et nous avons constaté qu'ils sont tous corrélés avec la chronologie de COVID. Et il y a un autre pic - vous pouvez voir que nous avons repéré un autre pic dans la ville de Chengdu qui reflète la tension diplomatique entre les États-Unis et la Chine lorsque les États-Unis ont décidé de fermer le consulat à Chengdu. Il s'agit donc d'un heureux hasard que nous avons découvert dans le cadre de cette étude.

Slide 7

Nous menons une autre expérience pour déterminer quelles sont les villes qui ont le plus changé au cours de différentes périodes. Nous classons donc les mouvements d'intégration du nœud suivi à chaque fois et nous constatons que, comme vous pouvez le voir, Wuhan est généralement la ville qui a le plus changé au fur et à mesure de l'évolution de COVID. Nous mettons en évidence le titre de l'actualité au cours de cette période afin d'avoir une idée de ce qui s'est passé à cet endroit.

Slide 8

Les principales conclusions sont les suivantes : nous disposons d'un algorithme d'intégration de nœuds très efficace qui peut capturer les intégrations dans de très grands graphes dynamiques. Nous avons ensuite étudié l'évolution du graphe de connaissances de Wikipédia en 2020 et découvert les changements de nœuds intéressants pendant la COVID. Pour plus de détails, veuillez vous référer à notre article et consulter les ressources publiées. [<https://arxiv.org/abs/2106.01570> et <https://github.com/zjlxgz/DynamicPPE>] Et oui, merci, merci !