Graphes multicouches pour mesurer la similarité entre des films

Majda Lafhel, Lylia Abrouk, Hocine Cherifi, Mohammed El Hassouni

▶ **To cite this version:**

Majda Lafhel, Lylia Abrouk, Hocine Cherifi, Mohammed El Hassouni. Graphes multicouches pour mesurer la similarité entre des films. French Regional Conference on Complex Systems, Jun 2022, Paris, France. hal-03671762

HAL Id: hal-03671762
https://hal.science/hal-03671762

Submitted on 18 May 2022

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.

Graphes multicouches pour mesurer la similarité entre des films

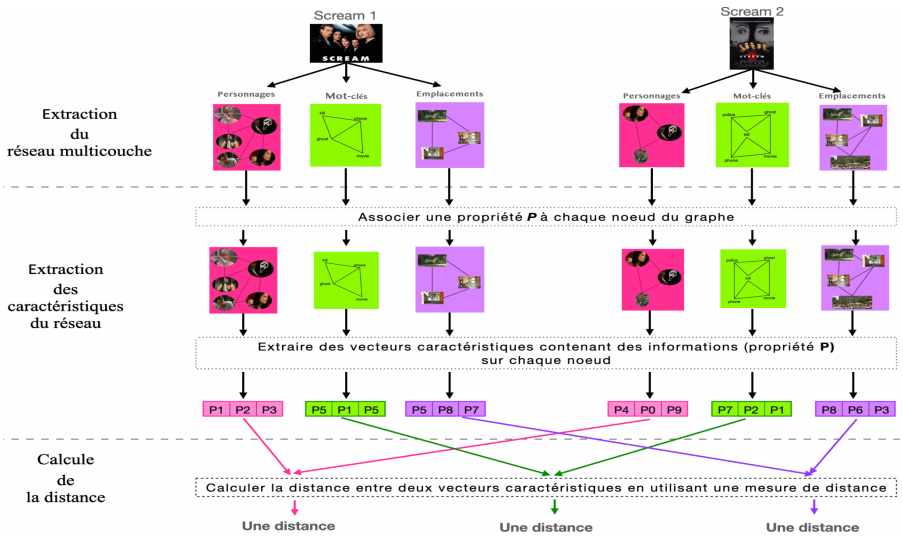
Majda Lafhel[1], Lylia Abrouk[2], Hocine Cherifi[2], Mohammed El Hassouni [1]. (1) LRIT URAC 29, Mohammed V University, Rabat, Morocco, majdalafhel1@gmail.com (2) LIB, University of Burgundy, France.

Un système de recommandation des films propose à ses utilisateurs un ensemble de films similaires à leurs séries préférées, en se basant sur le principe de filtration, par exemple la filtration de leurs données personnelles. Ces dernières années, les réseaux sociaux sont devenus de plus en plus utilisés pour représenter et analyser des histoires de films[1][2][3]. Une étude [6] s'est appuyée sur les réseaux sociaux pour classifier les films par genre. D'autres études [4][5] ont utilisé les réseaux sociaux pour analyser des films à partir des rôles des personnages. A notre connaissance, il n'existe pas une technique qui permet de trouver la similarité entre deux films en calculant la distance entre leurs réseaux. Les mesures de distances sont des approches qui permettent d'associer une propriété à un réseau afin de capturer ses informations structurelles, et calculer la distance entre eux.

Dans ce travail, notre objectif est d'investiguer la performance des mesures de distances à calculer la similarité entre des films, et l'incorporer aux systèmes de recommandations. Les réseaux modèles existants sont basés uniquement sur les personnages. Ces derniers ne représentent qu'une facette d'une histoire, ce qui les rend insuffisantes pour analyser un film. Récemment, un modèle de réseau multicouche a été introduit [7] pour présenter une histoire de film en prenant en considération trois aspects : les personnages, les mots-clés et les lieux. Le réseau multicouche se constitue de trois couches(personnages, mot-clés, lieux) et les interactions entre eux. Nous avons utilisé ce modèle pour comparer la similarité entre des films en tenant compte de plusieurs éléments de l'histoire. Dans ce travail, nous analysons les résultats de plusieurs mesures de distances sur différentes catégories de films. Cette analyse concerne les différentes couches définies dans le modèle multicouche. Nous présentons dans ce qui suit notre approche ainsi que les premiers résultats.

Approche Notre méthodologie est composée de trois étapes principales :

1. Extraction du réseau multicouche pour chaque film;
2. Extraction des caractéristiques du réseau (features);
3. Calcul de la distance.



La première étape consiste à extraire pour chaque film un réseau multicouche constitué de trois couches correspondant chacune à une entité (personnages, mots-clés, ou emplacements). Le réseau multilayer est composé de relations intra-couches (liant les nœuds de la même entité) et inter-couches (liant les nœuds de différentes entités). Au cours de cette étape, nous avons ignoré les relations inter-couches afin de comparer des couches de la même entité.

La deuxième étape consiste à extraire les caractéristiques de chaque réseau. Les caractéristiques jouent le rôle principal dans l'analyse d'un réseau puisqu'elles nous donnent des informations sur sa structure et sa topologie. Pour ceci, on associe une propriété à tous les nœuds du graphe. Puis, on extrait un vecteur caractéristique contenant les propriétés de chaque nœud.

La troisième étape vise à déterminer la différence entre les informations structurelles de deux graphes donnés ; en d'autres termes, cette étape vise à calculer la différence entre les vecteurs caractéristiques, ce qui permet d'obtenir une mesure de distance.

Résultats. Nous avons implémenté 5 mesures de distances : « Laplacian Spectra », « Network Portrait Divergence », « NetLSD », « NetMF » et «la distance euclidienne entre les spectres » sur les couches de personnages, mots-clés et lieux extraits de films d'horreur, de films romantiques, de films comiques, et de films d'aventures. Puis nous avons comparé les couches de personnages des films entre eux, les couches des mots-clés entre eux, et les couches de lieux entre eux.
Dans un premier temps, nous présenterons les résultats obtenus de la comparaison des couches de personnages. La table ci-dessous récapitule la performance des mesures de distance à comparer les films par catégories, selon l'approche utilisée. Il est clair que l'approche « Network Portrait Divergence » a réussi à trouver la similarité entre les personnages pour les quatre catégories.

Measures	Horror	Romance	Adventure	Comedy
NetLSD	✓	✗	✗	✓
NetMF	✓	✗	✗	✗
D-measure	✗	✗	✗	✓
Portrait Divergence	✓	✓	✓	✓
Laplacien Spectra	✓	✗	✗	✓

Table 6: Character checklist table

Références

[1] Park, S. B., Oh, K. J., & Jo, G. S. (2012). Social network analysis in a movie using character-net. *Multimedia Tools and Applications*, *59*(2), 601-627.

[2]Sang, J., & Xu, C. (2010, October). Character-based movie summarization. In *Proceedings of the 18th ACM international conference on Multimedia* (pp. 855-858).

[3]Sang, J., Liang, C., Xu, C., & Cheng, J. (2011, July). Robust movie character identification and the sensitivity analysis. In *2011 IEEE International Conference on Multimedia and Expo* (pp. 1-6). IEEE.

[4]Weng, C. Y., Chu, W. T., & Wu, J. L. (2009). Rolenet: Movie analysis from the perspective of social networks. *IEEE Transactions on Multimedia*, *11*(2), 256-271.

[5]Jung, J. J., You, E., & Park, S. B. (2013). Emotion-based character clustering for managing story-based contents: a cinemetric analysis. *Multimedia tools and applications*, *65*(1), 29-45.

[6]Adams, B., Dorai, C., & Venkatesh, S. (2002). Toward automatic extraction of expressive elements from motion pictures: Tempo. *IEEE Transactions on Multimedia*, *4*(4), 472-481.

[7]Mourchid, Y., Renoust, B., Cherifi, H., & El Hassouni, M. (2018, December). Multilayer network model of movie script. In *International Conference on Complex Networks and their Applications* (pp. 782-796). Springer, Cham.

[8]Messadi, M., Cherifi, H., & Bessaid, A. (2021). Segmentation and ABCD rule extraction for skin tumors Classification. In arXiv preprint arXiv:2106.04372, 2021