



HAL
open science

Méthodes d'extraction de réseaux sociaux et de diffusion de photos sociales

Michel Crampes, Michel Plantié

► **To cite this version:**

Michel Crampes, Michel Plantié. Méthodes d'extraction de réseaux sociaux et de diffusion de photos sociales. IC 2011, 22èmes Journées francophones d'Ingénierie des Connaissances, May 2012, Chambéry, France. pp.739-754. hal-00746735

HAL Id: hal-00746735

<https://hal.science/hal-00746735v1>

Submitted on 29 Oct 2012

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.

Méthodes d'extraction de réseaux sociaux et de diffusion de photos sociales

Michel Crampes, Michel Plantié

LGI2P Research Center, Alès School of Mines,
Site EERIE, Parc scientifique G. Besse, F 30 035 Nîmes, France
Michel.Crampes@mines-ales.fr
Michel.Plantie@mines-ales.fr

Résumé : Les photos sociales qui représentent des individus et qui ont été prises lors d'événements sociaux portent la trace des relations sociales entre les participants à ces événements. Nous proposons dans cet article plusieurs méthodes d'extraction de réseaux sociaux de proximité à partir de collections de photos sociales. Elles sont illustrées sur le cas d'un mariage puis comparées après réduction et visualisation des graphes extraits. Nous montrons ensuite comment une collection de photos sociales peut être décrite en tant qu'hypergraphe. Dans le but de procéder à une diffusion personnalisée des photos, nous proposons de transformer les graphes des réseaux en hypergraphes. Plusieurs méthodes sont possibles et nous détaillons une méthode simple. Il est ensuite possible de définir une politique de diffusion des photos à partir d'une combinaison de l'hypergraphe des personnes et de l'hypergraphe des photos.

Mots-clés : Photos sociales, analyse de Réseaux Sociaux, apprentissage de distances et similarités, apprentissage de graphes, visualisation

1. Introduction

1.1. Notion de photo sociale

Nous qualifions de « sociales » les photos qui montrent la présence de personnes isolées ou de groupes de personnes lors d'événements sociaux tels que des regroupements familiaux ou des soirées entre amis. Elles prennent ce qualificatif social du fait qu'elles sont associées à un événement social. Avec l'usage très répandu d'appareils photos numériques, ce genre de photos joue de plus en plus un rôle central de témoignage des échanges pendant les événements sociaux. De plus le partage de photos numériques, et en particulier de photos sociales, est au cœur du succès d'applications telles que FaceBook qui sont devenus ensuite des modèles de constitution de réseaux sociaux. Il y a de toute évidence un rapport particulier entre les

photos sociales, les réseaux sociaux et la diffusion de photos sociales sur les réseaux sociaux. Par ailleurs (Jain, 2010) considère que le problème de la diffusion personnalisée de photos est un problème ouvert auquel la recherche a porté peu d'intérêt jusqu'à présent.

Implicitement, une collection de photos sociales est représentative dans une certaine mesure de l'existence de relations sociales privilégiées puisque les photos montrent des personnes qui se sont rencontrées, ou tout au moins qui se sont croisées.

1.2. Hypergraphes et photos

Nous montrons qu'une collection photos peut être représentée formellement comme un hypergraphe dont les sommets sont les personnes et les arêtes les photos ou, de manière duale, un hypergraphe dont les individus sont les arêtes et les photos les sommets. Par analogie nous posons alors qu'un réseau social, qui est en général représenté par un graphe, peut aussi être vu formellement comme un hypergraphe dans lequel les individus sont à la fois des nœuds et des arêtes que nous qualifions de tribus. Nous proposons plusieurs méthodes d'extraction de cet hypergraphe social à partir de l'hypergraphe des photos. Nous extrayons d'abord différents graphes et appliquons des algorithmes de réduction et de visualisation pour distinguer la nature des relations et des réseaux extraits. Puis nous introduisons plusieurs politiques de construction d'hypergraphes sociaux. Une d'entre elles est détaillée avec sa politique de diffusion de photos basée sur la dualité de l'hypergraphe des photos et de l'hypergraphe des personnes. Pour comparer les résultats des différentes méthodes nous considérons une collection particulière de photos sociales prises lors d'une cérémonie de mariage. Notre objectif est de faire apparaître des groupes sociaux pour faciliter la diffusion de photos et organiser d'autres événements sociaux en fonction de ces groupes en les impliquant de différentes manières, et de donner aux gens une image d'un événement à partir duquel ils peuvent s'identifier et éventuellement partager leurs photos.

1.3. Contribution, motivation et Structuration de l'article

Notre contribution s'inscrit à l'intérieur de limites qui pourront être levées en tout ou en partie dans des recherches ultérieures. Nous nous limitons ici aux réseaux que nous qualifions de proximité, c'est-à-dire des réseaux d'individus qui ont l'occasion de se rencontrer physiquement lors d'événements. Ces réseaux sont de taille réduite, incluant rarement plus de 200 personnes (cas d'un mariage déjà assez important). De même les observations que nous avons pu faire montrent que le nombre de photos prises lors d'événements sociaux et qui peuvent faire l'objet d'une diffusion étant donné leur qualité dépassent rarement le nombre de 300 photos quand on exclut les photos tirées en rafale avec les mêmes personnes comme par exemple le couple des mariés.

Nous considérons des photos déjà indexées avec des noms ou des identifiants de personnes. En conséquence, nous n'utilisons pas d'autres informations comme l'âge, le sexe, ou les informations civiles ; nous faisons appel à ces dernières comme données externes de référence pour évaluer les résultats des réseaux construits. En mettant en œuvre une organisation de type sous hiérarchie de Galois, nous ne considérons que les individus sur les photos et ne faisons pas intervenir de critères techniques ou esthétiques sur la qualité des photos à diffuser. Ces aspects ne sont cependant pas totalement écartés dans la mesure où un choix subjectif peut être effectué a posteriori à partir de photos candidates à la diffusion.

L'article s'articule de la manière suivante. Après l'état de l'art de la section 2, l'article aborde les différentes méthodes mises en œuvre pour construire un graphe de relations entre individus dans la section 3. Une expérimentation est menée dans la section 4 qui nous permet de visualiser et comparer des résultats de réseaux. La section 5 propose des méthodes de construction d'hypergraphes de réseau social et une politique de diffusion de photos par tribus en utilisant un formalisme matriciel. La section 6 discute et conclut l'article.

2. Etat de l'art

Paradoxalement, le rapport entre photos sociales et réseaux sociaux est encore peu étudié. Les réseaux sociaux en ligne se construisent actuellement à partir des utilisateurs qui acceptent des "amis" au sens de Facebook ou de MySpace. A noter cependant que le service peut recommander certains "amis". La diffusion des photos est elle aussi laissée à l'initiative des utilisateurs avec quelques exceptions comme l'accès automatique à des photos taguées avec des noms d'amis dans FaceBook. L'analyse de photos sociales n'est devenue que récemment un sujet de recherche important. Les travaux portent essentiellement sur l'indexation, l'organisation et la fouille de collections photos. (Rooden & Wood, 2003) analyse les attentes d'utilisateurs pour montrer la nécessité d'aides techniques en la matière. La littérature fait souvent appel à des interfaces riches comme dans (Ames M. & Manguy 2006) ou à des techniques de reconnaissance d'images. L'utilisation des méthodes issues de l'Analyse Formelle de Concepts est plus rare (Eklund et al. 2006) (Ferré S. 2007) (Crampes et al. 2009).

Les réseaux sociaux par ailleurs constituent depuis longtemps un domaine de recherche important et l'intérêt est croissant. La forme essentielle en est le graphe et la recherche porte surtout sur l'analyse topologique (Newman, 2006) avec quelques incursions dans la visualisation pour révéler les structures d'intérêt (Freeman, 2000). L'extraction de réseaux sociaux devient un sujet important de recherche avec des contributions notables comme celles de (Mika 2005) qui croise individus, documents et concepts tandis que (Jung & Euzenat 2007) bâtissent des réseaux sociaux sémantiques basés sur des ontologies. Dans ces travaux le

réseau social reste vu comme un graphe ou bien une partition d'un ensemble. Des travaux plus récents généralisent les graphes aux hypergraphes pour représenter les réseaux et en particulier les réseaux sociaux dans le but de retrouver certaines propriétés topologiques (Estrada & Rodriguez-Velazquez, 2006). D'autres formes de représentation au delà du graphe sont explorées dont en particulier les treillis de Galois (Freeman & White 1993) (Roth & Bourguin 2006).

L'extraction de réseaux sociaux à partir de photos sociales est avancée par (Golder 2008). (Wu & Tretter 2009) combinent l'observation d'individus sur des photos sociales avec une analyse d'image limitée pour recomposer des groupes de référence. Dans (Planté & Crampes 2010) nous proposons un passage des photos sociales aux réseaux sociaux jusqu'à la diffusion. Nous approfondissons ici cette démarche en matière de formalisation, de visualisation, d'extraction d'hypergraphes de réseaux sociaux et de politique de diffusion basée sur un calcul matriciel.

3. Définitions et méthodes

3.1. Construction d'hypergraphes duaux

La première étape pour la construction de réseaux sociaux à partir de photos sociales consiste à s'abstraire de tout critère esthétique ou de technique photographique. A cette fin les photos préalablement indexées avec les identifiants des individus qui y figurent sont rassemblées par similarité de groupes d'individus. De manière simplifiée, étant donné l'ensemble $\{P, X\}$ constitué respectivement de l'ensemble P des photos de la collection et de l'ensemble X des individus présents sur les photos, nous considérons chaque photo $p_i \in P$ comme un objet et chaque individu $x_i \in X$ comme un attribut booléen associé à chaque photo. L'ensemble des photos et des individus (objets et attributs) constitue un "contexte" au sens de l'Analyse Formelle de Concepts (Ganter & Wille, 1999). Un groupe maximal de photos qui contient un sous-groupe d'individus forme un "concept" c'est-à-dire un groupe d'objets pour lequel l'ajout ou le retrait d'un attribut changerait la constitution. L'ensemble des objets d'un concept forme son "extension" et l'ensemble des attributs du concept son "intension". Une relation d'ordre partiel est appliquée sur les concepts pour établir une hiérarchie. Selon la définition des hiérarchies de Galois, un objet peut figurer dans plusieurs concepts puisqu'il peut partager un même groupe d'attributs avec d'autres objets appartenant à d'autres concepts. En ne retenant des concepts que ceux dans lesquels les objets figurent une seule fois au plus bas de la hiérarchie, on obtient une Sous Hierarchie de Galois Objet (Crampes et al 2009).

La collection se réduit alors à un ensemble de concepts C , chaque concept $c_k \in C$ étant un groupe de photos dont l'extension est l'ensemble des photos du groupe et l'intension l'ensemble unique d'individus présents dans

ces photos. N'importe laquelle des photos d'un groupe peut représenter le concept. Nous choisissons arbitrairement une photo pour chaque groupe. Lors de la phase de diffusion des photos, le choix d'une ou de plusieurs photos dans chaque groupe sur des critères techniques photographiques ou esthétiques pourra être envisagé, mais cet aspect ne fait pas l'objet de la recherche présentée ici. Pour simplification nous appellerons un groupe de photos un "concept", la photo représentative d'un groupe la "photo prototype".

Ainsi réduite à ses concepts objets, la collection photos peut être vue comme un hypergraphe (Berges 1973). Un hypergraphe H est un couple (V, E) où $V = v_1, v_2, \dots, v_n$ est un ensemble non vide (généralement borné) et $E = E_1, E_2, \dots, E_m$ est une famille de parties non vides de V . Les éléments de V sont les sommets de H . Les éléments de E sont les arêtes de H . On peut observer qu'un graphe est un cas particulier d'hypergraphe de rang 2 (toutes les arêtes ne relient que deux sommets). Dans notre cas nous définissons l'hypergraphe H de la manière suivante : chaque personne est un sommet ($x_j \equiv v_i$) et chaque concept (ou de manière équivalente chaque photo prototype) peut être vu comme une arête qui contient des personnes ($c_j \equiv E_j$).

Un hypergraphe est aussi un graphe bipartite : toute arête est reliée à un ensemble de sommets et tout sommet peut appartenir à plusieurs arêtes. En conséquence, à partir d'un hypergraphe H , il est possible de construire son hypergraphe dual H^t dont la matrice d'adjacence est la transposée de la matrice d'adjacence de H . Dans notre cas, l'hypergraphe dual a pour sommets l'ensemble des photos et pour arêtes les individus. L'arête d'un individu contient toutes les photos dans lesquelles cet individu est présent.

Dans la stratégie de diffusion de photos que nous présentons, nous partons de l'hypergraphe H des photos pour construire un hypergraphe des relations entre personnes K qui représente les tribus. Une étape intermédiaire est nécessaire : elle consiste à rechercher la proximité des individus selon leur cooccurrence dans les concepts.

3.2. Méthodes d'analyse de proximité entre individus

Dans (Crampes & Plantié 2010) nous avons défini trois manières d'évaluer la force du lien social entre individus à partir de leur appartenance commune à une arête (à un groupe de photos similaires). Nous reprenons ici ces trois forces en leur donnant un caractère plus formel.

Soient les définitions suivantes :

$[.]$ est un ensemble (aussi symbolisé par une lettre majuscule pour un ensemble particulier : $[x]$ est l'ensemble X des individus x).

Car $(.)$ est la cardinalité d'un ensemble.

x_i est un individu particulier ($x_j \in X$)

c est un concept (ou une photo prototype) ($c \in C$) dont l'intension est un ensemble d'individus

$[c|[x_i, x_j]]$ est l'ensemble des concepts qui contiennent le couple $[x_i, x_j]$.

$[c|x_i \vee x_j]$ est l'ensemble des concepts qui contiennent x_i ou x_j .

$\text{Car}(c[[xi, xj]])$ est la cardinalité du concept c contenant le couple $[xi, xj]$.

3.2.1. La force simple d'un couple

La force la plus simple d'attraction d'un couple résulte de la fréquence d'observation du couple parmi tous les concepts, c'est-à-dire dans des situations où le couple est mêlé avec d'autres personnes selon différentes combinatoires, ce qui permet d'en conclure une certaine stabilité.

$$\text{ForceSimple}(xi, xj) = \frac{\text{Car}[c[[xi, xj]]]}{\text{Car}(C)}$$

Ce mode de calcul est fréquemment retenu dans la littérature pour l'extraction de réseaux sociaux sous le terme de cooccurrence (Roth & Bourguine, 2006). Nous montrons dans les deux autres méthodes présentées ci-dessous que la simple cooccurrence oublie des indicateurs essentiels des relations entre individus et conduit à des biais importants.

3.2.2. La proximité ou force pondérée d'un couple

Si on se limite à la force simple, on conclut que deux personnes présentes sur une unique photo prise lors d'une manifestation comprenant plusieurs milliers de personnes (par exemple un policier et un manifestant) sont aussi proches que deux mariés seuls sur une photo. Pour corriger ce biais nous considérons que plus il y a de personnes dans une photo, moins les liens sont forts entre ces personnes. (Golder 2008) et (Wu & Tretter 2009) introduisent ce paramètre pour calculer une distance (respectivement une proximité) entre individus sous la forme d'une racine carrée au numérateur (respectivement au dénominateur) sans toutefois en donner une justification. Nous donnons une expression différente avec une justification de ce choix.

Etant donné un ensemble d'individus X (l'intension d'un concept dans notre cas), la probabilité de l'existence d'un couple particulier $[xi, xj]$ est égale à l'inverse du nombre de couples possibles, soit :

$$\text{Prob.}(xi, xj) = 2/(\text{Car}(X) * (\text{Car}(X)-1)).$$

Donc, plus le couple est dilué dans un grand nombre d'individus pour un concept, moins la contribution de ce concept à la proximité est forte. Quand un couple n'apparaît dans aucun concept, la force est nulle. Deux personnes n'apparaissant qu'en couple donnent la proximité maximum égale à 1. Dans la mesure où par la suite on s'intéressera aux proximités relatives entre individus, il suffit de ne considérer qu'un seul facteur ($2/\text{Car}(X)$). En se limitant à un seul facteur, nous obtenons :

$$\text{Proximité}(xi, xj) = \frac{\sum_{[c[[xi, xj]]} \frac{2}{\text{Car}(c[[xi, xj]])}}{\text{Car}(C)}$$

L'utilisation d'une racine carrée comme pour les auteurs cités ci-dessus se justifie aussi de manière empirique, mais alors la proximité de deux

personnes seules sur une photo unique n'est pas égale à 1, ce qui nous paraît moins pertinent. De plus leur calcul de proximité porte sur toutes les photos et non sur les concepts comme nous le proposons.

3.2.3. Cohésion d'un couple

Les deux méthodes précédentes s'intéressent à la fréquence d'observation de couples. Par contre elles n'apportent aucune information sur le caractère soudé ou non d'un couple, c'est-à-dire sur le fait que les deux membres d'un couple sont presque toujours vus ensemble ou bien fréquemment séparés. Nous appelons "cohésion" cette observation. Le couple $[x_i, x_j]$ a une forte cohésion lorsque x_i et x_j sont souvent représentés dans les concepts par rapport au nombre de fois où ils apparaissent dans un concept.

$$\text{Cohésion}(x_i, x_j) = \frac{\text{Car}([c|x_i, x_j])}{\text{Car}([c|x_i \vee x_j])}$$

Cette force de liaison est riche d'enseignement comme nous verrons plus loin et nous ne connaissons pas d'autres auteurs qui en ont fait usage pour extraire des réseaux sociaux. Pourtant elle correspond formellement à l'indice de Jaccard. En effet si on considère l'hypergraphe dual H^t dans lequel les individus sont les arêtes et les photos les attributs, un individu x_i est représenté par l'ensemble des photos dans lesquelles il figure. Or, étant donné deux ensembles A et B, l'indice de Jaccard $J(A, B)$ est le rapport entre la cardinalité de l'intersection des ensembles considérés $\text{Car}(A \cap B)$ et la cardinalité de l'union des ensembles $\text{Car}(A \cup B)$ ce qui correspond bien à la définition de la cohésion que nous avons donné. La représentation de la collection sous forme d'hypergraphe nous permet d'exploiter des méthodes bien connues pour les calculs de similarité entre deux ensembles.

4. Expérimentation: fouille d'un réseau social particulier

Dans cette section, nous appliquons les trois forces sur un corpus de photos sociales et analysons les résultats afin d'obtenir des interprétations sémantiques. Les photos sociales sont la plupart du temps prises au cours d'événements familiaux, en particulier les mariages, divers événements familiaux ou soirées entre amis. Une cérémonie de mariage est intéressante car de nombreuses photos v sont prises et par ailleurs les relations civiles sont connues à l'avance. Ces relations civiles peuvent être utilisées en tant que référents sémantiques décrivant le réseau social. Dans cette section nous décrivons comment les forces peuvent sonder ces relations sociales et dévoiler d'autres relations sociales représentant une sémantique alternative.

4.1. Le corpus et le graphe associé

Notre corpus expérimental contient 144 photos qui ont toutes été prises au cours d'une cérémonie de mariage. Le corpus initial contenait davantage de photos mais une sélection préalable a été faite pour éviter la redondance des photos prises en cascade.

Du fait que la prochaine étape sera d'extraire des réseaux sociaux à partir de ces photos, nous avons besoin d'une référence pour valider les résultats attendus du point de vue topologique et sémantique. Pour éviter un biais possible nous n'avons pas construit nous même cette référence. Nous avons délégué cette tâche à deux étudiants de troisième cycle à qui l'on a demandé de dessiner un graphe des relations civiles du mariage de manière indépendante. Les deux graphes civils qu'ils ont livrés sont très similaires. Les nœuds sont évidemment les mêmes parce que les deux expérimentateurs ont utilisé les 28 mêmes personnes, les rôles ont été définis avec presque les mêmes mots, et surtout la plupart des liens faits à la main étaient semblables dans les deux graphiques. Nous avons conclu que l'un des deux graphes pourrait être utilisé comme un référent pour les graphes de relations sociales

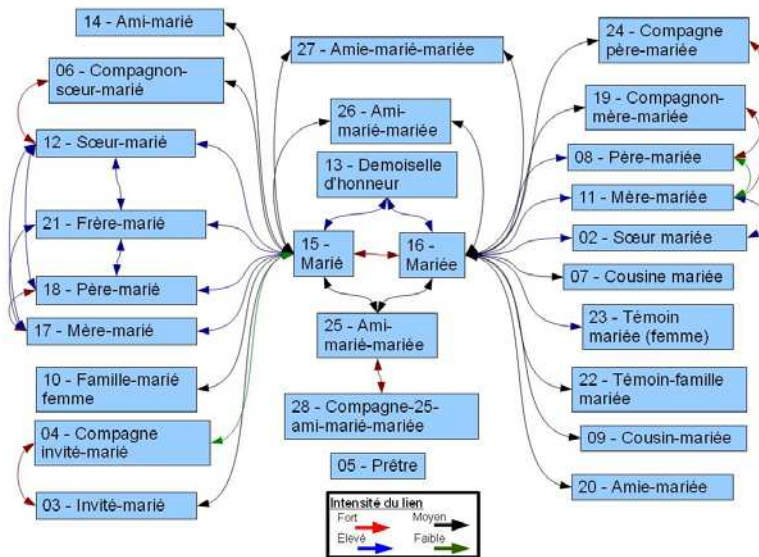


Fig -1 Le graphe civil de référence

à calculer dans les étapes suivantes. L'un d'eux a été choisi et appelé graphe civil. Ce graphe montré dans la Figure 1 donne une référence sémantique à la noce avec un rôle civil précis pour chaque personne.

4.2. Extraction de réseaux sociaux

L'analyse du corpus des photos du mariage finalement retenues a permis d'extraire 127 concepts représentés par 127 photos prototypes montrant des

groupes de personnes tous différents. Nous avons ensuite calculé la similarité (la force du lien social) pour tous les couples de personnes selon les différentes forces de lien social présentées dans la section précédente. Avec 28 personnes impliquées dans la cérémonie du mariage, l'application des liens sociaux donne trois matrices carrées 28 x 28, une pour chaque type de lien. Nous avons ensuite calculé un graphe pondéré pour chaque lien social. Les nœuds d'un graphe sont les personnes et les liaisons sont leurs distances sociales (ou dissimilarités comme précisé ci-dessous) calculées à partir des matrices avec les normalisations suivantes :

Normalisation. Elle est opérée sur chaque force. Le couple ayant la force de liaison la plus forte est pris pour référence avec une similarité égale à 1 (ou une distance égale à 0). Par exemple pour la force simple, on observe que les mariés ont la plus forte similarité égale à 0,50294. Nous la considérons comme une base égale à 1. Nous divisons alors la force simple de tous les couples par la valeur de la force du couple (ici 0,50294). Finalement nous appliquons un multiplicateur de 1000 pour obtenir une mesure entre 0 et 1000 avec des chiffres significatifs. En considérant la force simple, la dissimilarité (l'opposé de la similarité) pour un couple $[x_i, x_j]$ est donc :

$$1000 * (1 - \text{force-simple}([x_i, x_j]) / 0,50294).$$

Sémantiquement pour plus de clarté nous utiliserons le terme distance plutôt que le terme dissimilarité. D'un point de vue mathématique, c'est un abus de langage parce que les dissimilarités calculées ici ne sont pas des distances au sens mathématique. En effet, la loi de transitivité des distances mathématiques, n'est en général pas vérifiée dans notre cas. Une personne A peut ne pas être plus proche d'une personne C que la somme des proximités entre les personnes A et B les personnes B à C. Cependant, la loi de symétrie est vérifiée et implique qu'un graphe de réseau social dérivé à partir de photos est non orienté. Les trois graphes résultants montrent 146 liaisons pour la force simple, 138 liaisons pour la force de proximité et 149 liaisons pour la force de cohésion. Tous ces calculs ont été effectués par une équipe de trois étudiants à qui les définitions des forces avaient été fournies, et les résultats ont été vérifiés par les auteurs de cet article.

La Figure 2 montre une projection sur deux dimensions du graphe de la force de cohésion. Elle a été produite par notre logiciel de dessin graphique «Molage» avec lequel il est possible de définir de manière interactive différentes heuristiques d'affichage qui s'appliquent à des sommets de type différents. Dans ce cas précis il s'agit d'une projection MDS (MultiDimensional Scaling) basée sur la minimisation du stress avec une simulation d'un ensemble de ressorts (Eades, 1984). Lorsque l'on observe cette figure, elle paraît complexe et sa topologie difficile à interpréter en particulier au regard de la Figure 1. Une première possibilité pour l'analyse serait de mener une évaluation quantitative et utiliser des mesures comme les mesures de centralité pour extraire des connaissances sur l'importance de

chaque acteur. Toutefois, cette approche même si elle peut apporter quelques résultats intéressants, n'atteint pas notre objectif de donner aux gens une image de la cérémonie à partir de laquelle ils peuvent s'identifier et éventuellement partager leurs photos. Nous avons besoin d'un affichage qui ressemble un peu à la Figure 1 et en même temps donne d'autres traits sémantiques. La première chose à faire est évidemment de réduire le nombre d'arêtes et de tenter de réduire le nombre de croisement des arêtes restantes.

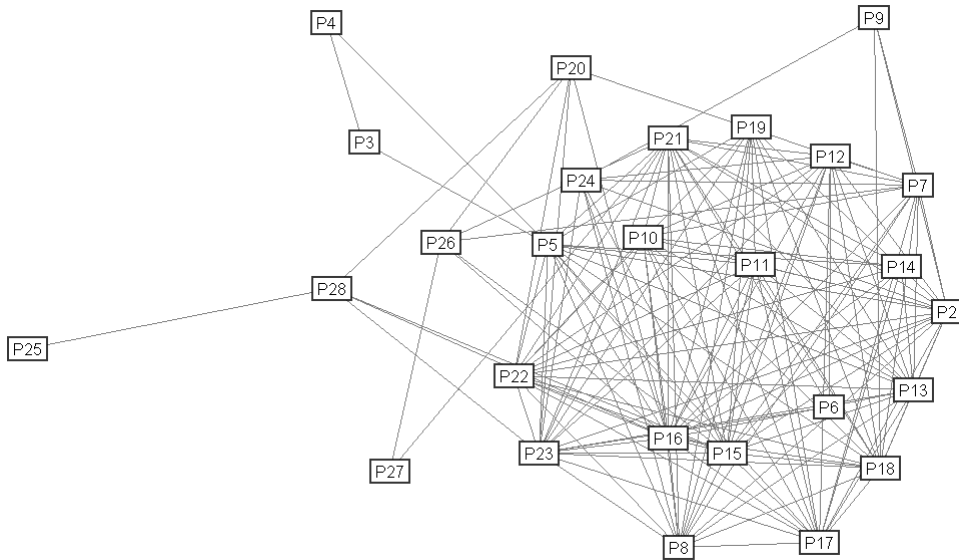


Fig. 2 -- Projection MDS de la force de cohésion avec 149 liaisons

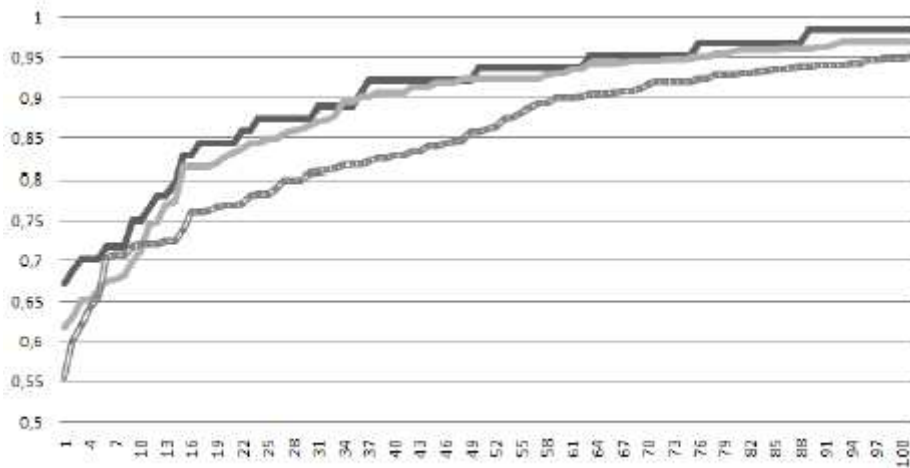
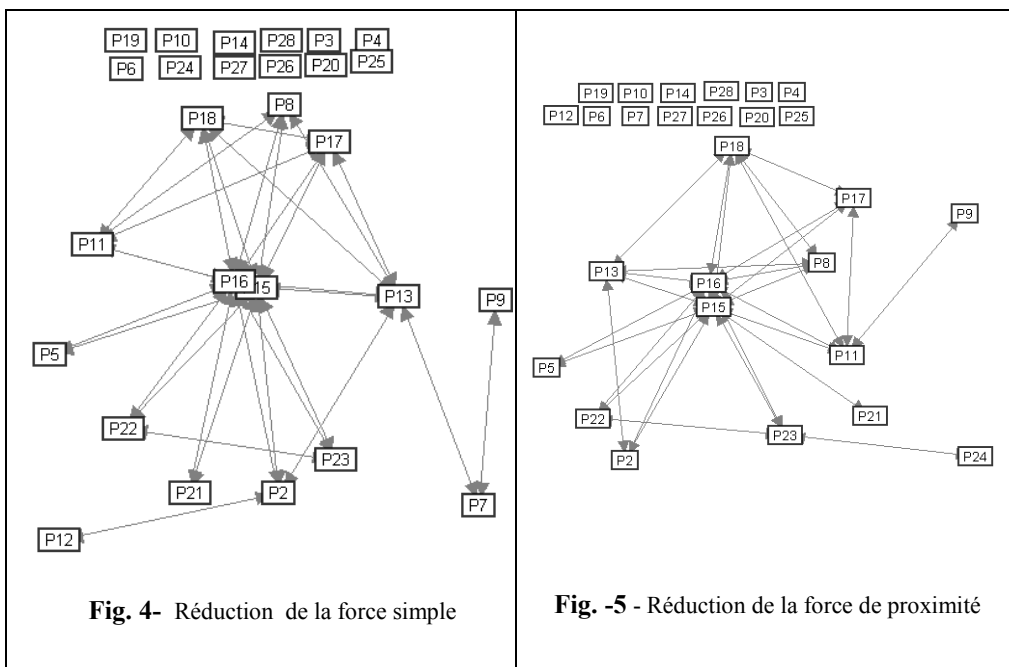


Fig. 3 -- Comparaison du nombre d'arêtes en fonction du seuil

4.3. Réductions du graphe de réseau social



Un meilleur affichage peut être obtenu si nous ne conservons que les

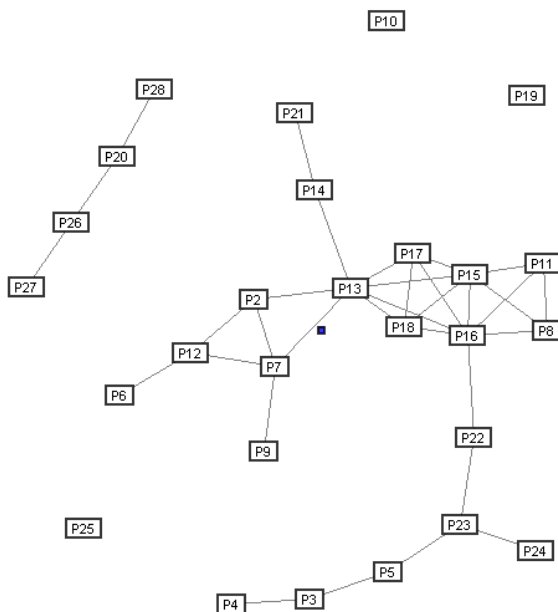


Fig. 6 -- Graphe réduit de la force de cohésion (33 arêtes)

arêtes qui représentent la sémantique du réseau dans sa meilleure configuration. La plupart des approches évidentes consistent à garder les plus petites arêtes dans leur distance avant projection car elles représentent les liens sociaux les plus forts. Le nombre d'arêtes à conserver doit respecter les contraintes qui suivent : pas trop d'arêtes pour éviter la complexité visuelle, un nombre d'arêtes suffisant pour obtenir une visualisation significative, et le même nombre d'arêtes entre les trois graphes réduits correspondant aux trois liens sociaux pour être en mesure de comparer leur sémantique.

Pour trouver cette valeur, nous pouvons comparer les trois graphiques en fonction de leur nombre d'arêtes par rapport à la valeur du seuil de distance indiqué dans la Figure 3. Dans cette figure pour chacune des trois forces le nombre d'arêtes acceptées est donné sur l'axe des x en fonction de la valeur de la distance normalisée à 1 sur l'axe des y. Lorsque l'on regarde cette figure, la force de proximité et la force simple se comportent pratiquement de la même façon comparativement au nombre d'arêtes. Rares sont les couples dont la distance est inférieure à 900. Le lien de cohésion est beaucoup plus plat et il indique plus de couples que les deux autres liens à de plus petites distances.

Synthétiquement, on voit que pour la cohésion il y a 149 liaisons pour un seuil de 1000, 65 liaisons pour un seuil de 900, et 33 liaisons pour un seuil de 800. Ce chiffre donne un nombre d'arêtes voisin du nombre de nœuds. Par ailleurs le choix d'un nombre optimal d'arêtes est aussi en partie guidé par le graphe référent de la Figure 1, car c'est de cette façon que la sémantique civile a été définie manuellement avec 30 arêtes. En prenant en compte ces deux points de vue nous avons choisi 33 comme le bon nombre d'arêtes pour les trois différents réseaux et nous avons sélectionné la valeur seuil de distance correspondante pour chaque graphe. Ils peuvent être comparés après visualisation.

4.4. Visualisation des graphes réduits des trois forces

La projection des graphes réduits n'est pas unique et de nombreuses heuristiques sont applicables. Il nous faut faire un compromis entre le maintien au mieux des longueurs initiales des arêtes, une disposition qui permet de comparer visuellement les trois graphes entre eux et avec le graphe de référence de la Figure 1. De plus il convient de respecter certains critères d'esthétisme tels que la symétrie, la séparation des nœuds et le minimum de croisement d'arêtes (Battista et al. 1999). Il se trouve déjà que le graphe de la Figure 1 réalisé à la main respecte certains critères esthétiques (symétrie et minimisation des croisements d'arêtes). Pour faire le compromis entre ces différentes exigences, nous avons fait appliquer par notre environnement de projection de graphes l'algorithme de Kamada-S. Kawai (Battista et al. 1999) en fixant les nœuds P16 et P15 qui

correspondent aux deux mariés et qui sont au centre du graphe de référence. Puis nous avons corrigé les projections en déplaçant quelques nœuds à la main pour améliorer la disposition en particulier en ce qui concerne les singletons qui constituent environ 40% des nœuds pour les forces simple et de proximité. Nous les avons regroupés en haut à des places similaires pour ces deux graphes afin de les comparer.

4.5. Analyse sémantique des réseaux

Le processus de réduction en fonction du nombre de liaisons dans le réseau référent et un choix empirique d'une distance seuil a permis de produire trois graphes de plus petite densité qui peuvent être analysés et comparés plus facilement (Figures 4, 5 et 6). Nous trouvons alors une première évidence : les graphes de la force simple et de la force de proximité ont une topologie différente mais très similaire. Environ 40% des personnes ne sont pas connectées et les personnes restantes, la plupart du temps la famille proche, se rassemblent en étoile autour de la mariée et du marié. Sémantiquement, ces deux liens sociaux extraient des réseaux qui mettent en lumière le sens profond du mariage : l'union de deux familles par l'union de deux personnes. Les autres participants sont moins couplés à la cérémonie via les liens sociaux. Le réseau de la force de cohésion est très différent. Un couple est fort si les deux personnes sont présentes sur les photos et rarement séparées, ceci indépendamment du nombre de concepts. Dans le graphe résultant réduit de la Figure 6, seules trois personnes sont déconnectées. Le graphe présente de longues chaînes de nœuds. Les héros du mariage, la mariée et le marié, sont toujours là, avec les degrés le plus importants. Ils appartiennent désormais clairement à deux petites cliques qui sont leur famille proche respective. Les autres participants sont maintenant des acteurs solidaires de la cérémonie. Dans le réseau référent la plupart des gens sont décrits en fonction de leurs relations civiles avec les familles du marié et de la mariée. Cette information est retrouvée dans les deux graphes des forces simples et de proximité. Le graphe de cohésion témoigne plus des rencontres à l'occasion de la cérémonie. Le graphe de cohésion est moins un miroir civil qu'un miroir des relations sociales entre les participants qui se sont rencontrés à la cérémonie. Nous avons effectué des mesures de rappel sur les liens détectés qui donne les résultats suivants : 34,1%, 36,7%, 40% respectivement de rappel pour la force simple, de proximité et de cohésion.

5. Hypergraphes et diffusion des photos

Toujours dans la perspective de la diffusion personnalisée des photos, l'étape suivante consiste à chercher quel usage peut être fait de ces différents graphes. Il serait possible de définir un algorithme qui combine directement les distances de chaque personne à chaque personne avec la présence des personnes sur les photos pour déterminer la pertinence de

chaque photo pour chacun. Mais une telle méthode masque un vécu social important. De différentes discussions menées avec des personnes qui se sont confrontées au problème de la diffusion des photos, il apparaît que le mariage est vécu comme un lieu d'appartenance à différents groupes d'individus qui se rencontrent pour la circonstance. Celui qui veut diffuser des photos souhaite envoyer des groupes de photos à des groupes d'individus susceptibles de s'y intéresser. Le terme «tribus» est souvent employé par les diffuseurs pour désigner les groupes d'individus.

Dans (Plantié & Crampes 2010) nous avons proposé différentes méthodes de construction d'un hypergraphe de tribus basées sur la connexité des graphes issus des trois modèles de forces, puis différentes méthodes de diffusion des photos dont nous avons évalué la «précision» en regard d'un modèle de diffusion de référence. Les résultats sont intéressants, mais l'organisation des tribus telle que nous l'avons définie est discutable sur plusieurs points. Par exemple les tribus sont vues comme des arêtes d'un graphe social unique et indépendant des personnes alors qu'une tribu est d'abord la perception par un individu d'un groupe de personnes et que les perceptions en général diffèrent selon les individus. De plus dans nos modèles les relations sociales sont symétriques : si A apprécie B, B apprécie aussi A. Ceci n'est pas toujours le cas dans la réalité. Par ailleurs la recherche de la connexité conduit à 2^n tribus possibles pour n individus avec des algorithmes NP-complets. Pour remédier à ces différentes observations, nous proposons ici un algorithme plus simple qui part de la perception des individus pour aboutir à une méthode de construction de tribus de complexité linéaire. Nous traitons ci-dessous le cas de la construction des tribus pour la force de cohésion.

Pour chaque individu nous considérons le sous ensemble des nœuds du graphe initial tel que le poids des arêtes qui relient les nœuds à l'individu soit inférieur à un seuil. Formellement l'ensemble $a_i = \{ x_i, s_i, G \}$, où G est l'un des graphes de distances, x_i un individu et s_i le seuil retenu pour cet individu, définit un sous-ensemble des nœuds de G centré sur x_i . L'ensemble des couples $\{ [x_i, a_i] \}$ définit un hypergraphe étiqueté sur G où $\{ x_i \}$ est l'ensemble des nœuds et $\{ a_i \}$ l'ensemble des arêtes étiquetées, chaque arête a_i étant étiquetée par x_i . Cet hypergraphe peut être représenté par une matrice carrée K qui a en ligne les arêtes et en colonnes les nœuds, les premières étant les individus tels qu'ils perçoivent leurs accointances, et les secondes étant les individus tels qu'ils sont perçus par les autres.

Ainsi chaque personne a une vision personnelle de ses accointances et cette vision n'est pas symétrique (« A se sent proche de B » n'entraîne pas nécessairement que « B se sent proche de A »). De plus chacun peut avoir son propre seuil de sociabilité s_i et ainsi paramétrer son disque de relations sociales. Cependant, une difficulté subsiste. Qu'est-ce qui détermine le seuil de chacun ? Nous ne traitons pas cet aspect dans la limite de cet article. Afin de mener à son terme le processus de diffusion des photos, nous considérons que tout le monde a la même valeur de seuil s_i . Nous la définissons sans le

justifier ici par la valeur à partir de laquelle le graphe G est totalement connexe, c'est-à-dire qu'il n'y a aucun individu non relié au groupe.

La diffusion des photos peut alors s'opérer en faisant appel à la combinaison de la transposée de la matrice d'adjacence de l'hypergraphe des photos H avec l'hypergraphe K des tribus (H^t a en ligne les individus et en colonnes les photos). Formellement, nous considérons la matrice D qui est le résultat du produit matriciel suivant : $D = K * H^t$, où $*$ est l'indice de Jaccard calculé entre chaque ligne K et chaque colonne de H^t . Les lignes de D sont donc les individus et les colonnes sont les photos. La valeur de chaque cellule est la perception de chacun pour chaque photo en fonction de la perception de sa tribu. Pour opérer concrètement la diffusion, il faut établir un seuil d'intérêt. De manière empirique, nous l'avons pris égal à 0,5, ce qui correspond à la valeur à partir de laquelle une majorité de personnes de la tribu est observable sur la photo.

Pour évaluer la performance de la diffusion nous avons mis en place une méthode basée sur la précision et le rappel qui sera présentée dans un autre article. La précision est supérieure à 91%, mais le rappel est décevant. Si le seuil d'intérêt que nous avons fixé à 0,5 est augmenté mécaniquement à 0,8 le rappel s'améliore à 63% et la précision diminue à 76%. Mais l'amélioration la plus notable consiste à remplacer le calcul final qui fait appel à l'indice de Jaccard par un autre. En effet cet indice introduit un biais évident : toute photo qui comporte peu d'individus a peu de chance d'être diffusée dans des tribus riches en individus. Une recherche en cours a pour but d'étudier un nouvel indice dont les performances sont plus intéressantes.

6. Conclusion

Cet article a montré comment une collection de photos sociales pouvait témoigner des relations vécues par des individus lors d'événements qui les rassemblent. Nous avons proposé une série de modèles et de méthodes pour extraire des réseaux sociaux d'abord sous forme de graphes, puis, arguant du fait que la collection est un hypergraphe, sous forme d'hypergraphes étiquetés. Ces derniers sont mis à profit pour proposer des stratégies de diffusion personnalisées des photos. Nous en avons synthétiquement montré une application dont les résultats sont prometteurs. L'étape suivante consistera à porter nos méthodes sur des réseaux sociaux élargis et accessibles en ligne.

Références

AMES M. & MANGUY L. (2006) PhotoArcs: A Tool for Creating and Sharing Photo-Narratives. Proceedings of ACM CHI.

- BATTISTA G., P. EADEAUS, R. TAMASSIA, AND I. TOLLIS. (1999). Graph drawing. Algorithms for the visualisation of graphs. Prentice Hall.
- BERGE, C. (1973). Graphs and Hypergraphs. New York: Elsevier,.
- CRAMPES M., DE OLIVEIRA-KUMAR J., RANWEZ S., VILLERD J. (2009). Indexation de photos sociales par propagation sur une hiérarchie de concepts. Actes de la conférence IC 2009. pp. 13-24.
- E. ESTRADA & J. A. RODRIGUEZ-VELAZQUEZ (2006). Subgraph centrality and clustering in complex hyper-networks. *Physica A: Statistical Mechanics and its Applications*, 364. pp. 581-594.
- EADES P., (1984). A Heuristic for Graph Drawing, *Congressus Numerantium*, vol. 42, pp. 149-160,.
- EKLUND P., DUCROU J., WILSON T. (2006) An Intelligent User Interface for Browsing and Search MPEG-7 Images using Concept Lattices. *Proc. of the 4th International Conference on Concept Lattices and Their Applications*, LNAI, Springer-Verlag.
- FERRÉ S. (2007) CAMELIS : Organizing and Browsing a Personal Photo Collection with a Logical Information System. *Proc. of the 5th International Conference on Concept Lattices and Their Applications*. pp. 112-123.
- FREEMAN L., WHITE D. (1993) Using Galois Lattices to Represent Network Data. *Sociological Methodology*, vol. 23, pp. 127-146.
- FREEMAN, L. C. (2000). Visualizing social networks. *Journal of Social Structure*, 1 (1).
- Ganter B. & Wille R. (1999). Formal Concept Analysis: Mathematical Foundations. Springer.
- GOLDER S. A. (2008) Measuring Social Networks with Digital Photograph Collections. *ACM Conference on Hypertext and Hypermedia*. June 19-21. Pittsburgh, Pennsylvania.
- JAIN R. (2010). Life = experiences (events) and vision. *Proceedings of the ACM international conference on Multimedia*. Firenze, Italy.
- JUNG J. J., EUZENAT J. (2007). Towards Semantic Social Networks. ESWC'2007. pp.267-280.
- MIKA P. (2005). Ontologies are us : A unified model of social networks and semantics. *International Semantic Web Conference*. p. 522-536.
- NEWMAN M. E. J. (2006). The Structure and Function of Complex Networks. *SIAM Rev.* 45, pp. 167- 256.
- PLANTIE M., CRAMPES M. (2010) Des réseaux de photos aux réseaux sociaux. création et utilisation d'un réseau social à partir de photos. *MARAMI, conférence sur les Modèles et l'Analyse des réseaux : Approches Mathématiques et Informatique*. Toulouse.
- ROODEN K., WOOD K. (2003) How Do People Manage Their Digital Photographs ? *Proc. of ACM CHI*.
- ROTH C., BOURGINE P. (2006) Lattice-based dynamic and overlapping taxonomies : The case of epistemic communities. *Scientometrics (impact factor : 1.74)*, VOL. 69(2), p. 429-447.
- WU P. & TRETTER D. (2009). Close & closer: Social clustering and closeness from photo collections. Proceedings of the seventeen ACM international conference on Multimedia.