



HAL
open science

Des réseaux de photos aux réseaux sociaux, Création et utilisation d'un réseau social à partir de photos

Michel Plantié, Michel Crampes

► To cite this version:

Michel Plantié, Michel Crampes. Des réseaux de photos aux réseaux sociaux, Création et utilisation d'un réseau social à partir de photos. *Revue I3*, 2011, Numéro spécial, volume 11 (numéro 1), pp.100-110. hal-00804709

HAL Id: hal-00804709

<https://hal.science/hal-00804709>

Submitted on 26 Mar 2013

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.

Des réseaux de photos aux réseaux sociaux. Création et utilisation d'un réseau social à partir de photos

Michel Plantié , Michel Crampes

EMA-LGI2P, Parc Scientifique Georges Besse, 30035 Nîmes Cedex, France
{ michel.plantie, michel.crampes } @ema.fr, { <http://www.lgi2p.ema.fr> }

Résumé

Avec l'avènement du monde numérique pour la photographie, les photos prises lors d'événements familiaux ou autres réunions sociales font apparaître des groupes cohérents de personnes intéressants dans l'optique des réseaux sociaux. Dans cet article, nous montrons les différentes analyses et conclusions que l'on peut tirer de ces photos sociales. Ces résultats peuvent contribuer à donner les réponses les plus appropriées pour différents objectifs sur les réseaux ou groupes de personnes identifiés. L'organisation du réseau de photos utilise des méthodes d'analyse formelle de concepts.

Mots-clés : Photos sociales , Réseaux sociaux, Analyse formelle de concepts, Diagrammes de Hasse. Techniques et outils de conception, Interfaces utilisateur, Environnements interactifs, Théorie des graphes, Algorithmes de Graphes.

Abstract

With the new digital photography world, photos taken at family events or other social gatherings reveal coherent groups of people having an interest for social networks concerns. In this paper, we layout the different analysis and conclusions that can be drawn from these social photos. These results may provide the most appropriate answers for different objectives on networks or identified groups of people. The network organization of photos uses methods of concepts formal analysis.

Key-words: Social Photos, Social Networks, Formal Concept Analysis, Hasse Diagrams. Design Tools and Techniques, User interface, Interactive environments, Graph Theory, Graph algorithms.

1 INTRODUCTION

Avec l'arrivée des photographies digitales et la disponibilité récente des appareils photos dans les téléphones mobiles, il est maintenant possible pour

une seule personne de prendre des milliers de photos en une seule année. Actuellement, chacun fait face à des problèmes soulevés par l'abondance : comment organiser, visualiser, rechercher et partager des milliers de photos. Dans ce papier, nous considérons uniquement les photos "sociales" indexées manuellement. Nous nous référons avec ce terme aux photos qui sont prises durant des événements familiaux ou lors de soirées entre amis. Ces photos représentent des individus ou des groupes de personnes. En considérant le rôle important que les photos numériques semblent prendre pendant des événements "sociaux", il est intéressant d'explorer leur capacité à agir en tant que nouveau langage social reflétant des relations sociales. Elles ont une charge émotionnelle pour les personnes représentées. A ce titre la diffusion de ces photos est un enjeu social. Par ailleurs le problème de la diffusion de photos en général s'avère difficile [8]. La contribution de cet article est de proposer des méthodes et des outils pour répondre à cet enjeu et aux difficultés qui l'accompagnent.

Il est aussi pertinent d'explorer de nouvelles applications dérivées de ce nouveau langage permettant de renforcer la socialisation. Des applications telles que "Facebook" ou Flickr manipulent déjà des photos dans des réseaux sociaux. Cependant, dans ces cas les réseaux sont construits en utilisant les personnes. Les photos sont uniquement des objets passifs qui sont volontairement partagés. Elles ne jouent pas un rôle actif dans la construction des réseaux sociaux. Des auteurs tentent d'utiliser les relations entre personnes pour l'aide à la reconnaissance des visages [15].

Nous nous intéressons ici à des événements sociaux comme des mariages ou soirées. Dans ce cadre les groupes d'individus sont toujours des communautés réduites de personnes. Les analyses que l'on peut effectuer sur ces groupes sont de nature différentes et plus précises que sur les grands réseaux sociaux de l'internet.

Pour considérer les photos comme des marqueurs sociaux, il est nécessaire d'interagir avec elles en les considérant comme des objets actifs qui par leur seule présence témoigne d'une relation sociale. Cette exigence doit être appliquée à différents niveaux dans le cycle de vie du réseau de photo et du réseau social.

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. 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 noeuds 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 diffé-

rents 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 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 excède rarement quelques centaines.

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 oeuvre 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. Chaque niveau est développé dans les sections ci-dessous.

La section 2 décrit les techniques pour organiser les photos sociales [2] de manière à révéler les relations sociales entre personnes. Pour remplir cet objectif, nous utilisons les méthodes d'analyse formelle de concepts pour construire les réseaux de photos expressifs, en particulier les hiérarchies de concepts et les diagrammes de Hasse. La section 3 présente plusieurs solutions pour extraire des réseaux sociaux de collections de photos organisées. La section 4 présente les différentes méthodes de formation des tribus. La section 5 présente une application pratique de constitution de réseaux. La section 6 présente la constitution d'albums personnalisés à partir des réseaux construits.

2 ETAT DE L'ART

Organiser socialement des photos et élaborer des réseaux sociaux à partir du contenu des photos est depuis peu un domaine très actif de recherche. Une méthode proche de nos travaux est présentée en [7]. L'auteur justifie le fait que les photos représentant des personnes peuvent être utilisées pour élaborer

rer des réseaux sociaux. L'auteur ne présente pas de réseau social déduit du réseau de photos. Par conséquent, il n'y a pas d'aide pour la construction et le partage d'albums personnalisés. Les photos sont également indexées dans un outil séparé qui n'utilise pas leur contenu social disponible. Il n'y a pas de réelle justification pour l'élaboration du réseau social à partir des photos. [2] présente une méthode différente incluant ce couplage entre deux outils, dont l'un est une aide à l'indexation manuelle, que nous étendons dans ce papier. Une autre approche présentée dans [17] est très proche de nos travaux construits des réseaux sociaux à partir d'un ensemble de photos en examinant l'apparition conjointe des individus sur les photos. Les auteurs utilisent des algorithmes pour maximiser la modularité de partition du graphe pour détecter des "clusters" sociaux. Notre travail utilise de nouveaux modèles de liens sociaux qui expriment mieux les relations sociales entre personnes. Dans ce contexte les techniques de "clustering" traditionnelles et les mesures de graphes comme celle de [6] sont utiles mais expriment que partiellement la sémantique des relations sociales entre les personnes. Dans l'article [2] les auteurs utilisent les hypergraphes pour des applications de sélection musicale pour représenter les données multiformes des personnes. Les hypergraphes comme nous allons le montrer sont très utiles pour représenter les données sociales et spécialement les réseaux sociaux déduits des photos.

L'utilisation des hiérarchies de Galois pour gérer des ensembles de photos, est décrite dans plusieurs applications telles que dans [3, 4]. L'analyse formelle de Concepts (FCA, Formal Concept Analysis) et les diagrammes de Hasse ont été introduits, il y a longtemps pour représenter les réseaux sociaux [5] et restent un domaine fructueux de recherche comme dans [14]. Nous introduisons une caractéristique principale de notre système : la capacité à organiser les photos sociales dans un réseau particulier. Un diagramme de Hasse est utilisé pour révéler le contenu social et élaborer le réseau social sous-jacent. Un diagramme de Hasse est une représentation bien connue d'un treillis de concepts (ou treillis de Galois) comme présenté dans [13]. Nous utilisons ce formalisme pour générer notre réseau de photos. La théorie des réseaux sociaux [16] s'intéresse à la morphologie des réseaux afin de trouver des "super-nodes" des éléments référents dans le réseau qui permettent la diffusion fluide de l'information à travers le réseau. Cependant, notre objectif est différent car nous cherchons à constituer des réseaux à partir de la connaissance, déduite des photos, que représente la présence conjointe de personnes et leurs fréquences dans un événement. Cette connaissance analysée est suffisante pour induire des liens sociaux entre ces personnes. Les techniques bien connues de clustering sont fondées sur des distances entre éléments. Notre objectif est différent et se situe en amont de ces techniques car il est à la fois de définir des distances liées aux liens sociaux et de raisonner à partir de ces liens qui permettent de conserver la sémantique du modèle.

3 GROUPES ET TRIBUS

3.1 Des réseaux de photos aux réseaux sociaux

Nous présentons pour rappel ci-après des éléments qui ont déjà été partiellement présentés dans [13] afin de clarifier notre approche. Chaque concept montre un groupe particulier de personnes (l'intension du concept) et contient une collection de photos (l'extension du concept) avec ce même groupe de personnes. Chaque concept correspond donc à un groupe de photos contenant les mêmes personnes. A partir de ce diagramme, nous pouvons déduire des réseaux sociaux par le calcul de distance, l'inverse de la proximité entre toute paire de personne. Différentes formules de proximité sont possibles et nous introduisons quelques unes d'entre elles. Nous définissons les variables suivantes :

x_i : un individu parmi les personnes trouvées dans la collection de photos,
 N : le nombre de personnes de la collection de photos,
 c : un concept dans le diagramme de Hasse,
 C : le nombre de concepts dans le diagramme de Hasse,
 n_c : le nombre de personnes dans un concept c ,
 $[c/x_i]$: l'ensemble des concepts qui contiennent x_i ,
 $[c/[x_i, x_j]]$: l'ensemble des concepts qui contiennent uniquement x_i et x_j ,
 $[c/x_i, x_j]$: l'ensemble des concepts qui contiennent au moins x_i et x_j ,
 $[c/x_i, x_j, x_k]$: l'ensemble des concepts qui contiennent x_i et x_j et x_k ,
 $[c/x_i \vee x_j]$: l'ensemble des concepts qui contiennent au moins x_i ou x_j ,
 $Car[.]$: la cardinalité d'un ensemble particulier,
 $[c/x_i \oplus x_j]$: l'ensemble des concepts qui contiennent x_i ou x_j mais non les deux,
 $[c/\oplus(x_i, x_j)]$: l'ensemble des concepts ne contenant ni x_i ni x_j .

3.2 Le concept de tribu

L'objectif de notre travail est de créer des albums personnalisés selon les liens entre les personnes présentes lors d'un événement et permettre le partage de photos entre les personnes concernées.

Plus formellement ils s'agit de trouver les individus x_i (les individus dans les photos) qui sont concernés par l'attribution d'un ensemble de photos. Nous devons pour cela introduire une nouvelle notion. Nous appelons Tribu un sous-ensemble de personnes x_i concernées par un album personnalisé. Tous les individus qui appartiennent à la même tribu possèdent un réseau social commun qui est mis en évidence par l'apparition de ces individus sur différentes photos sous des formes différentes. Nous raisonnerons ici sur un couple d'individus x, y , plus précisément sur le lien entre x et y . Le concept de tribu sera défini ultérieurement en utilisant les notations des FCA. Notre ensemble de photos et de personnes est un hypergraphe [1, 11] dont les sommets sont des personnes x_i et les arêtes représentent les tribus. Un hyper-

graphe est la généralisation d'un graphe, et les liens entre les personnes x_i seront toujours représentés par les photos dans l'hypergraphe. Dans cet article nous proposons plusieurs méthodes de formations des tribus ci-après présentant chacune des avantages et inconvénients. La notion de tribu sera approfondie dans nos travaux futurs.

3.3 Modèles de Forces de couples et tribus

Une tribu est définie par un sous ensemble d'individus obéissant à des règles. Il est possible de définir plusieurs règles d'appartenance. Ci-dessous nous définissons différentes lois qui constitueront des tribus différentes et nous comparerons ensuite ces différents découpages. Dans ce modèle nous considérons les occurrences de personnes en relation dans les concepts. Un concept est représenté par une photo qui caractérise "l'intension" d'un concept, c.a.d. le groupe de personnes. La notion de tribu et les notions de forces définies ci-après sont plus générales et vont au delà de la simple co-occurrence entre personnes. Ces notions prennent en compte des mesures sur l'ensemble des concepts et non pas uniquement la présence conjointe de deux personnes.

3.3.1 La force simple d'un couple

Ici nous définissons la "fréquence" d'un couple (deux individus dans une relation) comme le nombre d'occurrences de ce couple parmi les concepts divisé par le nombre de concepts :

$$FORCESIMPLE(x_i, x_j) = \frac{Car[c/x_i, x_j]}{C} \quad (1)$$

Cette métrique est intéressante car elle représente la fréquence d'apparition d'un couple capturé par le photographe dans différentes situations sociales. Si un couple est vu dans de nombreuses situations alors la relation entre ces deux personnes est plutôt stable. Par contre cette similarité n'exprime pas la force de la liaison entre deux personnes.

3.3.2 La proximité ou force pondérée d'un couple

En conformité avec [7] nous considérons que plus il y a de personnes dans une photo, moins les liens sont forts entre ces personnes. En conséquence pour exprimer la force d'un couple, le nombre de personnes qui sont présentes dans les photos avec ce couple doit être pris en compte. La proximité d'une tribu (respectivement d'un couple) est la somme des inverses des nombres d'individus présents dans les concepts qui contiennent la tribu, divisée par le nombre total de concepts. 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 une tribu (un couple) n'apparaît dans aucun concept, l'ensemble $c_k / ((x_i, x_j) \in c_k)$ est vide, donc la force est

nulle. Deux personnes apparaissant seules dans un seul concept, donne la proximité maximum de $2/2/1 = 1$.

$$PROXIMITE(x_i, x_j) = \frac{(\sum_{k, (k/(x_i, x_j) \in c_k)} \frac{2}{Car[c_k/x_i, x_j]})}{C} \quad (2)$$

3.3.3 Cohésion d'un couple

Intuitivement, on peut comprendre que le modèle social ci-dessus ne capture pas certaines observations qui peuvent être extraites automatiquement à partir du diagramme de Hasse. Parmi elles, il en existe une qui est particulièrement intéressante. Plus deux personnes sont vues séparément, plus elles sont indépendantes l'une de l'autre. Inversement quand deux personnes sont observées toujours ensemble, cela renforce l'hypothèse de liens particuliers entre eux. Nous l'appelons "cohésion". Le couple (x_i, x_j) a une forte cohésion lorsque x_i et x_j sont souvent représentés dans le même concept par rapport au nombre de fois où ils apparaissent dans un concept. La force de cohésion reprend l'indice bien connue de Jaccard que nous évoquerons plus tard dans cet article. On a donc :

$$COHESION(x_i, x_j) = \frac{Car[c/x_i, x_j]}{Car[c/x_i \vee x_j]} \quad (3)$$

4 MÉTHODES DE FORMATION DES TRIBUS

A partir des forces définies ci-avant, nous sommes en mesure de déterminer des liaisons entre personnes et donc de construire des graphes de personnes, c'est à dire un réseau social. Les tribus sont formées à partir des hypergraphes obtenus avec les calculs des forces entre personnes. Nous construisons un graphe valué en considérant les personnes comme les sommets et en insérant un arc entre chaque sommet dès que la force entre deux personnes est non nulle. Chaque arc comportera un poids qui est égal à la force de liaison entre les deux personnes reliées par cet arc. De nombreuses possibilités s'offrent à nous selon les graphes générés et leurs utilisations. La manière de diffuser les photos découlera directement de la méthode de calculs des groupes de personnes que nous nommons tribus. Dans ce chapitre, nous proposons différentes politiques et nous appliquerons l'une d'entre elles sur une application pratique réelle. Comme nous l'avons évoqué au début de ce document, les communautés que nous analysons sont toujours des ensembles réduits de personnes.

4.1 Calculs Introductifs : Densité et densité pondérée d'un sous graphe

Dans un graphe non valué (i.e. les arêtes n'ont pas de poids), la théorie des graphes définit la densité d'un graphe comme le rapport du nombre d'arêtes

sur le nombre d'arêtes du graphe complet (clique) correspondant. Le nombre d'arêtes d'un graphe complet est le nombre maximal d'arêtes dans un graphe et est égal à $\frac{(nb_sommets \times ((nb_sommets - 1)))}{2}$, $nb_sommets$ étant le nombre de sommets. Donc la densité est $\frac{2 \times nb_arêtes}{(nb_sommets \times ((nb_sommets - 1)))}$. Dans un graphe valué (i.e. les arêtes ont un poids), nous définissons la densité pondérée d'un graphe comme le rapport de la somme des poids des arêtes du graphe sur le nombre d'arêtes du graphe complet (clique) correspondant :

$$D_w = \frac{2 \times \sum_{nb_arêtes_{i=1}} poids_arête(i)}{(nb_sommets \times (nb_sommets - 1))} \quad (4)$$

4.2 Première méthode : extraction de tribus sur le graphe complet

Cette méthode procède en plusieurs étapes :

1. tracer le graphe à partir des concepts et l'une des forces présentées au paragraphe précédent,
2. énumérer tous les sous graphes possibles,
3. retenir les tribus les plus liées : les sous graphes qui ont la densité pondérée la plus forte.

4.3 Deuxième méthode : extraction des tribus en considérant chaque personne

Dans cette méthode, si le graphe était non valué, nous considérerions qu'une personne peut avoir des liens intéressants avec d'autres personnes si le nombre d'arcs qui les séparent est inférieur ou égal à 2. Au delà, nous considérons que ces deux personnes se connaissent beaucoup moins en général. Nous considérons cette limite en accord avec la théorie des petits monde [12] et en considérant qu'au delà d'un chemin de deux arcs entre deux personnes les liens entre celles-ci sont bien plus distants. Dans notre cas où les graphes sont valués, en plus de la considération précédente nous considérons qu'une personne peut avoir des liens intéressants avec d'autres personnes, si la somme des distances des arcs nécessaires pour la liaison entre deux personnes est au maximum égal à 2 fois la distance maximum des arêtes partant de ce sommet. Cela signifie qu'une personne peut être liée à une autre si elle la connaît directement et en ayant une valeur de distance maximum, ou bien si elle la connaît par l'intermédiaire d'une autre personne qui la connaît avec au maximum la même distance de liaison que cette personne intermédiaire. Cette méthode demande un calcul par personne.

1. Pour chaque personne :
 - (a) tracer le graphe selon les consignes précédentes à partir du sommet représentant la personne considérée,

- (b) retenir les tribus intéressantes : générer les sous-graphes et retenir ceux qui ont la densité pondérée la plus forte,
2. Comparer les tribus obtenues, et éliminer les doublons pour les tribus identiques.

4.4 Troisième méthode : extraction des tribus en procédant par filtrage sur les composantes connexes

Dans cette méthode, le principe est de classer les arêtes du graphe de concepts par distances croissantes. Puis à chaque valeur seuil de distance en débutant par la valeur de distance la plus faible, nous conservons les arêtes dont la distance est inférieure ou égale au seuil considéré ainsi que les sommets qui leurs sont reliés. Nous obtenons ainsi un graphe constitué des arêtes et sommets dont le seuil de distance est inférieur au seuil fixé. Pour chaque valeur seuil un graphe est construit. Les graphes sont de plus en plus étoffés au fur et à mesure que le seuil augmente. Cette méthode procède en plusieurs étapes :

1. classer les arêtes du graphe de concepts par distances croissantes,
2. choisir une valeur seuil, qui sélectionne un graphe construit avec les arêtes dont la valeur de distance est inférieure au seuil,
3. dans ce dernier graphe, deux alternatives sont possibles :
 - (a) soit nous limitons le graphe construit à 2 fois le poids maximum des arêtes partant de ce sommet,
 - (b) soit nous générons tous les sous graphes possibles à partir du graphe obtenu par seuillage, et retenons les tribus intéressantes : les sous graphes qui ont la densité pondérée la plus forte.

4.5 Quatrième méthode : extension des calculs de forces aux groupes

Dans cette méthode, nous étendons les concepts de forces à des groupes. Nous calculons alors pour chaque groupe possible la force de ce groupe au sens de la force choisie. Cette méthode procède en plusieurs étapes :

1. énumérer tous les sous graphes possibles,
2. Calculer la mesure de force choisie pour chaque sous graphe,
3. retenir les tribus intéressantes : les sous graphes qui ont la mesure de force la plus forte.

4.6 Cinquième méthode : combinaison des deux méthodes précédentes (4.4 et 4.5)

Dans cette méthode, nous procédons comme dans la troisième méthode ci dessus, mais en utilisant l'extension des calculs de forces de couples aux

tribus. Cette méthode procède en plusieurs étapes :

1. classer les arêtes du graphe de concepts par distances croissantes,
2. choisir une valeur seuil, et tracer le graphe associé,
3. dans ce graphe, retenir les tribus intéressantes : les sous graphes qui ont la mesure de force de tribu la plus forte.

4.7 Choix des tribus intéressantes

Le nombre de tribus possibles en fonction de la méthode choisie peut être important. Il est donc important de limiter les tribus à traiter. Nous pouvons considérer quelques méthodes de sélection :

1. classer les tribus par mesures de force ou de cohésion des plus fortes au plus faibles, puis choisir une valeur seuil en fonction :
 - (a) d'un nombre maximum de tribus,
 - (b) d'une valeur seuil sur la force considérée pour le graphe dont les tribus sont issues,
 - (c) un compromis des deux précédentes propositions.

Cette proposition nécessite donc de trouver une valeur seuil. La problématique se reporte alors sur le choix de la bonne valeur de seuil.

2. soit considérer les tribus ayant un nombre de personnes maximum : on considère alors que les tribus trop importantes ne regroupent pas des personnes très proches,
3. soit un compromis des deux précédentes propositions.

4.8 Combinaisons possibles

Plusieurs combinaisons des méthodes ci-dessus sont possibles. Nous avons déjà évoqué une combinaison dans la cinquième méthode. Chacune des méthodes ci-dessus pourra être utilisée en choisissant un type de force différent à chaque fois. Cela donnera donc au moins 15 méthodes différentes.

4.9 Avantages et inconvénients de chaque méthode

Nous avons proposé ci-dessus plusieurs méthodes pour déterminer les tribus intéressantes. Chacune a ses avantages et ses inconvénients.

1. La méthode 1 a l'avantage de pouvoir comparer tous les graphes possibles. Cependant l'inconvénient majeur est le nombre de graphes possibles, il est au nombre de 2^N , N étant le nombre de personnes. Pour une application réelle de 40 personnes comme l'application que nous présentons ci-dessous, le nombre de tribus est déjà très important.

2. La méthode 2 a l'avantage de limiter, à un facteur près, le nombre de tribus générées au nombre de personnes. Elle est également très adaptée à l'application de diffusion de photos. L'inconvénient est bien sûr de calculer des tribus différentes pour chaque personne. Dans cette méthode, on ne détermine pas de tribus pour l'ensemble des personnes et l'utilisation que l'on pourra faire des tribus ainsi trouvées sera plus limitée. L'inconvénient également est que nous devons choisir un seuil. Il faudra donc trouver des critères pertinents pour son choix.
3. La méthode 3 a l'avantage de ne pas proposer un calcul exponentiel mais un calcul linéaire. Cependant la deuxième alternative peut générer un nombre de tribus potentiellement important. Il sera cependant limité au graphe généré. L'inconvénient est que nous devons choisir un seuil qui déterminera le graphe construit.
4. L'inconvénient majeur de la méthode 4 est le nombre de graphes possibles, il est au nombre de 2^N , N étant le nombre de personnes. Son avantage est d'étendre le concept de force à des graphes.
5. La méthode 5 combine les avantages des méthodes trois et quatre tout en éliminant le problème du nombre exponentiel de graphes. Son avantage est d'étendre le concept de force à des graphes. L'inconvénient est que nous devons choisir un seuil qui déterminera le graphe construit. Il faudra donc trouver des critères pertinents pour son choix.

En terme de complexité et de temps de traitement, on distingue deux catégories, les algorithmes exponentiels selon le nombre de sommets, et les algorithmes linéaires selon le nombre de sommets.

4.10 Méthode choisie

Nous choisirons la méthode trois qui élimine le problème du nombre exponentiel de graphes. Cette méthode demande la sélection d'un seuil que nous évoquerons dans la partie suivante. La méthode cinq est relativement similaire à la différence qu'elle nécessite l'extension du calcul des forces aux tribus. Cela demande donc de redéfinir tous les concepts de forces. Nous ne le ferons pas dans le cadre de cet article.

5 EXPÉRIMENTATION PRATIQUE SUR LA TROISIÈME MÉTHODE

Nous avons considéré un événement réel de type mariage comprenant : 27 personnes et 127 photos. Conformément à la règle des treillis de concepts, nous considérons comme hypothèse que deux photos comprenant les mêmes personnes représentent le même concept. Nous calculons alors la matrice de présence 27×127 des personnes dans les concepts. Il y a en moyenne 3 personnes par photos réparties selon le graphique 1 . Puis pour chaque force nous calculons les différents graphes associés aux niveaux de seuil sur les

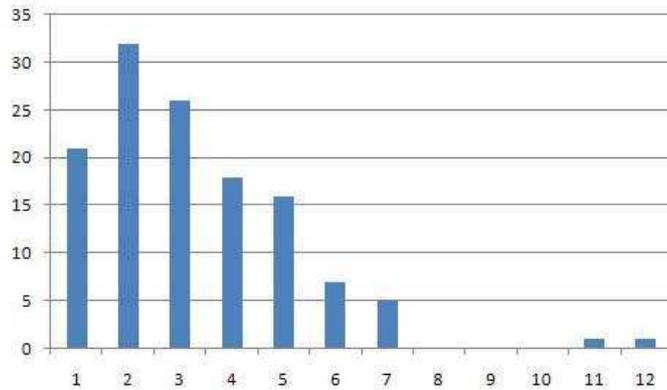


FIGURE 1 – répartition du nombre de personnes par photos

forces de liaisons entre sommets. Cela produit des graphes successifs de plus en plus complets au fur et à mesure que le seuil augmente. Afin de pouvoir valider les tribus obtenues par nos calculs, nous représentons ci-après (cf. figure 2) le réseau familial comprenant les personnes présentes sur les photos. Dans ce graphique, des personnes ne sont en relation avec aucun membre : ce sont les amis et autres relations.

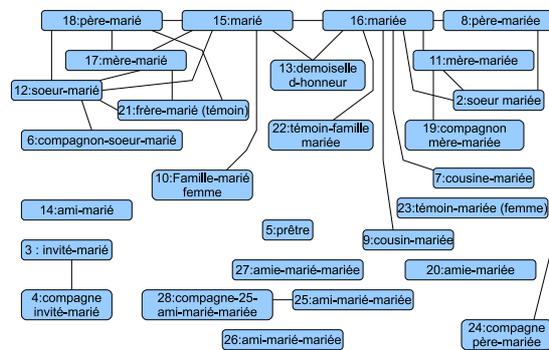


FIGURE 2 – réseau familial réel

5.1 Graphes des personnes selon les différentes forces

La force simple produit un graphe évolutif et toujours connexe (cf figure 3). Cette force fait apparaître une seule tribu. Pour obtenir davantage de tribus nous devons générer les sous-graphes de la composante connexe sélectionnée. La force de cohésion produit des tribus séparées qui deviennent de plus en plus connexes (cf figure 4). Cette force fait apparaître rapidement des

sous groupes intéressants. La force de proximité produit un graphe évolutif de plus en plus connexe (cf figure 5). Cette force fait apparaître deux tribus.

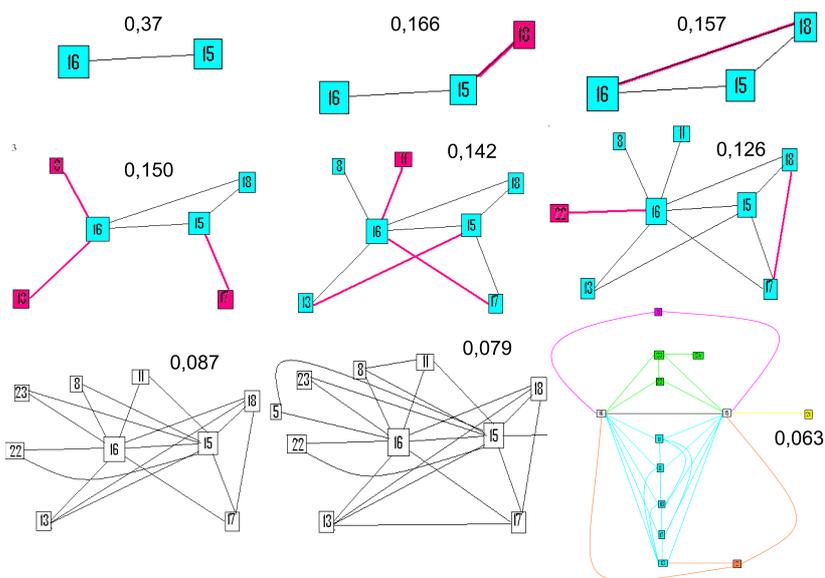


FIGURE 3 – graphes connexes de la force simple

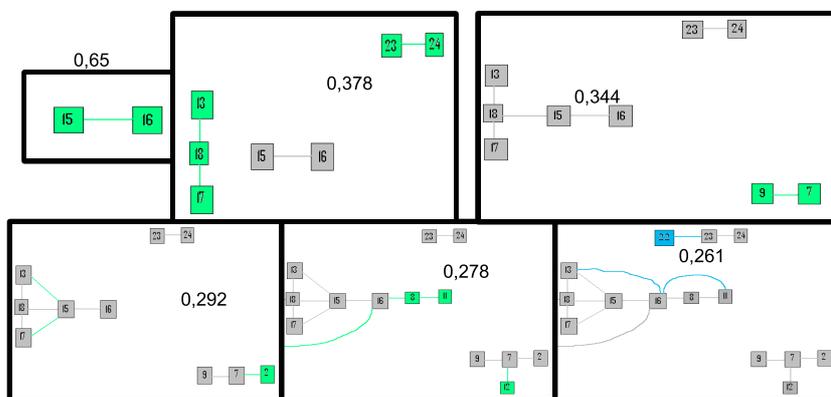


FIGURE 4 – graphes connexes de la force de cohésion

5.2 Choix du seuil des composantes connexes et sous-graphes

Afin de pouvoir comparer les différents effets des forces, nous choisissons de limiter le nombre de personnes à 13 connectées au moins une fois sur le graphe associé de chaque force, soit une personne de plus que le nombre de

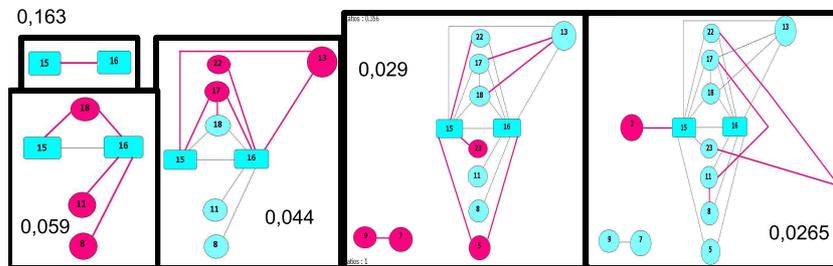


FIGURE 5 – graphes connexes de la force de proximité

personnes maximum sur une photo. Ce choix ne permettra pas automatiquement de représenter toutes les personnes de l'événement dans le graphe mais constitue un compromis entre complexité et utilité. Cette limite nous permet de fixer un seuil pour chaque force. Il est de 0,062 pour la force simple, de 0,277 pour la force de cohésion et 0,0265 pour la force de proximité. Nous choisissons la deuxième alternative proposée dans la méthode trois à savoir générer toutes les sous-graphes possibles à partir de la composante connexe sélectionnée par le seuil.

5.3 Sélection des sous-graphes intéressants de la force simple

Le tableau 1 montre les tribus générées et leur densité pondérée (cf. 4.1) pour la force simple. Nous avons choisi de citer uniquement les sous-graphes dont la connectivité brute a une valeur élevée. Nous avons volontairement éliminé les sous-graphes constitués d'un seul couple de personnes et de trois personnes. la sélection se fait selon le degré de connectivité en observant le graphe, car le nombre de sous-graphes est déjà important (2^{13}). Le dernier sous-graphe présenté correspond à la composante connexe complète. On remarquera que le calcul de densité pondérée est de plus en plus faible au fur et à mesure que le nombre de personnes augmente.

5.4 Sélection des sous-graphes intéressants de la force de cohésion

La force de cohésion pour une composante connexe comprenant 13 sommets (personnes), fait apparaître naturellement plusieurs sous-graphes non connectés (voir tableau 2). Nous limiterons les calculs à ces sous-graphes "naturels".

n° tribu	tribu	densité	densité pondérée	description
1	15-16-11-8	1	0,16	mariés, parents-mariée
2	15-16-22-23	1	0,14	mariés, témoins-mariée
3	15-16-11-8-18	0,9	0,1354	mariés, père marié, parents-mariée
4	15-16-11-8-17	0,9	0,1323	mariés, mère marié, parents-mariée
5	15-16-11-8-18-17	0,93	0,0945	mariés, parents-mariés
6	15-16-11-8-18-17-13	0,714	0,0937	mariés, parents-mariés demoiselle d'honneur
7	15-16-22-23-24	0,7	0,08	mariés, témoins-mariée..
8	15-16-11-8-18-17 13-22-23-24-5-2-21	0,423	0,001	composante connexe

TABLE 1 – densité des tribus de la force simple

n° tribu	tribu	densité	densité pondérée	description
1	23-24	1	0,444	femme-témoin-mariée et compagne-père-mariée
2	9-7-2-12	0,5	0,154	cousin-cousine-mariée soeur-marié-soeur-mariée
3	13-18-17-15-16-8-11	0,6	0,139	mariés, parents-mariés demoiselle d'honneur

TABLE 2 – densité des tribus de la force de cohésion

5.5 Sélection des sous-graphes intéressants de la force de proximité

La force de proximité pour une composante connexe comprenant 13 sommets (personnes), fait apparaître naturellement deux graphes non connectés (voir tableau 3). Nous générons également les sous-graphes du graphe le plus grand dont la connexité brute a une valeur élevée. On remarquera que le calcul de densité pondérée n'est pas systématiquement proportionnel au nombre de personnes de la tribu.

n° tribu	tribu	densité	densité pondérée	description
1	15-16-22-23	0,66	0,0473	mariés, témoins-mariée
2	8-11-15-16-18-17	0,6	0,0368	mariés, parents-mariés
3	8-11-15-16-18-17-13	0,571	0,0346	mariés, parents-mariés demoiselle d'honneur
4	9-7	1	0,03	cousin-cousine-mariée
5	8-11-15-16-18-17-13 2-5-23-22	0,381	0,0176	mariés, parents-mariés demoiselle d'honneur prêtre, témoins mariée

TABLE 3 – densité des tribus de la force de proximité

5.6 Discussion sur la validité des tribus

Le classement décroissant par densité pondérée des tribus mesure la solidité des tribus détectées. L'évaluation que nous pouvons faire des groupes détectés est essentiellement qualitative comparativement à une référence qu'est le graphe familial réel. Dans un événement de type mariage, nous pouvons déduire des groupes issus du bon sens :

1. la famille du marié,
2. la famille de la mariée,
3. les amis du marié,
4. les amis de la mariée,
5. dans un second temps : réseau de proximité familial du marié, réseau de proximité familial de la mariée,
6. dans un troisième temps : amis des frères et soeurs du marié et de la mariée.

la force simple reproduit assez fidèlement les relations familiales mais peine à faire ressortir les relations amicales,

la force de cohésion fait émerger en priorité les relations périphériques amicales puis les relations familiales,

la force de proximité donne des résultats similaires à la force simple.

Les forces simple et de proximité sont complémentaires à la force cohésion. L'ensemble des trois forces font ressortir une grande partie des relations réelles.

6 POLITIQUE DE DIFFUSION DE PHOTOS, CONSTITUTION D'ALBUMS PERSONNALISÉS

Nous utiliserons une technique originale pour constituer des albums. Le principe est de mesurer la distance de chaque photo à une tribu. Cette mesure nous donnera la mesure de l'opportunité de diffuser une photo à un groupe de personnes.

Nous utilisons la distance de Jaccard [9] qui mesure la dissimilarité entre deux ensembles. Elle est obtenue en divisant la différence de taille de l'union et de l'intersection de deux ensembles : $J_\delta(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}$. Les deux ensembles sont les individus présents dans une photo et les individus présents dans une tribu. Nous pondérons alors la distance d'une photo à une tribu par la densité pondérée des tribus associées à chaque force. Nous obtenons pour chaque force un classement décroissant des couples photos-tribus en fonction de leur distance pondérée.

6.1 Diffusion de photos

Cinq alternatives de diffusion sont possibles de la plus restrictive à la plus sociale :

1. Diffusion des photos aux seules personnes présentes sur les photos
2. Diffusion des photos aux personnes présentes sur les photos et aux tribus dont la distance photo-tribu est inférieure à une valeur donnée selon une règle sémantique. Fixons comme règle pour diffuser une photo à tous les membres d'une tribu, le fait que la distance de Jaccard photo-tribu soit inférieure à 0,5, qui indique qu'une majorité des personnes de la photo font partie de la tribu.
3. Diffusion des photos aux personnes présentes sur les photos et aux tribus dont la distance photo-tribu est inférieure à une valeur donnée établie selon un critère économique. C'est à dire le nombre d'exemplaire de photos à diffuser, ici 1000 par exemple.
4. Cette solution est une variante sémantique de la solution 2. Nous classons les couples "photo-tribu" par ordre décroissant selon leur distance et nous diffusons les photos en déterminant un seuil à partir duquel on ne diffuse plus les photos aux tribus désignées. Ce seuil peut revêtir un critère plus sémantique comme le fait qu'un nombre majoritaire de membres de la tribu est présent sur la photo (différent du cas 2), la photo est diffusée à toute la tribu.
5. Diffusion de toutes les photos à toutes les personnes.

6.2 Pertinence des méthodes de diffusion

Le critère de choix entre différentes alternatives peut être fondé sur un critère économique : le nombre total d'exemplaires de photos diffusés. La détermination de la pertinence de la solution, est également donnée par le fait que chaque personne reçoit des photos des personnes avec qui elle a été en contact étroit (lien personnel, présence dans le même groupe, etc.) pendant la soirée et ne reçoit pas celles des personnes avec qui elle est très peu ou pas en contact. Etablir la pertinence de la diffusion des photos est difficile à moins de faire une enquête subjective auprès des personnes à qui les photos sont diffusées. Cette enquête sera de toutes façons soumise aux aléas de la subjectivité des personnes.

6.2.1 Critères objectifs

Pour obtenir des critères plus objectifs, le dialogue avec les utilisateurs et le bon sens nous permettent de définir les critères objectifs de validité qui suivent.

1. C1 : chacun reçoit toutes les photos dans lesquelles il est présent,

2. C2 : chacun reçoit les photos dans lesquelles il est présent et les photos dans lesquelles il n'est pas présent contenant des personnes proches,
3. C3 : C1 plus les photos de groupes de personnes dont il est moins proche mais qui ensemble présentent un intérêt pour lui.
4. C4 : tout le monde reçoit la photo de groupe s'il y en a une.

6.2.2 Affinement de ces critères

Le critère C4 est par définition respecté par le critère C1 puisque tout le monde est sensé être dans la photo de groupe.

Concernant les notions de "personnes proches" dans le critère C2, nous pouvons définir deux types de relations : proches et lointaines. Le critère C2 considérera uniquement les relations proches. Les relations proches sont : personnes mariés ou en couple, relation parents-enfants, relation frères-soeurs, relation amis proches. Les relations plus éloignées sont : famille (autre que précédemment), proche de la famille, proche des amis, inconnus.

La notion de "proche" pour le critère C3 sera discutée dans un article ultérieur. Nous limiterons donc l'évaluation au critère C2.

Nous évaluerons la fiabilité de la méthode de diffusion par la notion de précision. Etant donné n objets retournés par une requête, la précision est la proportion exprimée en pourcentage d'objets pertinents divisé par le nombre d'objets retournés. Dans notre cas : la proportion parmi toutes les photos diffusées aux personnes de celles qui vérifient le critère C2.

6.2.3 Evaluation des différentes solutions

1. La solution 1 (diffusion des photos aux personnes présentes sur les photos) représente 415 photos et est déjà abondamment pratiquée. Le degré de satisfaction (100% du critère C1 et C4) de chaque participant correspond au minimum que l'on peut attendre d'un outil de diffusion personnalisé. La précision est de 100%.
2. La solution 2 (diffusion des photos aux personnes présentes sur les photos augmenté des photos à distance de Jaccard photo-tribu inférieure à 0,5, soit 100% du critère C1 et C4 et pourcentage partiel du critère C2) a été évaluée avec les différentes forces. Chaque force définit des tribus différentes. En calculant la mesure photo-tribu nous obtenons donc des photos diffusées différentes uniquement du fait de la différence de définition des tribus :
 - (a) Pour la force simple le nombre de photos est de 445 en considérant les tribus 6 et 7 qui représentent les sous-graphes précédents avec le nombre de sommets maximum. En effet elles couvrent les tribus 1,2,3,4,5, avec des scores élevés de densité et en même temps elles comportent presque tout l'ensemble des personnes de la composante connexe. Cependant aucune tribu de densité élevée ne contient les trois personnes manquantes de la composante

connexe. Cette solution est intéressante : chaque personne reçoit les photos où elle se trouve et également les photos où elle n'est pas présente mais dont les membres sont des personnes proches en terme de fréquence conjointe d'apparition. Sans prendre en compte les photos de la partie correspondant à la solution 1, La précision est de 28/30 soit 93,4% . La précision globale de cette solution représente 437/445 soit 98,2%.

- (b) Pour la force de cohésion le nombre de photos est de 435 en prenant les tribus "naturelles" 1,2 et 3. Cette solution prend en compte les avantages de la force de cohésion. Chaque personne reçoit les photos où elle se trouve et également les photos où elle n'est pas présente mais dont les membres sont des personnes proches au sens de la force de cohésion. Les personnes ne recevront pas de grandes quantité de photos et seront plus satisfaites car elles obtiendront les photos des personnes avec lesquelles elles ont été très proches pendant l'événement. Sans les photos de la partie correspondant à la solution 1, la précision est de 19/20 soit 95%. La précision globale de cette solution représente 434/435 soit 99,8%.
- (c) Pour la force de proximité le nombre de photos est de 455 en prenant les tribus 1, 3 et 4 qui couvrent toutes les personnes de la composante connexe. Cette solution prend en compte les caractéristiques de la force de proximité. Chaque personne reçoit les photos où elle se trouve et également les photos où elle n'est pas présente mais dont les membres sont des personnes proches au sens de la force de proximité. Le degré de satisfaction des personnes est important car elles retrouvent les photos des personnes avec lesquelles elles ont eu des liens plus personnels dans l'événement. Sans les photos de la partie correspondant à la solution 1, la précision est de 27/40 soit 67,5%. La précision globale de cette solution représente 442/455 soit 97,1%.

Avec cette méthode on obtient un nombre plus important de photos que la solution 1, sans qu'il ne soit trop élevé. Les photos diffusées sont plus pertinentes. On observe une différence entre les forces : la force de cohésion extrait plus facilement les personnes proches, la force de proximité a tendance à extraire davantage les personnes éloignées et diminue la précision et la force simple se situe entre les deux.

- 3. La solution 3 (critère économique : 1000 exemplaires diffusés maximum) représente 1000 photos par définition. Le degré de satisfaction des participants est moins important que dans les solutions précédentes : l'arbitraire dans le choix du seuil (lié à un critère économique), diminue l'intérêt de l'ensemble des photos diffusées. Une majorité des photos sont pertinentes, mais un ensemble non négligeable de photos augmente peu la satisfaction des participants tout en ayant le défaut

d'augmenter le nombre de photos inutiles et donc de noyer les photos utiles et agréables avec les moins utiles.

4. La solution 4 (diffusion des photos aux personnes présentes sur les photos augmenté des photos dont une majorité des membres des tribus sont présents sur la photo) a été évaluée avec les différentes forces :
 - (a) Pour la force simple le nombre de photos est de 454 en considérant les tribus 6 et 7. Chaque personne reçoit les photos où elle se trouve et également les photos où elle n'est pas présente mais dont les membres sont des personnes appartenant en majorité aux mêmes tribus que la personne concernée. Sans prendre en compte les photos de la solution 1, la précision est de 39/39 soit 100%. La précision globale de cette solution représente aussi 100%.
 - (b) Pour la force de cohésion le nombre de photos est de 506 en prenant les tribus "naturelles" 1,2 et 3. Cette solution prend en compte les avantages de la force de cohésion. Chaque personne reçoit les photos où elle se trouve et également les photos où elle n'est pas présente mais dont une majorité de membres sont des tribus proches au sens de la force de cohésion. Cette solution diffuse à chaque personne des photos de personnes qui ont montré une fréquence d'apparition simultanée importante avec la personne concernée. Sans prendre en compte les photos de la solution 1, la précision est de 78/91 soit 85,7%. La précision globale de cette solution représente aussi 97,4%.
 - (c) Pour la force de proximité le nombre de photos est de 558 en prenant les tribus 1, 3 et 4 qui couvrent toutes les personnes de la composante connexe. Cette solution prend en compte les caractéristiques de la force de proximité. Chaque personne reçoit les photos où elle se trouve et également les photos où elle n'est pas présente mais dont une majorité de membres sont des tribus proches au sens de la force de proximité. Le degré de satisfaction des personnes est important car elles retrouvent les photos des membres de tribus avec lesquelles elles ont eu des liens plus personnels dans l'événement. Sans prendre en compte les photos de la solution 1, la précision est de 143/143 soit 100%. La précision globale de cette solution représente aussi 100%.

Avec cette méthode on obtient un nombre plus important de photos que la solution 2, sans qu'il ne soit trop élevé. Les photos diffusées sont plus pertinentes au sens des tribus. Dans cette solution, la force de cohésion favorise davantage la diffusion de photos de personnes plus éloignées et diminue la précision et les forces simple et force de proximité extraient plus facilement les personnes proches.

5. La solution 5 (diffusion de toutes les photos à tout le monde) représente 3429 photos cumulées et est déjà abondamment pratiquée. C'est

la solution extrême qui oblige chacun à trier localement les photos qui l'intéressent et que les travaux que nous présentons ici se propose d'éviter. Parmi ces photos, un bon nombre ne leur apporte aucune valeur ajoutée.

Les stratégies évaluées fondées sur la sémantique permettent globalement de diffuser un nombre de photos plus restreint tout en augmentant le degré de satisfaction des participants. La diffusion des photos est habituellement un casse-tête pour les personnes en charge de cette diffusion. Elles ne souhaitent pas envoyer toutes les photos mais plutôt les plus intéressantes. Nous proposons ici un ensemble de solutions élégantes et efficaces.

7 CONCLUSION

Nous avons présenté des méthodes d'extraction de réseaux sociaux à partir de photos pour leur diffusion, dans l'intention d'utiliser également ces réseaux pour d'autres objectifs. Les différentes stratégies présentées sont modifiables et montrent la richesse du modèle présenté. Nous avons évalué quelques stratégies sur un cas réel assez important, en précisant que nous traitons toujours des communautés réduites de personnes associées à des événements sociaux de type mariage.

Les modèles peuvent être étendus pour prendre en compte plus efficacement la sémantique des relations et utiliser les réseaux extraits à des fins plus sémantiques. Nos travaux futurs vont explorer ces modèles sur des cas réels plus volumineux en approfondissant les aspects sémantiques. Dans les options que nous avons traitées, l'une d'entre elle est encore un problème très ouvert, c'est le choix du seuil pour l'alternative trois de construction des graphes. C'est à dire le choix de la distance pour le filtrage des arêtes du graphe. Nous traiterons ce problème important dans nos travaux futurs.

RÉFÉRENCES

- [1] Claude Berge. *Hypergraphes, combinatoires des ensembles finis*. Gauthier-Villars, 1987.
- [2] Michel Crampes, Jeremy de Oliveira-Kumar, Sylvie Ranwez et Jean Villerd. Indexation de photos sociales par propagation sur une hiérarchie de concepts. *Actes de la conférence IC 2009, Hammamet Tunisie*, pages 13–24, 2009.
- [3] P. Eklund, J. Ducrou et T. Wilson. 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*, 2006.
- [4] S. Ferré. 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, pages 112–123, 2007.

- [5] L.C. Freeman et D.R. White. Using galois lattices to represent network data. *Sociological Methodology*, 23, pages 127–146, 1993.
- [6] Linton C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40/1, pages 35–41, 1977.
- [7] Scott A. Golder. Measuring social networks with digital photograph collections. *ACM Conference on Hypertext and Hypermedia. June 19-21. Pittsburgh, Pennsylvania*, 2008.
- [8] Ramesh Jain. Life = experiences(events) and vision. *Proceedings of the ACM international conference on Multimedia , Firenze, Italy, October 2010*, 2010.
- [9] L. Kaufman, P. Rousseeuw et J.G. Battista P. *Finding groups in data : An introduction to cluster analysis*. WILEYInterscience, 1990.
- [10] Sonia Lajmi. Annotation et recherche contextuelle des documents multimédias socio-personnels. 2011.
- [11] Peter Mika. Ontologies are us : A unified model of social networks and semantics. *In International Semantic Web Conference*, pages 522–536, 2005.
- [12] S. Milgram. The small-world problem. *Psychology Today*, pages May/62–67, 1967.
- [13] M. Plantié et M. Crampes. From photo networks to social networks, creation and use of a social network derived with photos. *Proceedings of the ACM international conference on Multimedia , Firenze, Italy, October*, 2010.
- [14] C. Roth et P. Bourguine. Lattice-based dynamic and overlapping taxonomies : The case of epistemic communities. *Scientometrics* 69(2), pages 429–447, 2006.
- [15] Z. Stone, T. Zickler et T. Darrell. Autotagging facebook : Social network context improves photo annotation. *Proc. of the 1st IEEE Workshop on Internet Vision (CVPR 2008)*, 2008.
- [16] Lior J. Strahilevitz. A social networks theory of privacy. *American Law & Economics, Association Annual Meetings*, 2005.
- [17] Peng Wu et Dan Treter. Close & closer : Social cluster and closeness from photo collections. *Proceedings of the seventeen ACM international conference on Multimedia*, pages 709–712, 2009.