



HAL
open science

**Carrières d'algorithmes : la détection automatique de motifs dans des graphes (années 1950–1970).
Contribution à l'histoire des premiers apports des sciences sociales à l'informatique**

Sébastien Plutniak

► **To cite this version:**

Sébastien Plutniak. Carrières d'algorithmes : la détection automatique de motifs dans des graphes (années 1950–1970). Contribution à l'histoire des premiers apports des sciences sociales à l'informatique. ARCS - Analyse de réseaux pour les sciences sociales / Network analysis for social sciences, 2024, Concepts and methods in network analysis, 10.46298/arcs.10756 . hal-03861335v3

HAL Id: hal-03861335

<https://hal.science/hal-03861335v3>

Submitted on 12 Apr 2024

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.



Distributed under a Creative Commons Attribution 4.0 International License



Carrières d'algorithmes : la détection automatique de motifs dans des graphes (années 1950–1970). Contribution à l'histoire des premiers apports des sciences sociales à l'informatique

Sébastien Plutniak

CNRS, CITERES-LAT – UMR 7324, Tours, France,
sebastien.plutniak_at_cnrs.fr

Cet article est une version traduite et révisée de : Plutniak, S. (2021). « Assyrian merchants meet nuclear physicists: History of the early contributions from social sciences to computer science. the case of automatic pattern detection in graphs (1950s–1970s) » (M. Verdicchio, Éd.). *Interdisciplinary Science Reviews*, 46(4). *Computing in the World*, 547-568. <https://doi.org/10.1080/03080188.2021.1877502>

Résumé

La détection de communautés est une question centrale en analyse de réseaux. Cet article combine une approche socio-historique à la reconstruction expérimentale de programmes informatiques afin d'éclairer l'histoire des premiers algorithmes de détection de cliques, problème qui compte encore aujourd'hui parmi les problèmes NP-complets non résolus. Restituer les recherches menées par l'archéologue Jean-Claude Gardin depuis les années 1950 sur le traitement de l'information non numérique et l'analyse de graphes met en évidence ces contributions précoces à l'informatique réalisées depuis les sciences humaines et sociales. Ces applications originales de l'informatique aux humanités ont reçu une réception et une reconnaissance limitées. Ce fait est éclairé par deux facteurs : 1) les politiques de financement, qui ont motivé le transfert des efforts de recherche sur les graphes depuis un éphémère espace interdisciplinaire vers des organisations de recherche en informatique, domaine alors émergent ; 2) les carrières erratiques des algorithmes, où l'efficacité, les erreurs, les corrections et le statut des auteurs ont été des facteurs déterminants. Ces facteurs se combinent aux effets des historiographies et des bibliographies sur la conservation, la découvrabilité et la réutilisation des résultats scientifiques.

Mots-clefs : histoire de l'informatique, détection de cliques, détection de communautés, réseaux sociaux, sociométrie, analyse de graphe, algorithme

Introduction

Au cours des vingt dernières années, l'application de méthodes basées sur la théorie des graphes s'est largement diffusée dans de nombreux domaines scientifiques, y compris les sciences humaines et sociales (SHS). Parmi les promoteurs les plus visibles de ces méthodes, certains physiciens et informaticiens sont allés jusqu'à revendiquer une « science des réseaux » en tant que domaine scientifique à part entière (Barabási, 2002 ; Lewis, 2009). Cependant, l'utilisation de graphes pour l'étude de « réseaux » et, plus généralement, le recours à des algorithmes sont loin d'être nouveaux en SHS, contrairement à ce que suggèrent les récits contemporains signalés ci-dessus, qui insistent sur la nouveauté de la « science des réseaux ». Les promoteurs de cette idée négligent les premières applications et le perfectionnement des méthodes d'analyse de graphes développées par des chercheurs inscrits en SHS dès les années 1950.

Cet article traite de la manière dont ces premières applications ont été réalisées et des raisons pour lesquelles elles ont été oubliées ou rejetées. La détection de cliques, c'est-à-dire l'identification de sous-ensembles de nœuds tous connectés entre eux dans un graphe, illustre ces innovations algorithmiques précoces. Ce cas exemplaire est exploré à partir des recherches collectives animées par le français Jean-Claude Gardin (1925–2013), archéologue, linguiste et spécialiste de documentation. Cette étude prolonge les recherches récentes ayant souligné le rôle majeur joué par Gardin durant la seconde moitié du xx^e siècle en matière d'épistémologie (Plutniak, 2017 ; Plutniak, 2019) et d'application en SHS de travaux menés en recherche d'information, analyse automatique du discours, et systèmes experts (Dallas, 2015 ; Léon, 2015, p. 135-155).

1 Le problème des cliques

Étant donné un phénomène modélisé par un graphe (par exemple, des relations d'amitié entre individus, des relations entre protéines, des relations trophiques entre animaux), il peut être intéressant de détecter les sous-ensembles particulièrement denses de ce graphe. La plupart des méthodes actuelles identifient des classes de nœuds en déterminant des sous-graphes présentant une plus grande densité d'arêtes (ce type de méthodes est aujourd'hui communément appelée « détection de communautés »¹) ou en détectant des motifs élémentaires, tels que les configurations triadiques (Wasserman et Faust, 1994). La clique compte parmi ces motifs. Néanmoins, la longue histoire de ce concept doit être recherchée non pas en mathématiques, mais en psychométrie et en sociométrie. Ces deux domaines

1. Beaucoup de ces méthodes sont basées sur l'optimisation de la modularité (Newman et Girvan, 2004).

de recherche, qui se recouvraient partiellement, ont été particulièrement investis par les mathématiciens (Freeman, 2004).

La sociométrie est un champ d'investigation créé au début des années 1930 par le psychologue roumano-américain Jacob Levy Moreno (1889–1974). Ce domaine s'est développé à partir de son livre majeur publié en 1934, *Who shall Survive? A New Approach to the Problem of Human Interrelations* (Moreno, 1934) puis autour de la revue *Sociometry* (publiée sous ce nom de 1937 à 1977). Dans ces travaux, Moreno employait des diagrammes nœuds-arêtes pour représenter et analyser les relations sociales observées empiriquement et recourait occasionnellement à la notion de clique, mais au sens de sous-ensemble dense.

Cette notion s'est généralisée dans la littérature sociométrique (Forsyth et Katz, 1946, par exemple) après que R. Duncan Luce (1925–2012) et Albert D. Perry, étudiants au département de mathématiques du *Massachusetts Institute of Technology*, publièrent, dans la revue *Psychometrika*, la première définition formelle du concept de clique en théorie des graphes (Luce et Perry, 1949). Cette définition est restée, jusqu'aux années 1970 (Alba, 1973), l'unique définition formelle disponible et demeure aujourd'hui une base des travaux menés dans ce domaine.

Claude Flament (1930–2019), psychologue français et spécialiste de la théorie des graphes, la reprenait également, une quinzaine d'années après sa publication, en rappelant l'origine sociologique du concept employé en théorie des graphes : une clique est « la systématisation d'une notion courante en *sociométrie* : tous les individus d'une clique se choisissent mutuellement les uns les autres » (Flament, 1965, p. 53).

D'un point de vue mathématique, une clique est usuellement définie comme un « sous-graphe complet maximal d'un graphe » (Moon et Moser, 1965) ou, formellement :

$G' = (X'; V')$, sous-graphe de $G = (X; V)$, est une *clique* si $(x, y) \in V' \Leftrightarrow x \text{ et } y \in X'$, c'est-à-dire, si tous les arcs possibles existent en G' . (Flament, 1965, p. 53.)

Puisqu'une clique peut contenir une clique, Flament précisait le concept de clique maximale, c'est-à-dire « une clique qui n'est pas un sous-graphe d'une clique ». Notons que la publication de définitions formelles n'a pas empêché les variations dans les emplois et dans la sémantique du concept de clique au cours des décennies suivantes : en témoigne, par exemple, la définition d'une clique comme un « sous-ensemble de membres qui s'identifient davantage les uns aux autres qu'ils ne le font avec les autres membres de leur groupe »², c'est-à-dire comme un synonyme du concept actuel de « communauté »³.

2. « subset of members who are more closely identified with one another than they are with the remaining members of their group » (Hubbell, 1965, p. 377).

3. Cliques et sous-groupes sont ainsi traités comme équivalents dans MacRae, 1960. Plus généralement, voir Lankford, 1974 pour un panorama des méthodes de détection ou

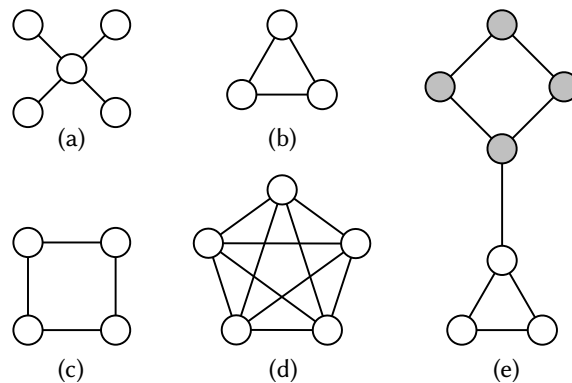


FIGURE 1 – Exemples de motifs de graphes mentionnés dans ce texte : une étoile (a), des cycles (b, c), des cliques (b, d), un graphe comportant deux communautés (e).

Malgré la simplicité de cette définition, automatiser la détection des cliques par un algorithme soulève de sérieuses difficultés computationnelles, principalement dues à l’explosion combinatoire qu’implique l’examen de chacune des paires de sommets d’un graphe. Richard Karp (1935–) a ainsi inclus la détection de cliques parmi ses 21 problèmes NP-complets, des problèmes difficiles n’étant pas calculables en un temps déterministe (Karp, 1972). Les recherches menées à partir de la fin des années 1950 par Gardin et ses collaborateurs ont donc compté parmi les premiers efforts pour aborder l’un de ces problèmes, automatiser la détection de cliques.

2 L’étude du réseau commercial assyrien de Gardin et Garelli

En 1961, Gardin et l’assyriologue Paul Garelli (1924–2006) publièrent dans la revue *Annales* (Gardin et Garelli, 1961) ce qui fut probablement la première application automatisée de la théorie des graphes à l’analyse de sources historiques. Ils visaient à reconstituer un réseau commercial actif au XIX^e siècle avant J.-C. dans l’actuelle Cappadoce. Il s’agissait pour cela : 1) d’automatiser la déduction des localisations géographiques des marchands mentionnés ; 2) d’identifier des groupes de marchands et la structure générale de leur réseau de relations commerciales ; 3) de discuter l’interprétation des aspects synchroniques et diachroniques du graphe obtenu ; 4) de proposer des représentations graphiques lisibles de ce graphe ; et 5) de déterminer le type et le degré de spécialisation de chaque marchand. Ces objectifs, qui recouvrent la plupart des principaux aspects actuels de

« cliques » et « communautés » sont considérées comme synonymes.

l'analyse empirique de réseaux, ainsi que le jeu de données utilisé dans cette étude, étaient inhabituels à l'époque : les objectifs, en raison de leur ambition ; les données, du fait de leur volume.

Cette reconstruction du réseau commercial assyrien s'appuyait sur le corpus de tablettes cunéiformes découvert sur le site de Kültepe (près de Kayseri en Turquie actuelle⁴). Ce corpus présente deux caractéristiques notables : d'une part, il documente par des textes des activités humaines très anciennes (2^e millénaire avant J.-C.) ; d'autre part, il comporte une quantité d'informations très importante, du fait que ces tablettes aient été collectées, traduites et publiées depuis le xix^e siècle, d'abord par des orientalistes puis par des assyriologues spécialisés. Au début des années 1920, environ 500 tablettes avaient été publiées ; à la fin des années 1950, ce nombre atteignait environ 2 600. Dès lors, cette abondance d'informations constituait un obstacle pour la recherche⁵.

Gardin envisagea de surmonter ce problème à la fois par l'échantillonnage des données et par le recours aux ordinateurs. L'échantillonnage pouvait être effectué à deux niveaux, celui des tablettes (les textes), puis celui des noms de marchands identifiés dans ces textes. Lors d'une réunion tenue en mars 1961, Gardin indiquait que l'achèvement du projet présupposait le codage des 2 000 tablettes étudiées par Garelli. Les résultats obtenus jusqu'alors reposaient sur l'analyse d'environ 200 tablettes⁶. Toutefois, aucune information n'était donnée à propos des critères utilisés pour sélectionner les tablettes à analyser. Dans une présentation à Paris au « Séminaire sur les modèles mathématiques dans les sciences sociales », Gardin parlait d'« environ » 1 000 tablettes (Gardin, 1961, p. 23). Le nombre de noms de marchands mentionnés dans les tablettes était estimé à 1 500 dans l'article de 1961, les auteurs soulignant les possibles problèmes d'homonymie⁷. Dans l'article de 1961, les analyses effectuées jusqu'alors portaient sur une « trentaine » de noms⁸. Chaque marchand était caractérisé par le nombre de transactions dans lesquelles il était impliqué, les auteurs considérant que ce nombre constituait une approximation robuste de l'importance relative des marchands (Gardin et Garelli, 1961, p. 854). Dans l'article de 1961, l'analyse était concentrée sur les trente commerçants les plus importants⁹ ; c'est à cet

4. Pour une carte de situation, voir celle adaptée d'après Gardin et Garelli, 1961, doi : [10.5281/zenodo.3934015](https://doi.org/10.5281/zenodo.3934015).

5. Pour une présentation plus détaillée de ces données, voir Plutniak, 2018, p. 7-11.

6. « Comité de direction. Réunion du 16 mars 1961 », 16-03-1961, JCG 1, MAE.

7. Gardin et Garelli, 1961, p. 847. Toutefois, lors du « Séminaire sur les modèles mathématiques dans les sciences sociales », Gardin parlait d'« environ » 3 000 noms (Gardin, 1961, p. 23) alors que dans Gardin, 1965a, p. 380, republication ultérieure de l'article des *Annales* en anglais, le nombre de 20 000 est donné.

8. Gardin et Garelli, 1961, p. 876.

9. Gardin, 1961, p. 25. Les auteurs considéraient que les analyses ultérieures porteraient sur une centaine de noms (Gardin et Garelli, 1961, p. 876 ; Gardin, 1962a, p. 457). Commentant postérieurement ces travaux, Gardin signalait des calculs effectués sur une matrice carrée correspondant aux relations entre 200 marchands (Gardin, 1965c, p. 389).

échantillon qu'ils appliquèrent l'algorithme de détection de cliques qu'ils développèrent.

Paradoxalement, malgré l'importance des données analysées et de la méthode employée, cette étude reste rarement citée dans la littérature historique et archéologique sur l'Orient ancien. Elle ne l'est pas davantage dans les initiatives ultérieures d'approches informatisées en assyriologie¹⁰, ni dans les études ayant à nouveau abordé les relations commerciales assyriennes par une formalisation fondée sur la théorie des graphes¹¹. Elle est également absente des quelques publications disponibles sur l'histoire de la théorie des graphes et de ses algorithmes (Lankford, 1974; Freeman, 1988; Barthélemy et Guénoche, 1988, p. 201-205; Pardalos et Xue, 1994). De plus, les travaux disponibles sur cette histoire se limitent à la chronique des différentes méthodes, sans considération pour leurs contextes sociaux et intellectuels de production. Les rares citations de l'étude de Gardin et Garelli dans les publications ultérieures ont contribué à son invisibilisation, quoique d'autres facteurs doivent être pris en compte.

Cet article adopte une perspective d'histoire sociale et intellectuelle¹² de l'informatique (De Mol et Primiero, 2014; Ensmenger, 2010). La première partie rend compte de la portée scientifique des recherches de Gardin. Sont ensuite abordés deux facteurs ayant limité la réception et le développement ultérieurs de ces applications informatiques en SHS : 1) un facteur organisationnel, lié au manque d'intérêt des historiens et des philologues pour cette approche, dont une conséquence fut le transfert du programme de recherche sur les graphes, que Gardin menait jusqu'alors depuis une organisation de recherche en SHS, vers des organisations de mathématiques appliquées et de physique ; 2) un facteur épistémique, lié aux capacités et aux potentielles défaillances et erreurs des différents algorithmes, lesquels déterminent également la possibilité et les modalités de leur utilisation.

10. Notamment dans la revue *CARNES – Computer Aided Research in Near Eastern Studies*. Voir par exemple le logiciel général d'assyriologie présenté dans Rouault, 1984.

11. L'étude de Gardin et Garelli est ainsi absente dans Constantine *et al.*, 1993; Bamman *et al.*, 2013; Wagner *et al.*, 2013, Waerzeggers, 2014, Barjamovic *et al.*, 2017; Gonçalves, 2021a; Gonçalves, 2021b. Une exception demeure : une thèse de doctorat, non publiée, dans laquelle un développement y est consacré (Anderson, 2017, p. 103-107). Pour une revue bibliographique détaillée de la réception des travaux de Gardin et Garelli, voir Plutniak, 2018, p. 30-37.

12. Archives consultées : archives Jean-Claude Gardin à la Maison Archéologie-Ethnologie, Université Paris X, Nanterre, France (abrégé ci-après MAE); fonds EURATOM aux Archives historiques de l'Union européenne, Fiesole, Italie (AHUE); fonds CADA au laboratoire d'archéologie CEPAM, Nice, France (CEPAM); fonds de l'Institut Blaise Pascal, archives privées de Pierre-Éric Mounier-Kuhn (IBP).

3 Un déplacement organisationnel : d'un espace interdisciplinaire vers la discipline informatique

L'analyse du réseau commercial assyrien est le fruit de la coopération entre deux nouvelles organisations de recherche axées sur les méthodes et l'interdisciplinarité : le Centre d'analyse documentaire pour l'archéologie (CADA) dirigé par Gardin à Paris et le Centre européen de traitement de l'information scientifique (CETIS), un service créé en 1961 dans le cadre du Centre commun de recherche de la Communauté européenne de l'énergie atomique (EURATOM). Dans cette section, après avoir présenté ces organisations et leur collaboration, je montrerai 1) comment le développement du traitement des données non numériques a constitué l'arrière-fond et la motivation de cette coopération et 2) comment cet effort interdisciplinaire inhabituel a pris fin, pour être ensuite poursuivi au sein d'organisations disciplinaires spécialisées en informatique.

3.1 Un espace interdisciplinaire précaire

3.1.1 Le Centre d'analyse documentaire pour l'archéologie

Le CADA était un laboratoire du Centre national de la recherche scientifique (CNRS) français, créé à l'initiative de Gardin et implanté à Paris et à Marseille. Malgré son nom, les activités qui y étaient menées ne se limitaient pas à l'archéologie. Elles concernaient des travaux méthodologiques en recherche d'information et en mathématiques, appliqués à un large éventail de disciplines dont l'archéologie, l'histoire, la littérature, la philologie, la linguistique et la médecine.

La composition du conseil d'orientation scientifique du CADA reflète le caractère peu commun de ce centre, tant par ses thématiques scientifiques que par sa position dans l'organisation institutionnelle de la recherche en France¹³. Il rassemblait des personnalités de la politique scientifique française, des orientalistes célèbres (par exemple Claude Schaeffer) ainsi que des chercheurs ouverts aux innovations méthodologiques (Henri Seyrig, orientaliste, ou Claude Lévi-Strauss, anthropologue). Ces acteurs, politiquement influents dans les institutions scientifiques, soutenaient les projets méthodologiques ambitieux de Gardin ; ces soutiens expliquent également l'existence et le financement du CADA. Par ailleurs, Gardin bénéficiait de relations privilégiées avec des membres des 5^e et 6^e sections de l'École pratique des hautes études (EPHE), en particulier avec certains des créateurs de la Maison des sciences de l'homme (MSH), dont Clemens Heller (1917–2002), proche collaborateur de Fernand Braudel (1902–1985). La MSH en était alors à ses débuts et n'existait que sous la forme d'une Association Marc Bloch. Ces bonnes relations aidèrent Gardin à obtenir un contrat de recherche

13. Pour une description détaillée, voir Plutniak, 2018, p. 11-14.

entre cette association et le Centre commun de recherche EURATOM en 1959, lui permettant de développer ses travaux sur ordinateur.

3.1.2 Le contrat de recherche entre EURATOM et l'association Marc Bloch

En juillet 1959, l'EURATOM mit sur pied un « Groupe de Recherche sur l'information scientifique automatique » (GRISA), placé sous la direction du mathématicien Paul Braffort (1923–2018). Une politique de financement de la recherche par contrat fut immédiatement développée¹⁴. Divers domaines scientifiques étaient concernés, dont la « sociologie »¹⁵ : à ce titre, un projet sur le « Traitement automatique de l'information dans les sciences humaines »¹⁶ fut financé, sous la direction de Gardin. Le budget initial était de 227 400 francs¹⁷ (438 355 € en 2023¹⁸). L'objectif de cette recherche était de développer un système de documentation automatique, partant du principe que « le langage des sciences sociales s'écarte assez peu du langage naturel, lequel est utilisé pour toutes les sciences exactes pour divulguer les résultats »¹⁹, ce qui justifiait le soutien d'EURATOM. Le groupe de recherche dirigé par Braffort avait pour objectif de développer un langage documentaire universel. Ses principales applications devaient être les domaines prioritaires d'EURATOM, à savoir la physique et les mathématiques, domaines où l'expression scientifique est fortement standardisée. Braffort et Gardin pensaient que, s'il était possible de développer un langage documentaire pour des domaines scientifiques où l'expression est beaucoup moins standardisée et plus dépendante du langage naturel (comme la sociologie), alors ce langage serait *a fortiori* applicable aux sciences moins complexes dont l'expression est plus standardisée :

C'est ainsi que nous désirons participer à la mise au point d'un langage documentaire adapté au domaine des sciences de l'homme, langage conçu de telle façon qu'il puisse s'inscrire dans le cas général du langage documentaire universel qui nous intéresse tous.

[...]

14. Comme en attestent les 22 contrats listés dans MEYER-UHLENRIED, Karl-Heinrich, « Organisation de la collaboration entre la Commission et les institutions scientifiques des pays de la Communauté pour l'élaboration d'un langage documentaire (principe des contrats) », 15-01-1960, BAC-059-1980 0209, AHUE.

15. « Sociologie » en un sens large, qui incluait la physio-psychologie.

16. Contrat 001-60-1 CETF du 10 mars 1960.

17. « Contrat de recherche entre la Communauté européenne de l'énergie atomique et l'Association Marc Bloch », 21-12-1959, BAC-118-1986 1442, AHUE.

18. Pouvoir d'achat équivalent en euros en 2023 compte tenu de l'érosion monétaire due à l'inflation, calculé grâce au convertisseur de l'INSEE (<https://insee.fr/fr/information/2417794>).

19. GARDIN, Jean-Claude, « Programme d'études sémiologiques et documentaire (1961–1965) », Octobre 1960, boîte Gardin 6, CEPAM, p. 3.

La personne la mieux qualifiée pour conduire une telle recherche est, sans contestation possible, Monsieur Jean Gardin [sic], Directeur du Centre d'Analyse Documentaire pour l'Archéologie pour le Centre National de la Recherche Scientifique.²⁰

Les calculs relatifs aux tablettes assyriennes ne concernaient pas directement ce projet (dont l'aboutissement fut la mise au point du langage de documentation SYNTOL²¹). Cependant, ces calculs ont été rendus possibles grâce aux ressources relationnelles et financières obtenues à travers ce contrat, autorisant l'accès aux calculateurs et la prise en charge du coût des heures de calcul (à savoir, 21,52 heures-machine sur des ordinateurs IBM 650)²².

3.1.3 L'accès à des calculateurs d'avant-garde

L'approche conçue par Gardin revenait à analyser systématiquement les entités mentionnées dans les textes akkadiens de manière à pouvoir les coder sur cartes perforées à l'aide d'un nombre élevé de descripteurs. Les informations de chaque « affaire économique » documentée dans les textes akkadiens ont été enregistrées à l'aide de 13 variables codées dans les 80 colonnes des cartes perforées²³. Ces données permettaient de générer des matrices représentant les relations commerciales. Une procédure d'analyse appliquée à ces matrices devait permettre de détecter des sous-graphes particuliers, dont les cycles, les étoiles et les cliques (Figure 1). Quoique le nombre de relations et de marchands soit relativement faible, ce type d'analyse génère rapidement une explosion combinatoire, nécessitant le recours à un ordinateur. Dans ce cas, les calculs furent réalisés par André Debroux (1932–) et Peter Ihm (1926–2014) (Gardin et Garelli, 1961, p. 875). Debroux travaillait pour IBM Belgique, tout en collaborant avec le GRISA, dont Ihm, un statisticien allemand, était membre. En 1960, le GRISA était situé à Bruxelles et ne disposait pas encore de ses propres ordinateurs. Les heures de calcul devaient donc être louées à d'autres organismes ou prêtées (Braffort et Gazzano, 1961, p. 56). L'Université libre de Bruxelles (ULB) disposait justement d'un ordinateur IBM 650²⁴ : ce fut donc soit cet ordinateur qui fut utilisé pour analyser les relations commerciales assyriennes, soit celui d'IBM Belgique.

Après la publication de 1961, Gardin envisagea de poursuivre l'expérience en intégrant davantage de commerçants dans l'analyse, tout en soulignant les limites imposées par les capacités informatiques insuffisantes disponibles en Belgique :

20. Source : Lettre de Paul Braffort à Louis Zieglé (de la DGRST) datée de février 1959, MAE.

21. *Syntagmatic Organization Language*, Cros et al., 1964.

22. « Travaux du CETIS », p. 8–9, 28-11-1960, BAC-118-1986 1431, AHUE.

23. Voir la description des variables dans Gardin et Garelli, 1961, p. 848-850.

24. Acquis en 1957 (Halleux et Xhayet, 2007, p. 159).

Le « calcul » de ces groupes, à partir de la matrice des relations observées entre une centaine d'individus pris deux à deux, est cependant une opération compliquée, qui dépasse déjà les possibilités d'un ordinateur moyen comme l'IBM 650; et un nouveau programme est sur le point d'être achevé, pour une machine plus puissante (IBM 7090), avec la collaboration du Centre européen pour le traitement de l'information scientifique (EURATOM). (Gardin, 1962a, p. 457.)

Ce problème mathématique et informatique fut ainsi « transmis » aux chercheurs du GRISA. En 1961, ce groupe de recherche avait été intégré au CETIS, dirigé par Braffort et implanté sur le complexe EURATOM d'Ispra, près de Varèse, dans le nord de l'Italie. Ainsi, le contrat de recherche entre l'association Marc Bloch et l'EURATOM facilita pour Gardin l'accès aux ordinateurs utilisés par les chercheurs du GRISA et aux mathématiciens capables de définir et de programmer les méthodes de calcul.

3.2 Le développement du traitement de l'information non numérique²⁵

En plus de ses activités dans le domaine des SHS, Gardin s'est impliqué dans une communauté scientifique émergente à l'intersection des mathématiques, de la logique, de la linguistique et de la documentation. Diplômé en linguistique, il avait mené ses premiers travaux scientifiques en documentation automatique et fréquenté le Centre français de recherche opérationnelle. Dès la fin des années 1950, il participa aux premiers groupes de recherche européens consacrés à la recherche d'information et, plus généralement, au traitement automatisé de l'information non numérique. L'analyse de graphe comptait alors parmi les problèmes d'analyse de l'information non numérique. Jacques Arsac (1929–2014), membre de l'Institut de programmation de Paris (créé en 1963), indiquait par exemple que sous le nom d'information non numérique « on enseignait principalement la théorie des graphes » (Arsac, 1988). De ce fait, pour les chercheurs investis dans ce domaine tel que Braffort, l'analyse du réseau commercial assyrien, pour laquelle Gardin avait sollicité leur concours, constituait un cas d'étude pertinent pour leurs préoccupations scientifiques du moment.

Le GRISA menait des travaux relatifs tant au traitement de l'information numérique (notamment le calcul des réacteurs nucléaires) que non numérique (notamment l'automatique documentaire). En 1960, Ihm comptait parmi les chercheurs qui y menaient des travaux de « mathématiques appliquées »²⁶. Cette année-là, son rapport d'activités précisait l'enjeu de ses recherches concernant le « Problème des marchands assyriens (dans le cadre du contrat Gardin) » :

25. Cette section résume Plutniak, 2018, p. 22-25.

26. « Rapport GRISA », 2, document daté de mai 1960, CEAB12-640, AHUE.

Les résultats de cette étude seront utilisés comme exemples-pilotes pour le problème plus général consistant dans la détermination dans un ensemble d'objets quelconques – qui se trouvent dans une relation quantifiable – le degré de dépendance, le point de gravité, etc.

On recherche un algorithme d'extraction automatique de l'information sur la base de la relation éventuelle existant entre les éléments de chaque couple d'objets.²⁷

Le problème posé à partir du cas du réseau commercial assyrien fut ainsi reformulé par Ihm en un problème mathématique plus général.

En somme, lorsque l'analyse du réseau commercial assyrien a été réalisée, Gardin pensait qu'elle démontrait l'intérêt d'automatiser le traitement d'informations non numériques en SHS. L'analyse de graphes n'était un objectif scientifique en soi que du point de vue des mathématiciens. Cette division s'est accrue au cours des années suivantes.

3.3 Le déplacement des travaux sur les graphes vers des organisations informatiques

Les projets de recherche ultérieurs à l'analyse du réseau commercial assyrien de 1961 manifestent un déplacement progressif des travaux sur les graphes vers des organisations de recherche dédiées aux mathématiques appliquées et à l'informatique.

3.3.1 Une étude complémentaire : les réseaux sociaux aux Nouvelles-Hébrides

Les méthodes développées pour le réseau commercial assyrien furent immédiatement réutilisées au CADA pour une autre recherche portant sur les réseaux sociaux contemporains aux Nouvelles-Hébrides (un archipel du Pacifique Sud, aujourd'hui État du Vanuatu, mais condominium franco-britannique jusqu'en 1980).

L'ethnographe Jean Guiart (1925–2019) menait des travaux sur la structure sociale dans ces îles, à travers l'étude du système par lequel les individus pouvaient recevoir des « titres » de prestige, liés à des lieux. Guiart avait collecté un ensemble de données comptant environ 1 200 titres et 500 lieux, associés à des individus par environ 2 700 relations (Espirat *et al.*, 1973, p. 370). La première mention de cette recherche en rapport avec le CADA date de 1962, après que Guiart ait demandé un soutien pour la réalisation d'analyses expérimentales sur ordinateur (Gardin, 1962b)²⁸. Deux

27. Rubrique consacrée aux activités de Peter Ihm dans « Travaux du CETIS », document daté du 28-11-1960, BAC-118-1986 1431, AHUE.

28. « Orientation des travaux à partir du 2e semestre 1962 », juin 1962, boîte Gardin 6, CEPAM, p. 3.

collaboratrices de Gardin supervisèrent cette collaboration : Marie-Salomé Lagrange (1935–2011), membre du CADA, et Monique Renaud, membre d'un autre groupe de recherche dirigé par Gardin, la Section d'automatique documentaire (SAD)²⁹ de l'Institut Blaise Pascal (IBP), alors principal lieu de recherche en informatique à Paris (Collinot et Mounier-Kuhn, 2010). Une personne supplémentaire fut recrutée pour coder les données sur cartes perforées, Madame A.-M. Nougaret³⁰. Des tabulations préliminaires furent réalisées par le Service mécanographique du CNRS et les principaux calculs furent exécutés en 1963 sur les ordinateurs IBM 1401 et 704 de l'IBP (Espirat *et al.*, 1973, p. 387). Lagrange et Renaud rédigèrent un rapport interne en 1965³¹ mais les résultats finaux n'ont été publiés qu'en 1973. L'analyse des graphes représentant les relations entre individus, titres et lieux comprenait la détection de structures « arborescentes », les chaînes et les cycles de différentes longueurs ; les cliques n'étaient, par contre, pas étudiées (Figure 2). Comme ce cas l'illustre, les organismes de recherche dédiés à l'informatique commençaient à jouer un rôle de premier plan, tant pour ce qui concerne les acteurs que les instruments de calcul.

3.3.2 Le langage de programmation développé à la SAD pour l'analyse de graphes

Fort des travaux menés à propos des cas assyrien et polynésien, Gardin comptait pousser plus loin leurs recherches sur les graphes. Dans un article présentant les recherches menées à la SAD, il insistait sur le caractère générique des problèmes relatifs aux graphes et l'intérêt de les approfondir :

Une seconde catégorie d'applications est également apparue, proche d'un aspect au moins de la documentation automatique. Celle-ci consiste souvent, on le sait, à manipuler en machine des structures de graphes : arborescence sémantiques, schèmes syntaxiques, etc... (le SYNTOL est d'ailleurs un bon exemple de langage documentaire où les données sont exprimées sous forme de graphes, sur l'axe paradigmatique comme sur l'axe syntagmatique). Or, il existe de fort nombreux problèmes, en particulier dans les sciences humaines, qui se ramènent également à la recherche de certaines configurations particulières dans un graphe plus ou moins étendu, différemment interprété : liens de parenté, relations sociales, transactions économiques, etc. (Gardin, 1965b, p. 15.)

Pour mener ces recherches, Gardin privilégia le cadre de la SAD plutôt

29. NSF, 1959, p. 121.

30. Comme c'est souvent le cas, les sources archivistiques conservent peu d'informations à propos des personnels subalternes.

31. LAGRANGE, Marie Salomé et RENAUD, Monique. 1965. *Étude d'un réseau sociologique aux Nouvelles-Hébrides, sur calculateur*, 55 p., mentionné dans Gardin, 1965b.

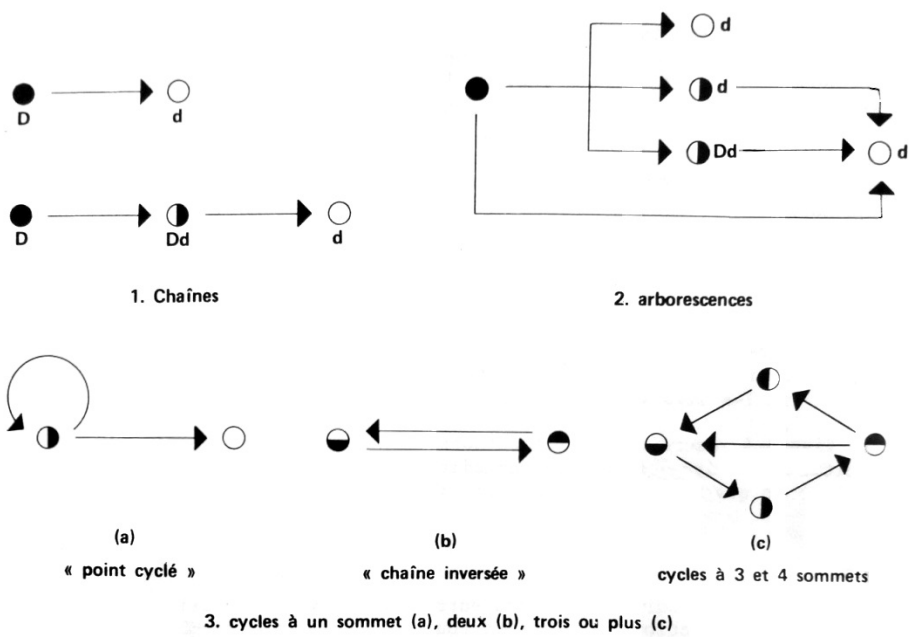


FIGURE 2 – Motifs d'intérêts dans les graphes de l'étude du réseau social des Nouvelles-Hébrides (d'après Espirat *et al.*, 1973, p. 390). « D » fait référence aux titres dominants et « d » aux titres dominés.

que celui du CADA, alors qu'il dirigeait toujours ces deux organisations de recherche. En 1964, il obtint une subvention de la toute récente Délégation générale à la recherche scientifique et technique (DGRST, créée en 1961³²) pour développer à la SAD un langage de programmation spécifique à l'analyse des graphes³³. Il en défendait l'intérêt dans l'article mentionné ci-dessus :

Cependant, la programmation des algorithmes correspondants [aux graphes] souffre parfois de certaines insuffisances propres aux langages symboliques usuels, conçus plutôt pour le calcul numérique ; et l'idée vint de chercher à définir un langage de programmation spécial pour le traitement de ces problèmes de graphes, dont l'expérience montrait qu'ils formaient une famille assez nombreuse pour justifier un tel projet. (Gardin, 1965b, p. 15.)

Dirk Muysers, alors ingénieur-programmeur à la SAD, fut chargé de ces recherches. Il bénéficia de l'appui d'un jeune mathématicien grec, Ion Arghiridis (1944–), recruté en 1965 à cet effet. Muysers mit au point ce langage, qu'il présenta dans un rapport communiqué à la DGRST en 1966³⁴. Le rapport s'achevait sur deux exemples d'algorithmes exprimés à l'aide du langage : le premier concernait l'énumération de tous les cycles d'un graphe orienté et le deuxième permettait de déterminer le plus court chemin entre deux sommets. Quoique prometteuses, ces recherches prirent toutefois rapidement fin, pour au moins deux raisons : d'une part Muysers cessa ses activités de recherche et quitta le CNRS en 1968 ; d'autre part, la DGRST acta le financement d'un autre projet sur les graphes, devant être mené à l'Institut de programmation de Paris.

3.3.3 Le « contrat graphes » de l'Institut de programmation

L'Institut de programmation avait été pensé comme un pendant à l'IBP dédié au développement de l'enseignement de la programmation et de l'informatique³⁵. Des recherches y étaient toutefois également menées, notamment en traitement de l'information non numérique. En 1964, un an après sa fondation, l'Institut bénéficia ainsi d'un contrat de recherche avec la DGRST pour des travaux sur les graphes. Ils furent menés par Jean Berstel (1941–) et Jean-François Perrot (1941–), deux anciens élèves de Marcel-Paul

32. Duclert, 2004.

33. Contrat entre la SAD et la DGRST n° 64 FR 175. Voir aussi : SAD, « Rapport d'activités situation 1963–1964 », 6-03-1964, boîte Gardin 6, CEPAM.

34. MUYSERS, Dirk, *Projet d'un langage de programmation numérique et non numérique, adapté plus particulièrement au traitement des données structurées par réseau*. Rapport à la DGRST, avril 1966, JCG 156, MAE.

35. Mounier-Kuhn, 2012, p. 428 ; Collinot et Mounier-Kuhn, 2010, p. 81-83.

Schützenberger³⁶, qui en cosignèrent le rapport final³⁷.

Mathématiciens de formation, Arzac, Perrot et Berstel se consacraient à différents aspects de l’informatique. Cette configuration contraste avec celle du CADA où la recherche en informatique était développée en conséquence et en vue d’applications à des données en SHS. Outre leur intérêt proprement mathématique, les recherches sur les graphes avaient des conséquences pratiques dans les domaines où elles pouvaient être appliquées, par exemple pour l’optimisation de processus en recherche opérationnelle. L’informatique devenait ainsi un domaine clé pour le développement national, comme l’illustre le « Plan Calcul » lancé par le gouvernement de De Gaulle en 1966 (Mounier-Kuhn, 2010). D’autres pôles de recherche se constituaient alors, comme le Centre de recherche en informatique de Nancy, où des travaux de théorie des graphes étaient également menés³⁸.

Ce contexte, et la réorientation de l’effort financier de la DGRST sur les organismes de recherche spécialisés³⁹, permettent de comprendre pourquoi Gardin et ses collaborateurs abandonnèrent leurs recherches sur les graphes. Par la suite, Gardin concentra leurs travaux sur des problèmes de recherche d’information. Dans ses écrits, il ne mentionna plus l’analyse du réseau commercial assyrien jusqu’au début des années 1970 lorsque, orientant ses recherches vers la simulation du raisonnement, il réinterpréta ses premières expériences sur les graphes comme une préfiguration de son nouvel intérêt pour l’intelligence artificielle⁴⁰. Néanmoins, un autre facteur explicatif s’ajoute à ceux mentionnés ci-dessus : les difficultés et échecs rencontrés sur le plan strictement informatique.

4 Carrières d’algorithmes, efficacité et correction

Après un résumé chronologique des méthodes de détection de cliques, cette section présente Bernard Jaulin et l’algorithme qu’il développa pour l’analyse des relations commerciales assyriennes. Par la suite, les résultats d’une reconstitution de cet algorithme et de sa comparaison avec d’autres méthodes sont présentés. Enfin, les effets du caractère erroné ou incorrect des algorithmes sur leur devenir sont discutés.

36. Le mathématicien Marcel-Paul Schützenberger (1920-1996), proche collaborateur de Paul Braffort, alors en poste à l’IPB, joua un rôle essentiel dans le transfert des approches mathématisées et formalisées en linguistique depuis les États-Unis vers la France (Léon, 2015, p. 123-125; Collinot et Mounier-Kuhn, 2010, p. 83). En outre, durant la guerre, il participa comme Gardin à la résistance extérieure en gagnant Londres en 1943.

37. PERROT Jean, BERSTEL, Jean-François. 1967. *Rapport final de la Convention de Recherche DGRST 65 FR 002*, « Contrat Graphes », polycopié, Institut de Programmation.

38. Pair, 1990; Mounier-Kuhn, 2014, p. 16.

39. À propos de la politique scientifique de la DGRST, voir Aust, 2016.

40. Voir à ce propos Plutniak, 2018, p. 26-30.

4.1 Une chronologie des méthodes de détection de cliques

En 1949, année de la première définition formelle du concept de clique (Luce et Perry, 1949), le psychosociologue Leon Festinger (1919–1989) suggéra l'idée d'une méthode algébrique de détermination de cliques (en passant une matrice au cube) (Festinger, 1949, p. 156). Il n'en détaillait toutefois pas d'algorithme. Dès le début des années 1950, des procédures d'ordonnement de matrice furent employées pour déterminer des propriétés structurelles d'un graphe, quoique sans être appliquées à la détection de cliques (Beum et Brundage, 1950). Le premier algorithme, désigné dans ce qui suit par « méthode Harary », reposait sur l'algèbre matricielle (Harary et Ross, 1957). Deux ans plus tard, une méthode basée sur la construction de graphes successifs fut proposée (Paull et Unger, 1959). Ces méthodes restaient toutefois strictement théoriques, sans que des implémentations logicielles furent publiées, et les chercheurs se satisfaisaient le plus souvent d'une identification visuelle des cliques, à partir de la représentation d'un graphe sous forme de diagramme (Carlson, 1960, p. 331). Ceci, jusqu'en 1960, lorsque les sociologues Duncan MacRae (1921–2008) et James Coleman (1926–1995) publièrent un programme de détection de *clusters*⁴¹ (c'est-à-dire de communautés), écrit pour une machine Univac 1 (Coleman et MacRae, 1960).

Ce survol résume l'état des recherches au moment où Gardin et Jaulin commencèrent à travailler sur le réseau commercial assyrien. La taille de leur jeu de données les obligeait à recourir à une méthode automatique sur ordinateur :

The mathematical problem of clique detection has been approached by many authors, from varying angles. The main issue, here, is in the size of the matrix which has to be analysed to that end (ca. 200 rows x 200 columns); for any given detection method, it is unrealistic to consider its application on so large a matrix without the help of a computer.
(Gardin, 1965c, p. 389.)

La méthode d'Harary leur était connue mais Gardin et Jaulin considéraient qu'elle pouvait être améliorée en termes d'efficacité et de rapidité (Jaulin, 1961, p. 51; Gardin et Garelli, 1961, p. 865). Jaulin développa donc son propre algorithme.

4.2 Bernard Jaulin et son algorithme de détection de cliques

En 1960, après avoir été promu ingénieur à l'École nationale supérieure d'arts et métiers et avoir obtenu une licence ès sciences, Jaulin fut recruté au « Bureau d'études pour le traitement automatique de l'information dans les sciences humaines ». Cet organisation venait d'être créée par Gardin dans

⁴¹. Pour des revues de littérature, voir Lankford, 1974 et Wu et Hao, 2015. Hubbell, 1965, p. 377, déjà cité, signalait également un programme de détection de cliques écrit en Fortran.

le cadre du contrat entre l'association Marc Bloch et l'EURATOM. Gardin, chercheur en SHS intéressé par les approches formelles, et Jaulin, mathématicien intéressé par les SHS, débutèrent une relation serrée, tant sur le plan scientifique qu'amical⁴². Ensemble, ils organisèrent en 1959 le premier programme de formation sur le traitement de l'information non numérique à l'IBP. Gardin demanda par la suite à Jaulin de contribuer à l'étude du réseau commercial assyrien. En 1964, Jaulin prit la direction du Centre de mathématiques appliquées de la Maison des sciences de l'homme (récemment créée à partir de l'association Marc Bloch). En juillet 1966, les deux hommes organisèrent à Rome une conférence intitulée « Les applications du calcul dans les sciences de l'homme »⁴³. Cette conférence était soutenue par le Centre international de calcul de l'UNESCO (Icc⁴⁴), également situé à Rome et dirigé par le mathématicien français Claude Berge (1926–2002). Berge était alors une figure de proue de la théorie des graphes et avait également participé, comme Gardin, au « *Non-numerical data processing symposium* » à Blaricum en 1961. La conférence de Jaulin et Gardin à Rome fut immédiatement suivie d'un « Séminaire international sur la théorie des graphes et ses applications », organisé par Berge et l'Icc, démontrant l'intérêt soutenu pour la théorie des graphes à cette époque.

Jaulin présenta son algorithme de détection de cliques maximales en 1961 au « Séminaire sur les modèles mathématiques en sciences humaines » de l'EPHE à Paris. Les documents de travail de ce séminaire sont la seule source connue décrivant cet algorithme (Jaulin, 1961, reproduit en [Annexe A](#)). Il existait un rapport interne du CADA intitulé « Sur une méthode de détermination des cliques dans un graphe symétrique », mais il n'a pas été conservé dans les archives⁴⁵. Jaulin présentait sa méthode comme une amélioration de l'algorithme d'Harary et Ross. Il suggérait une approche différente, fondée sur l'algèbre booléenne et sur la notion de points uncliqaux (procédant par réduction du graphe pour traiter les nœuds qui n'appartiennent qu'à une seule clique). Selon Jaulin, cette méthode aurait été adaptée aux graphes symétriques et non pondérés. Gardin souligna que cette dernière caractéristique, l'absence de pondération, constituait un manque pour ses besoins⁴⁶, car elle empêchait de tenir compte des intensités relatives des relations entre marchands :

Une objection vient à l'esprit : cette définition des cliques a le tort de n'envisager que la présence ou l'absence d'un lien entre

42. Des informations biographiques détaillées sur Jaulin peuvent être trouvées dans l'autobiographie de son collègue et ami, le mathématicien et poète Jacques Roubaud (Roubaud, 2008, p. 71–73 et 173–205).

43. Publiée ultérieurement comme Gardin et Jaulin, 1968.

44. Pour une histoire de ce centre, voir Nofre, 2014.

45. Ce rapport est catalogué dans les archives MAE mais le document est absent ; il n'est pas non plus conservé dans les archives du CEPAM et l'AHUE.

46. Une méthode de détection des cliques dans un graphe pondéré fut publiée pour la première fois en 1969 (Doreian, 1969).

deux marchands, sans lui accorder aucun « poids » selon le nombre ou la nature des relations qui l’expriment, dans l’analyse. (Gardin et Garelli, 1961, p. 864.)

Cependant, Gardin faisait aussi valoir 1) que l’absence d’échelle qui permettrait de définir les valeurs des pondérations limitait leur usage ; et 2) qu’ils souhaitaient se concentrer en premier lieu sur la détection des cliques et des cycles, sans tenir compte des pondérations des relations (Gardin et Garelli, 1961, p. 864).

L’algorithme de Jaulin a été implémenté deux fois. Dans les deux cas, les codes de programmation n’ont pas été retrouvés ou conservés. La première implémentation fut réalisée en 1960, par André Debroux (1932–?) d’IBM Belgique et Otto Hermann du CETIS⁴⁷. Ce programme avait été écrit en SOAP (*Symbolic Optimal Assembly Program*), le langage d’assemblage de l’ordinateur IBM 650, et exécuté sur ce type de machine chez IBM Belgique. La seconde implémentation date de 1962, pour un ordinateur IBM 7043 :

l’exploitation des archives de Kültepe (Anatolie) à l’aide d’un ordinateur [...] s’est poursuivie en 1961-1962 : analyse de quelques centaines de tablettes supplémentaires, traduites de l’akkadien par M. GARELLI, et mise au point d’une programmation nouvelle, pour ordinateur 704.⁴⁸

Cette seconde implémentation semble avoir été rapidement abandonnée. Je me suis donc appuyé sur les documents de travail du séminaire de 1961 pour reconstituer l’algorithme de Jaulin.

4.3 Une archéologie des algorithmes

À l’instar des archéologues pratiquant l’expérimentation pour reproduire des objets matériels et leurs usages, les historiens et historiennes de l’informatique s’attachent aujourd’hui à reconstituer les ordinateurs et les programmes⁴⁹. Les informaticiens se livrent parfois également à cet exercice. Par exemple, les mathématiciens français Jean-Pierre Barthélemy (1945–2010) et Alain Guénoche (1947–) ont implémenté l’algorithme de détection de cliques de Paull et Unger, trente ans après sa publication en 1959 (Barthélemy et Guénoche, 1988, p. 201-205). Soulignons que Guénoche a été membre du CADA à partir de 1971. Il a, par la suite, collaboré durant une dizaine d’années à d’autres groupes de recherche associés à Gardin.

Dans le cadre de cette étude, j’ai cherché à implémenter l’algorithme de Jaulin, avec le soutien de Guénoche. Toutefois, la description figurant

47. CETIS, *Travaux du CETIS*, 28-11-1960, BAC-118-1986 1431, AHUE.

48. Centre national de la recherche scientifique, *Rapport d’activité. Octobre 1961–Octobre 1962*, p. 417.

49. Voir par exemple la reconstitution des premières tentatives, menées en 1946, de résolution de problème numériques (De Mol et Bullynck, 2008), ou celles, datant de 1948, de simulations de réactions nucléaires (Haigh *et al.*, 2014).

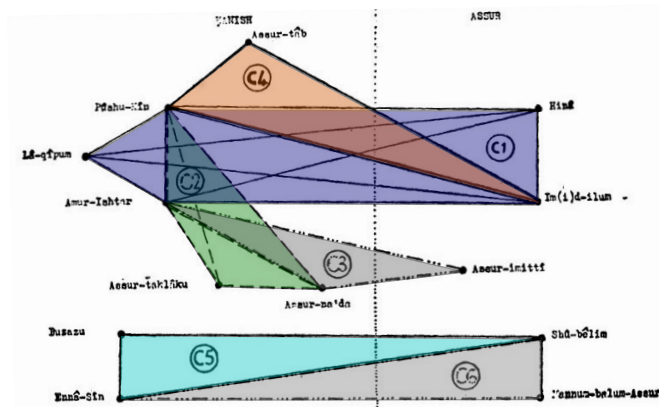


FIGURE 3 – Illustration de la détection de cliques dans le réseau commercial assyrien (adapté à partir de Gardin et Garelli, 1961, p. 868, Figure 10).

dans le document de travail de 1961 était si incomplète et confuse que cette tentative a échoué. Une solution alternative a consisté à extraire l'échantillon de données présenté dans l'article de 1961 à partir du diagramme illustrant la détection des cliques (Gardin et Garelli, 1961, p. 868), puis à traiter ces données avec d'autres méthodes afin d'en comparer les résultats. Cette approche s'apparente à celle de Hubbell, 1965, qui comparait les résultats de plusieurs méthodes appliquées aux données de MacRae, 1960. Soulignons que certains résultats de l'article de Gardin et Garelli sont contradictoires. Par exemple, la matrice d'adjacence présentée dans la figure 7 contient six cliques alors que la légende indique trois cliques et un sommet isolé, alors que ce dernier fait en réalité partie d'une clique. Par conséquent, les données ont été extraites non pas à partir de la matrice d'adjacence mais à partir du diagramme (Figure 3). J'ai ensuite appliqué cinq méthodes de détection de cliques, dont l'algorithme de Harary, que j'ai implémenté en langage R pour les besoins de cette recherche⁵⁰. Toutes les méthodes ont identifié les mêmes six cliques que celles rapportées dans l'article de Gardin et Garelli (Tableau 1). Cependant, la méthode Harary détecte deux cliques supplémentaires, qui constituent des sous-graphes d'autres cliques et ne sont donc pas des cliques maximales. Ce résultat erroné, ainsi que la légende contradictoire de l'article de Gardin et Garelli, invitent à considérer avec plus d'attention la question des erreurs dans le recours aux méthodes formelles.

50. Voir [Annexe B](#).

TABLEAU 1 – Identifiant et taille des six cliques indiquées dans Gardin et Garelli, 1961, comparées avec les cliques détectées ou non par les différentes méthodes.

| Id | Taille (sommets) | Gardin & Garelli 1961 | Harary 1957 | Bron 1973 | Makino 2004 | Osertgard 2001 | Eppstein 2010 |
|----|------------------|-----------------------|-------------|-----------|-------------|----------------|---------------|
| 1 | 5 | x | x | x | x | x | x |
| 2 | 4 | x | x | x | x | x | x |
| 3 | 3 | x | x | x | x | x | x |
| 4 | 3 | x | x | x | x | x | x |
| 5 | 3 | x | x | x | x | x | x |
| 6 | 3 | x | x | x | x | x | x |
| - | 3 | - | x | - | - | - | - |
| - | 3 | - | x | - | - | - | - |

4.4 Des algorithmes défaillants

Considérant que les différentes méthodes reposent sur la même définition des cliques, la différence des résultats générés par la méthode Harary laisse supposer une défaillance dans cet algorithme. Retracer l’histoire de cette méthode confirme cette hypothèse. Frank Harary (1921–2005) était un mathématicien américain spécialisé en théorie des graphes. Il nourrissait un intérêt particulier pour le développement d’applications à un large éventail de domaines, y compris l’anthropologie sociale (Hage et Harary, 1983). Après sa publication en 1957 (Harary et Ross, 1957), la méthode a été étendue aux graphes pondérés environ douze ans plus tard (Doreian, 1969). Dans l’intervalle, personne ne remarqua que l’algorithme de 1957 était défectueux, avant qu’Harary lui-même ne le fasse, treize années plus tard :

To set the record straight, that algorithm determines not only all the cliques of a graph, but sometimes a few other subgraphs as well. Correct algorithms for clique detection have subsequently been derived independently by several experts in computer programming. (Harary, 1970, p. 6.)

La méthode fut ainsi modifiée et implémentée en langage PL/1 (Dixon, 1973). Ce cas illustre comment une erreur algorithmique peut persister sur une longue période, sans que sa diffusion scientifique en soit limitée⁵¹.

Comme signalé précédemment, la détection des cliques n’a pas été appliquée à l’étude du réseau social des Nouvelles-Hébrides et a pratiquement

51. Un autre exemple d’erreur est donné par Augustson et Minker, 1970, p. 577, note 7, qui proposèrent une implémentation corrigée de l’algorithme de recherche de sous-graphe maximum de Raymond E. Bonner, notant que ce dernier avait omis d’en décrire une étape importante dans sa publication (Bonner, 1964).

été oubliée par Gardin et ses collègues dans leurs recherches ultérieures. La méthode de Jaulin est absente de la littérature spécialisée sur les graphes, à l'exception d'une brève mention dans l'ouvrage de Flament *Théorie des graphes et structures sociales*⁵². Les auteurs de l'ouvrage *Introduction to Finite Mathematics* reprennent les données de Gardin et Garelli pour discuter du recours à la détection de cliques afin d'identifier des marchands homonymes. Toutefois, ils n'abordent pas la définition des algorithmes (Kemeny *et al.*, 1971, p. 401-406).

Contrairement au cas de la méthode Harary, qui bénéficia d'une diffusion et d'un usage importants (en janvier 2024, *Google Scholar* signalait 282 citations de l'article de 1957), le caractère imparfait et l'efficacité limitée de la méthode Jaulin ont probablement empêché son auteur de la publier. Comme évoqué, la seule description disponible de l'algorithme était confuse. De plus, Jaulin avait lui-même signalé une limite de sa méthode, qui ne pouvait être appliquée aux graphes sans sommets uncliqaux ni sommets réductibles ; Jaulin indiquait que, dans de tels cas, la méthode Harary devait être utilisée (Jaulin, 1961, p. 56). Les carrières contrastées de ces deux algorithmes défailants soulignent l'importance significative des facteurs externes qui déterminent l'adoption et l'utilisation de méthodes formelles.

Conclusion

En combinant une approche socio-historique et une reconstruction expérimentale, cet article présente plusieurs contributions au développement actuel de l'histoire de l'informatique. Il met en lumière les premières applications automatisées de la théorie des graphes en SHS, développées en Europe mais tombées dans l'oubli. Il souligne ce faisant le rôle important que ces disciplines ont joué à cet égard. Les contributions de Gardin et de ses collègues ont été particulièrement innovantes. Cependant, comme l'illustre le cas de l'étude du réseau commercial assyrien, leurs travaux ont fait l'objet d'une réception et d'une reconnaissance extrêmement faibles eu égard leur originalité et leur valeur méthodologique. Deux facteurs possibles de cette sous-évaluation ont été discutés dans cet article, outre les effets des bibliographies sur la découvrabilité et la réutilisation des productions scientifiques :

- les effets des politiques de financement de la recherche, illustrés ici par le transfert de l'effort de recherche sur le traitement de l'information non numérique et les applications de la théorie des graphes depuis des espaces interdisciplinaires temporaires (le CADA et le

52. Flament, 1965, p. 53 ; voir aussi la version anglaise de ce livre, parue précédemment, Flament, 1963, p. 37. Les revues de littérature ultérieures – y compris celles publiées en français – ne mentionnent pas la méthode Jaulin, par exemple Schneider, 1973.

CETIS) vers des organisations disciplinaires liées à l’informatique, alors émergente ;

- les carrières erratiques des algorithmes, dans lesquelles la correction mathématique, l’efficacité computationnelle, les révisions et corrections, ainsi que le statut des auteurs sont des facteurs déterminants, mais imprévisibles.

Les travaux abordés dans cet article constituaient indéniablement des innovations de pointe en leur temps. Leur analyse détaillée montre que leur valeur et leurs effets scientifiques à long terme sont moins contrôlables que ce que supposent parfois les décideurs politiques des sciences, lorsqu’ils sont tentés d’organiser la science comme on gère l’innovation technologique.

Remerciements

Je remercie Pierre-Éric Mounier-Kuhn pour ses conseils et la mise à disposition des archives de l’Institut Blaise Pascal, Solène Chevalier et Olivier Rouault pour leur aide dans l’accès à certains documents, Alain Guénoche pour son soutien dans la compréhension de l’algorithme de Bernard Jaulin, Gary MacDonald pour m’avoir signalé l’ouvrage de John G. Kemeny *et al.*, et Laurent Beauguitte pour sa relecture au bénéfice de la revue ARCS.

Références

- Alba, R. D. (1973). « A Graph-theoretic Definition of a Sociometric Clique ». *Journal of Mathematical Sociology*, 3(1), 113-126. <https://doi.org/10.1080/0022250X.1973.9989826>
- Anderson, A. (2017). *The Old Assyrian Network: An Analysis of the Texts from Kültepe-Kanesh (1950–1750 b.c.e.)* [thèse de doct., Harvard University]. <http://nrs.harvard.edu/urn-3:HUL.InstRepos:39947158>
- Arsac, J. (1988). « Des ordinateurs à l’informatique (1952–1972) ». In P. Chatelin (Éd.), *L’histoire de l’informatique en France* (p. 31-43, T. 1). Institut national polytechnique de Grenoble.
- Augustson, J. G., & Minker, J. (1970). « An Analysis of some Graph Theoretical Cluster Techniques ». *Journal of the ACM*, 17(4), 571-588. <https://doi.org/10.1145/321607.321608>
- Aust, J. (2016). « Des appels d’offres pour la recherche : genèse et transformations d’un dispositif de gouvernement ». In P. Castel, L. Hénaut & E. Marchal (Éd.), *Faire la concurrence : Retour sur un phénomène social et économique* (p. 79-98). Presses des Mines.
- Bamman, D., Anderson, A., & Smith, N. A. (2013). « Inferring Social Rank in an Old Assyrian Trade Network ». *CoRR*, abs/1303.2873. <https://doi.org/10.48550/arxiv.1303.2873>

- Barabási, A.-L. (2002). *Linked: The New Science of Networks*. Perseus Publishing.
- Barjamovic, G., Chaney, T., Coşar, K. A., & Hortaçsu, A. (2017). *Trade, Merchants, and the Lost Cities of the Bronze Age* (Working Paper N° 23992). National Bureau of Economic Research. <https://doi.org/10.3386/w23992>
- Barthélemy, J.-P., & Guénoche, A. (1988). *Les arbres et les représentations des proximités*. Masson.
- Beum, C. O., & Brundage, E. G. (1950). « A Method for Analyzing the Sociomatrix ». *Sociometry*, 13(2), 141-145. <https://doi.org/10.2307/2784940>
- Bonner, R. E. (1964). « On Some Clustering Techniques ». *IBM Journal of Research and Development*, 8(1), 22-32. <https://doi.org/10.1147/rd.81.0022>
- Braffort, P., & Gazzano, A. (1961). « Le calcul électronique à Euratom ». *L'onde électrique*, 16(406), 54-56.
- Carlson, E. R. (1960). « Clique Structure and Member Satisfaction in Groups ». *Sociometry*, 23(4), 327-337. <https://doi.org/10.2307/2785687>
- Coleman, J. S., & MacRae, D. (1960). « Electronic Processing of Sociometric Data for Groups up to 1,000 in Size ». *American Sociological Review*, 25(5), 722-727. <https://doi.org/10.2307/2090146>
- Collinot, A., & Mounier-Kuhn, P.-É. (2010). « Forteresse ou carrefour : l'Institut Blaise-Pascal et la naissance de l'informatique universitaire parisienne ». *Revue pour l'histoire du CNRS*, 28, 79-88.
- Constantine, A. G., Gower, J. C., & Zielman, B. (1993). « A Cappadocian Tablet Problem ». In J. Andresen, T. Madsen & I. Scollar (Éd.), *Computing the Past. Computer Applications and Quantitative Methods in Archaeology* (p. 303-316). Aarhus University Press.
- Cros, R.-C., Gardin, J.-C., & Lévy, F. (1964). *L'automatisation des recherches documentaires : Un modèle général « Le SYNTOL »*. Gauthier-Villars.
- Dallas, C. (2015). « Jean-Claude Gardin on Archaeological Data, Representation and Knowledge: Implications for Digital Archaeology ». *Journal of Archaeological Method and Theory*, 23(1), 305-330. <https://doi.org/10.1007/s10816-015-9241-3>
- De Mol, L., & Bullynck, M. (2008). « A Week-end off: The First Extensive Number-Theoretical Computation on the ENIAC ». In A. Beckmann, C. Dimitracopoulos & B. Löwe (Éd.), *Logic and Theory of Algorithms* (p. 158-167). Springer. https://doi.org/10.1007/978-3-540-69407-6_19
- De Mol, L., & Primiero, G. (2014). « Facing Computing as Technique: Towards a History and Philosophy of Computing ». *Philosophy & Technology*, 27(2), 321-326. <https://doi.org/10.1007/s13347-014-0169-4>

- Dixon, A. H. (1973). *On the Efficiency of Clique Detection in Graphs* [thèse de doct., The University of British Columbia]. <https://doi.org/10.14288/1.0052075>
- Doreian, P. (1969). « A Note on the Detection of Cliques in Valued Graphs ». *Sociometry*, 32(2), 237-242. <https://doi.org/10.2307/2786266>
- Duclert, V. (2004). « La naissance de la Délégation générale à la recherche scientifique et technique. La construction d'un modèle partagé de gouvernement dans les années soixante ». *Revue française d'administration publique*, 112(4), 647-658. <https://doi.org/10.3917/rfap.112.0647>
- Ensmenger, N. (2010). *The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise*. MIT Press.
- Espirat, J.-J., Guiart, J., Lagrange, M.-S., & Renaud, M. (1973). *Système des titres, électifs ou héréditaires dans les Nouvelles-Hébrides centrales, d'Efate aux îles Shepherd*. Institut d'ethnologie.
- Festinger, L. (1949). « The Analysis of Sociograms using Matrix Algebra ». *Human Relations*, 2(2), 153-158. <https://doi.org/10.1177/001872674900200205>
- Flament, C. (1963). *Applications of Graph Theory to Group Structure* (M. Pinard, R. Bretan & F. Fontaine, Trad.). Prentice-Hall Inc.
- Flament, C. (1965). *Théorie des graphes et structures sociales*. Mouton ; Gauthier-Villars.
- Forsyth, E., & Katz, L. (1946). « A Matrix Approach to the Analysis of Sociometric Data: Preliminary Report ». *Sociometry*, 9(4), 340-347. <https://doi.org/10.2307/2785498>
- Freeman, L. C. (1988). « Computers Programs and Social Network Analysis ». *Connections*, 11(2), 26-31.
- Freeman, L. C. (2004). *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press.
- Gardin, J.-C. (1961). « Détermination du réseau économique des Marchands Assyriens en Cappadoce (19e siècle avant J.C.) d'après l'analyse de leurs archives (Exploitation sur ordinateur IBM 650) ». In *Comptes-rendus du Séminaire sur les Modèles Mathématiques dans les Sciences Sociales. Année 1960-1961* (p. 23-26, T. 4). École pratique des hautes études : Groupe de mathématique sociale et de statistique.
- Gardin, J.-C. (1962a). « Cartes perforées et ordinateurs au service de l'archéologie ». *La Nature*, 3331, 449-457.
- Gardin, J.-C. (1962b). « De l'archéologie à l'information automatique ». *Euratom : Bulletin d'Information de la Communauté européenne de l'énergie atomique*, 4, 25-29.
- Gardin, J.-C. (1965a). « A Typology of Computer Uses in Anthropology ». In D. H. Hymes (Éd.), *The Use of Computers in Anthropology* (p. 103-117). Mouton & co.

- Gardin, J.-C. (1965b). « Les travaux de la Section d'Automatique Documentaire (Institut Blaise Pascal, C.N.R.S.), 1961–1965 ». *T.A. informations : Revue internationale du traitement automatique du langage*, 1–2, 9-24.
- Gardin, J.-C. (1965c). « Reconstructing an Economic Network in the Ancient East with a Computer ». In D. H. Hymes (Éd.), *The Use of Computers in Anthropology* (p. 377-391). Mouton & co.
- Gardin, J.-C., & Garelli, P. (1961). « Étude des établissements assyriens en Cappadoce par ordinateur ». *Annales : Économies, sociétés, civilisations*, 16(5), 837-876. <https://doi.org/10.3406/ahess.1961.420758>
- Gardin, J.-C., & Jaulin, B. (Éd.). (1968). *Calcul et formalisation dans les sciences de l'homme*. Éditions du CNRS.
- Gonçalves, C. (2021a). « Social Network Analysis and Kinship in the Old Babylonian Diyala. Fathers and Sons in the Archive of Nūr-Šamaš ». *H2D. Digital Humanities Journal*, 3(1). *A Digital Ancient Near East: Challenges and Opportunities*. <https://doi.org/10.21814/h2d.3470>
- Gonçalves, C. (2021b). « Social Network Analysis, Homonyms, and Aliases in the Old Babylonian Diyala: A Study of the Archive of Nūr-Šamaš ». In C. Gonçalves & C. Michel (Éd.), *Interdisciplinary Research on the Bronze Age Diyala* (p. 83-101). Brepols. <https://doi.org/10.1484/M.SUBART-EB.5.126529>
- Hage, P., & Harary, F. (1983). *Structural Models in Anthropology*. Cambridge University Press.
- Haigh, T., Priestley, M., & Rope, C. (2014). « Los Alamos Bets on ENIAC: Nuclear Monte Carlo Simulations, 1947-1948 ». *IEEE Annals of the History of Computing*, 36(3), 42-63. <https://doi.org/10.1109/MAHC.2014.40>
- Halleux, R., & Xhayet, G. (2007). *La liberté de chercher : Histoire du Fonds national belge de la recherche scientifique*. Les éditions de l'Université de Liège.
- Harary, F. (1970). « Graph Theory as a Structural Model in the Social Sciences ». In B. Harris (Éd.), *Graph Theory and its Applications* (p. 1-16). Academic Press.
- Harary, F., & Ross, I. C. (1957). « A Procedure for Clique Detection using the Group Matrix ». *Sociometry*, 20(3), 205-215. <https://doi.org/10.2307/2785673>
- Hubbell, C. H. (1965). « An Input-output Approach to Clique Identification ». *Sociometry*, 28(4), 377-399. <https://doi.org/10.2307/2785990>
- Jaulin, B. (1961). « Sur une méthode de détermination des cliques dans un graphe symétrique ». In *Comptes-rendus du Séminaire sur les Modèles Mathématiques dans les Sciences Sociales. Année 1960–1961* (p. 50-57, T. 4). École pratique des hautes études : Groupe de mathématique sociale et de statistique.

- Karp, R. M. (1972). « Reducibility Among Combinatorial Problems ». In R. E. Miller & J. W. Thatcher (Éd.), *Complexity of Computer Computations* (p. 85-103). Plenum. https://doi.org/10.1007/978-1-4684-2001-2_9
- Kemeny, J. G., Snell, J. L., & Thompson, G. L. (1971). *Introduction to Finite Mathematics* (2^e éd.). Prentice-Hall.
- Lankford, P. M. (1974). « Comparative Analysis of Clique Identification Methods ». *Sociometry*, 287-305. <https://doi.org/10.2307/2786382>
- Léon, J. (2015). *Histoire de l'automatisation des sciences du langage*. ENS Éditions.
- Lewis, T. G. (2009). *Network Science: Theory and Applications*. John Wiley & Sons.
- Luce, R. D., & Perry, A. D. (1949). « A Method of Matrix Analysis of Group Structure ». *Psychometrika*, 14(2), 95-116. <https://doi.org/10.1007/bf02289146>
- MacRae, D. (1960). « Direct Factor Analysis of Sociometric Data ». *Sociometry*, 23(4), 360-371. <https://doi.org/10.2307/2785690>
- Moon, J. W., & Moser, L. (1965). « On Cliques in Graphs ». *Israel Journal of Mathematics*, 3(1), 23-28. <https://doi.org/10.1007/BF02760024>
- Moreno, J. L. (1934). *Who Shall Survive? A New Approach to the Problem of Human Interrelations*. Nervous & mental disease publishing co.
- Mounier-Kuhn, P.-É. (2010). *L'informatique en France de la seconde Guerre Mondiale au Plan Calcul : L'émergence d'une science* (2^e éd.). Presses de l'Université Paris-Sorbonne.
- Mounier-Kuhn, P.-É. (2012). « Computer Science in French Universities : Early Entrants and Latecomers ». *Information & Culture : A Journal of History*, 47(4), 414-456. <https://doi.org/10.1353/lac.2012.0025>
- Mounier-Kuhn, P.-É. (2014). « Algol in France: From Universal Project to Embedded Culture ». *IEEE Annals of the History of Computing*, 36(4), 6-25.
- National Science Foundation (U.S.) (Éd.). (1959, mars). *Science Information Notes*, 1(1), 121.
- Newman, M. E. J., & Girvan, M. (2004). « Finding and Evaluating Community Structure in Networks ». *Physical Review E*, 69, 026113. <https://doi.org/10.1103/PhysRevE.69.026113>
- Nofre, D. (2014). « Managing the Technological Edge: The UNESCO International Computation Centre and the Limits to the Transfer of Computer Technology, 1946-61 ». *Annals of Science*, 71(3), 410-431. <https://doi.org/10.1080/00033790.2013.827075>
- Pair, C. (1990). « CRIN: The History of a Laboratory ». *Annals of the History of Computing*, 12(3), 159-166. <https://doi.org/10.1109/MAHC.1990.10033>
- Pardalos, P. M., & Xue, J. (1994). « The Maximum Clique Problem ». *Journal of Global Optimization*, 4, 301-328. <https://doi.org/10.1007/BF01098364>

- Paull, M. C., & Unger, S. H. (1959). « Minimizing the Number of States in Incompletely Specified Sequential Switching Functions ». *IRE Transactions on Electronic Computers*, 3, 356-367. <https://doi.org/10.1109/TEC.1959.5222697>
- Plutniak, S. (2017). « Une contribution archéologique à la théorie des sciences sociales est-elle possible? Aspects de la controverse entre Jean-Claude Passeron et Jean-Claude Gardin » (P. Boissinot, Éd.). *Palethnologie*, 9. *Archéologie et sciences sociales*, 7-21. <https://doi.org/10.4000/palethnologie.279>
- Plutniak, S. (2018). « Aux prémices des humanités numériques? La première analyse automatisée d'un réseau économique ancien (Gardin & Garelli, 1961). Réalisation, conceptualisation, réception » (L. Beauguitte & C. Lemerrier, Éd.). *ARCS : Analyse de réseaux pour les sciences sociales*, 1. *Concepts et méthodes en analyse de réseaux : circulations et barrières disciplinaires*. <https://doi.org/10.46298/arcs.9236>
- Plutniak, S. (2019). « À distances raisonnables des structuralismes : logique, langage, formalisation et sciences de l'homme. Une dispute du 20e siècle finissant ». *Zilsel : Science, technique, société*, 6, 70-115. <https://doi.org/10.3917/zil.006.0070>
- Rouault, O. (1984). « Éléments pour un logiciel assyriologique » (G. Buccellati & O. Rouault, Éd.). *Computer Aided Research in Near Eastern Studies*, 1(2), 1-81.
- Roubaud, J. (2008). *Impératif catégorique : récit*. Éditions du Seuil.
- Schneider, M. (1973). « Méthodes pour recenser toutes les cliques maximales et q-maximales d'un graphe ». *Revue française d'automatique, informatique, recherche opérationnelle. Recherche opérationnelle*, 7(V3), 21-33.
- Waerzeggers, C. (2014). « Social Network Analysis of Cuneiform Archives – a New Approach ». In H. D. Baker & M. Jursa (Éd.), *Documentary Sources in Ancient Near Eastern and Greco-Roman Economic History: Methodology and Practice* (p. 207-233). Oxbow Books. <https://doi.org/10.2307/j.ctvh1dn9m.14>
- Wagner, A., Levavi, Y., Kedar, S., Cohen, Y., & Zadok, R. (2013). « Quantitative Social Network Analysis (SNA) and the Study of Cuneiform Archives: A Test-Case Based on the Murašū Archive ». *Akkadica*, 132(2), 117-134.
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- Wu, Q., & Hao, J.-K. (2015). « A Review on Algorithms for Maximum Clique Problems ». *European Journal of Operational Research*, 242(3), 693-709. <https://doi.org/10.1016/j.ejor.2014.09.064>

Annexe A : Jaulin, B. 1961, Sur une méthode de détermination des cliques dans un graphe symétrique

Nous nous proposons de décrire une méthode permettant de déterminer toutes les cliques d'un graphe fini et non orienté. Rappelons qu'un graphe est un ensemble Y , ensemble de marchands*, etc..., et une application de Y dans Y , les relations commerciales entre ces marchands par exemple. Le graphe est fini c'est-à-dire que l'ensemble Y est fini, et non orienté, c'est-à-dire que la relation est symétrique, si x traite avec y , y traite avec x . Si nous avons distingué la vente de l'achat dans les transactions entre marchands le graphe eût alors été orienté.

On notera Γx les éléments qui sont en relation avec x . Une clique est un ensemble $C \subset Y$ tel que si $x \in C$ et $y \in C$ on a $y \in \Gamma x$ quels que soient y et x dans C .

On appellera élément uniclinal et élément pluriclinal ceux des éléments qui appartiennent à une et plusieurs cliques respectivement.

Différentes méthodes ont été proposées pour la détermination des cliques d'un graphe. BERGE dans (1) ramène le problème à celui plus général de la recherche d'un ensemble stable intérieurement maximum, HAVARY [sic] dans (2) utilisant un résultat donné par FESTINGER permettant d'identifier les éléments uniclinaux décrit un algorithme conduisant à toutes les cliques du graphe. Le procédé que nous décrirons diffère de celui d'HAVARY [sic] par la façon de déterminer les points uniclinaux et les cliques correspondantes et aussi du fait qu'il permet de résoudre directement les cliques passant par certains points pluriclinaux d'un type particulier.

Soit X la matrice symétrique associée au graphe

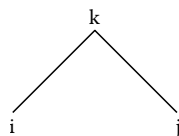
$$X = (x_{ij})$$

$x_{ij} = 0$ si l'élément i n'est pas en relation avec l'élément j

$x_{ij} = 1$ si i est en relation avec j

et $x_{ii} = 0$

Nous dirons qu'un sommet k du graphe est intermédiaire entre i et j si $x_{ik} = x_{jk} = 1$. Soit si l'on a la représentation



Soit a_{ij} les éléments de la matrice X^2

$$a_{ij} = \sum_{k=1}^n x_{ik}x_{kj}$$

*. Voir compte-rendu Fascicule I : M. Gardin. NDE : Jaulin fait référence à Gardin, 1961.

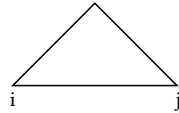
d'après la définition précédente a_{ij} représente le nombre d'intermédiaires entre i et j .

Appelons $b_{ij} = a_{ij} \cdot x_{ij}$ et B la matrice des b_{ij} .

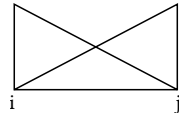
Alors $b_{ij} = 0$ si i n'est pas relié à j

et $b_{ij} = a_{ij}$ si i est relié à j

ainsi si $b_{ij} = 1$ on a



si $b_{ij} = 2$ on a



Détermination des points uncliqualux et des points réductibles

Appelons $r(i)$ le nombre de b_{ij} différents de zéro dans la ligne i de la matrice B et $b_i^1 = \sup(b_{ij})$

$$j = 1 \text{ à } n$$

remarquons que b_i^1 ne peut pas être supérieur à $r(i)$, il existerait sinon des éléments k intermédiaires à i et à j où j est défini par $b_{ij} = b_i^1$ et pour lesquels $b_{ik} = 0$ ce qui est impossible car k étant intermédiaire entre i et j , j est intermédiaire entre i et k donc $b_{ik} \neq 0$. On a donc $b_i^1 \leq r(i) - 1$.

Distinguons deux cas :

1°) $b_i^1 = r(i) - 1$

a) Points uncliqualux

Propriété 1 Pour que l'élément i appartienne à une seule clique d'ordre p il faut et il suffit que l'on ait quel que soit j , $b_{ij} = b_i^1 = p - 1$

En effet l'un quelconque des éléments j en relation avec l'élément i a $p - 1$ intermédiaires avec i , et comme $r(i) = b_i^1 + 1 = p$ représente le nombre d'éléments pouvant être en clique avec i , l'élément j est relié à tous les éléments reliés à i et ceci quel que soit j . Ces éléments forment donc une seule clique passant par i .

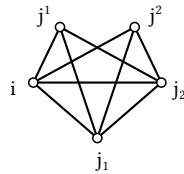
b) Points réductibles

Appelons j_1, j_2, \dots, j_k les éléments pour lesquels $b_{ij_1} = \dots = b_{ij_k} = b_i^1$

Propriété 2 Les éléments j_1, j_2, \dots, j_k font partie de toutes les cliques passant par i . Le raisonnement est analogue au précédent, le couple (i, j_i) ayant $b_i^1 = r(i) - 1$ intermédiaires, j_i est relié à tous les éléments pouvant être en clique avec i de même pour j_2, \dots, j_k .

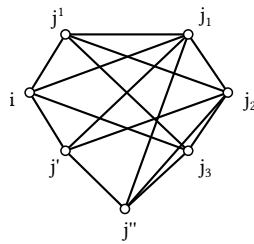
Nous appellerons (j_1, \dots, j_k) le noyau des cliques passant par i . Les individus qui sont en relation avec i le sont avec j_1, \dots, j_k mais la proposition inverse est fautive. Les éléments liés à l'un des j_1, \dots, j_k ne sont pas tous en relation avec i .

Conséquence 1 Les b_{ij} différents de zéro autres que b_i^1 sont au moins égaux à k , ceux pour lesquels $b_{ij} = k$ forment avec i une clique à $k + 2$ éléments formés de i, j, j_1, \dots, j_k . Ainsi l'on résout les éléments i tels que ceux indiqués sur la figure où le nombre des j^i et des j_i n'est pas limité.



Conséquence 2 Si par le procédé précédent on a trouvé une ou plusieurs cliques i, j^1, j_1, \dots, j_k on déduira de la ligne i une nouvelle ligne i' de la façon suivante : les éléments $b_{i'j_1}, \dots, b_{i'j_k}$ correspondant à j_1, j_2, \dots, j_k et $b_{i'j^1}$ seront égaux à zéro, les autres éléments $b_{i'j}$ seront égaux à $b_{ij} - k$. Sur cette nouvelle ligne on en essaiera l'analyse en utilisant soit la propriété 1 soit la conséquence 2, si elle aboutit à une clique il faudra lui adjoindre les éléments j_1, j_2, \dots, j_k .

On résoudra en utilisant cette conséquence des éléments i tels que ceux indiqués sur la figure, comme précédemment le nombre des j_i et des j^i n'est pas limité.



Les éléments i pouvant être résolus entièrement en utilisant la conséquence 1 ou 2 nous les appellerons des points réductibles. Lorsqu'un point est résolu entièrement la ligne i de la matrice B est nulle. Si un point a été résolu partiellement il faut en vue d'une itération laisser inchangée la ligne correspondante de cette matrice. On peut montrer qu'il existe 2^{k-1} types différents de points réductibles pour $b_i^1 = r(i) - 1 = k$, mais le nombre

de configurations possibles est beaucoup plus grand. À titre d'exemple si $b_i^1 = 4$, il existe 10 configurations possibles dont 8 sont réductibles à $b_i^1 = 5$, seulement 16 sont réductibles sur 32 environ.

2°) $b_i^1 < r(i) - 1$

Il existe malheureusement des graphes tels que pour chacun de leurs éléments on a cette inégalité : pas de points uniclinaux ni de points réductibles. Dans ce cas on peut opérer d'une façon analogue à celle indiquée par HAVARY [sic] dans (2) et considérer une partie du graphe total formée des éléments pouvant entrer en clique avec un élément i quelconque, celui pour lequel, par exemple, $r(i)$ est maximum. Résoudre cette partie du graphe c'est déterminer toutes les cliques passant par i . La matrice X_i correspondant à cet ensemble est obtenue en enlevant les lignes et les colonnes correspondant aux éléments j n'entrant pas en clique avec i . Remarquons que dans la matrice B_i déduite de X_i pour tous les éléments autres que i on a l'égalité $b_i^1 = r(i) - 1$. Cela ne veut pas dire pour autant que tous les points sont réductibles, il se peut même qu'aucun ne le soit.

Les différentes étapes du procédé

a) on calcule B et l'on évalue $r(i)$ et b_i^1 pour chaque ligne de la matrice. Si $b_i^1 < r(i) - 1$ pour tous les points on passe à l'étape c. Si $b_i^1 = r(i) - 1$ pour certaines lignes, on utilise les 2 critères relatifs aux points uniclinaux et réductibles pour résoudre cette ligne. Une ligne complètement résolue est une ligne nulle après analyse.

b) On calcule X' déduite de x en enlevant les lignes et les colonnes relatives aux points résolus au cours de l'étape précédente et on en revient en a où X est remplacé par X' . L'itération est arrêtée quand X est devenu nul.

c) On recherche un des points pour lequel $r(i)$ est le plus grand. On calcule X_i obtenu en enlevant cette fois à la matrice X les lignes et les colonnes correspondant aux éléments pour lesquels $b_{ij} = 0$ dans la ligne i de la matrice B et on retourne en a après avoir considéré à nouveau le graphe total itérant la chaîne avec une matrice X' déduite de X en enlevant la ligne et la colonne correspondant au point i .

2 juin 1961

Bibliographie

1. BERGE : Théorie des graphes et ses applications.
2. HAVARY [sic] : A procedure for clique detection using the group matrix.

3. FESTINGER : The analysis of sociograms using matrix algebra.

Annexe B : Code de calcul

Chargement des paquets et des données

Quatre paquets sont installés et chargés afin de disposer d'implémentations des algorithmes listés dans le Tableau 2 :

TABLEAU 2 – Algorithmes de détection de cliques utilisés et références de leur implémentation dans des paquets R.

| algorithme | fonction | package | référence |
|-------------------------------|-----------------|---------|----------------------------|
| Bron et Kerbosch, 1973 | maxClique() | RBGL | Carey <i>et al.</i> , 2020 |
| Makino et Uno, 2004 | clique.census() | sna | Butts, 2020 |
| Östergård, 2001 | qpGetCliques() | qpgraph | Castelo et Roverato, 2009 |
| Eppstein <i>et al.</i> , 2010 | max_cliques() | igraph | Csárdi et Nepusz, 2006 |

```
if (! requireNamespace("igraph", quietly = TRUE)){
  install.packages("igraph")
}
if (! requireNamespace("sna", quietly = TRUE)){
  install.packages("sna")
}
if (! requireNamespace("RBGL", quietly = TRUE)){
  install.packages("BiocManager")
  BiocManager::install("RBGL")
}
if (! requireNamespace("qpgraph", quietly = TRUE)){
  BiocManager::install("qpgraph")
}
library(igraph)
library(sna)
library(RBGL)
library(qpgraph)
```

Implémentation de l'algorithme d'Harary et Ross

Le code suivant constitue notre proposition d'implémentation de la méthode de détection de cliques décrite dans Harary et Ross, 1957. Ce code est également disponible en ligne (doi : [10.5281/zenodo.3932104](https://doi.org/10.5281/zenodo.3932104)).

```
.getC <- function(M, unicliquals, Mrowsums){
  Cgroup.list <- list()
  CgroupIndex.list <- list()
  CgroupPrime.list <- list()
```

```

for(i in 1:nrow(M) ){
  # skip if the point already belongs to a clique:
  if(i %in% unlist(CgroupIndex.list)) next
  # skip if no uncliquial points:
  if(! i %in% unicliquals ) next
  CgroupIndex <- sort(c(i, which(M[, i] != 0)))
  Cgroup <- sort(colnames(M)[c(i, which(M[, i] != 0))])
  CgroupPrime <- which( Mrowsums == Mrowsums[i] )
  CgroupPrime <- CgroupPrime[CgroupPrime %in% CgroupIndex]
  # add to results
  Cgroup.list <- append(Cgroup.list, list(Cgroup))
  CgroupIndex.list <- append(CgroupIndex.list, list(CgroupIndex))
  CgroupPrime.list <- append(CgroupPrime.list, list(CgroupPrime))
}
list(Cgroup.list = Cgroup.list,
     CgroupPrime.list = CgroupPrime.list)
}

.substract.matrix <- function(mat, CgroupPrime.list){
  i <- 1:nrow(mat)
  i <- i[ ! 1:nrow(mat) %in% unlist(CgroupPrime.list) ]
  mat[i, i]
}

.extract.cliques <- function(mat){
  cliques.list <- list()
  M <- mat * mat %*% t(mat)
  np <- apply(mat, 1, sum) # similar to vertices' degrees
  Mrowsums <- rowSums(M)
  unicliquals <- which( Mrowsums == np * (np - 1) )

  if(length(unicliquals) > 0){
    res <- .getC(M, unicliquals, Mrowsums)
    cliques <- res$Cgroup.list
    # get only cliques with at least 3 vertices:
    cliques <- cliques[sapply(cliques, function(x) length(x) > 2 )]
    cliques.list <- append(cliques.list, cliques)
    mat <- .substract.matrix(mat, res$CgroupPrime.list)
  }
  else{
    CgroupIndex <- sort(c(1, which(M[, 1] != 0)))
    Cgroup <- c(1, which(mat[, 1] != 0))
    submat1 <- mat[Cgroup, Cgroup]
    submat2 <- mat[-1, -1]
    mat <- list(submat1, submat2)
  }
  list(mat, cliques.list)
}

haross.cliques <- function(mat){
  # initial tests:
  if( ! is.matrix(mat) ) stop("The argument is not a matrix.")
  if( ncol(mat) != nrow(mat) ) stop("A square matrix is required.")
  if( is.null(colnames(mat)) & is.null(rownames(mat)) ){

```

```

colnames(mat) <- 1:ncol(mat)
rownames(mat) <- 1:nrow(mat)
}
# set variables:
cliques.list.final <- list()
mat.list <- list(mat)

repeat{ # repeat while the sum of the matrix values > 0
  # run the main function:
  res <- lapply(mat.list, .extract.cliques)
  # sort results:
  # 1) extract and add the cliques to the list:
  cliques.list.final <- append(cliques.list.final,
                              lapply(res, function(x) x[[2]]))
  )
  # 2) extract the list of matrices:
  mat.list <- lapply(res, function(x) x[[1]] )
  # if the list is too nested, unnest:
  if( is.list(mat.list[[1]]) & length(mat.list[[1]]) > 1 ) {
    mat.list <- unlist(mat.list, recursive = F)
  }
  # keep only the matrices with more than 2 points:
  mat.list <- mat.list[ sapply(mat.list, function(x) sum(x) > 2 ) ]
  # if there is no more matrices, break
  if( length(mat.list) == 0 ) break
}

unlist(cliques.list.final, recursive = F)
}

```

Données

Le graphe présenté en Figure 10 dans Gardin et Garelli, 1961, p. 868, permet de reconstituer un graphe et la liste des cliques identifiées par les auteurs.

```

# Reproduction du tableau tiré de Gardin et Garelli 1961, fig. 10 p. 868:
GardinGarelli.df <- as.matrix(rbind(
  c("amur-ishtar", "laqipun"),
  c("hina", "amur-ishtar"),
  c("hina", "im(i)d-ilum"),
  c("hina", "laqipun"),
  c("im(i)d-ilum", "amur-ishtar"),
  c("im(i)d-ilum", "laqipun"),
  c("pushu-kin", "hina"),
  c("pushu-kin", "laqipun"),
  c("amur-ishtar", "assur-nada"),
  c("assur-imitti", "amur-ishtar"),
  c("assur-imitti", "assur-nada"),
  c("assur-taklaku", "amur-ishtar"),
  c("assur-taklaku", "assur-nada"),
  c("assur-taklaku", "pushu-kin"),

```

```

c("pushu-kin", "assur-nada"),
c("assur-tab", "im(i)d-ilum"),
c("assur-tab", "pushu-kin"),
c("buzazu", "enna-sin"),
c("buzazu", "shu-belim"),
c("mannum-balum-assur", "enna-sin"),
c("mannum-balum-assur", "shu-belim"),
c("amur-ishtar", "pushu-kin"),
c("im(i)d-ilum", "pushu-kin"),
c("enna-sin", "shu-belim")
))
colnames(GardinGarelli.df) <- c("from", "to")

# cliques indiquées par Gardin et Garelli 1961:
GardinGarelli.res <- list(
  c("amur-ishtar", "pushu-kin", "hina", "im(i)d-ilum", "laqipun"),
  c("amur-ishtar", "pushu-kin", "assur-nada", "assur-taklaku"),
  c("amur-ishtar", "assur-imitti", "assur-nada"),
  c("pushu-kin", "assur-tab", "im(i)d-ilum"),
  c("buzazu", "enna-sin", "shu-belim"),
  c("enna-sin", "mannum-balum-assur", "shu-belim")
)

```

Ce graphe est converti en différents formats.

```

# matrix:
GardinGarelli.g <- igraph::graph_from_data_frame(GardinGarelli.df, directed=F)
GardinGarelli.df <- igraph::as_adjacency_matrix(GardinGarelli.g, sparse=F)
# graphNEL:
GardinGarelli.nel <- igraph::igraph.to.graphNEL(GardinGarelli.g)
# sna network:
GardinGarelli.net <- network::network(GardinGarelli.df, directed=F)

```

Exécution des détections de cliques

Les cinq algorithmes de détection de cliques sont appliqués sur le graphe.

```

# Harary and Ross 1957
harary.res <- haross.cliques(GardinGarelli.df)
harary.res <- harary.res[ order(sapply(harary.res, length), decreasing=T) ]
# Bron and Kerbosch 1973 (RBGL)
bron.res <- RBGL::maxClique(GardinGarelli.nel)$maxCliques
# Makino and Uno 2004 (sna)
makino.res <- sna::clique.census(GardinGarelli.net, mode="graph",
                                tabulate.by.vertex=F)$cliques
makino.res <- lapply(makino.res, function(x)
  lapply(x, function(y) network.vertex.names(GardinGarelli.net)[y]) )
makino.res <- unlist(makino.res, recursive = F)
# Ostergard 2001 (qpgraph)
ostergard.res <- qpgraph::qpGetCliques(GardinGarelli.df)

```



```
## 1/13 (max 1) 0.00 s (0.00 s/round)
## 13/13 (max 1) 0.00 s (0.00 s/round)

ostergard.res <- lapply(ostergard.res,
  function(i) rownames(GardinGarelli.df)[i])
# Eppstein et al. 2010 (igraph)
eppstein.res <- igraph::maxCliques(GardinGarelli.g, min=3)
eppstein.res <- lapply(eppstein.res, names)
```

Comparaison des résultats

Les cliques détectées par les différents algorithmes sont identifiées en triant et concaténant les noms des sommets qu'elles comportent.

```
res.list <- list(
  "Gardin 1961" = GardinGarelli.res,
  "Harary 1957" = harary.res,
  "Bron 1973" = bron.res,
  "Makino 2004" = makino.res,
  "Osertgard 2001" = ostergard.res,
  "Eppstein 2010" = eppstein.res
)
# tri et concaténation des noms de sommets:
res.list <- lapply(res.list, function(x) lapply(x, sort))
res.list <- lapply(res.list, function(x) lapply(x, paste, collapse="/"))
```

Ces différents résultats sont comparés à ceux obtenus à l'aide de la méthode d'Harary et Ross (Tableau 1).

```
res.tab <- lapply(res.list, function(x) res.list[[2]] %in% x)
res.tab <- do.call("rbind", res.tab)
res.tab <- t(res.tab)
res.tab <- cbind(id = c(1:6, NA, NA),
  size = sapply(harary.res, length),
  res.tab)
```

Informations de session R

- R version 4.0.3 (2020-10-10), x86_64-pc-linux-gnu
- Running under : Arch Linux
- Matrix products : default
- BLAS : /usr/lib/libblas.so.3.9.0
- LAPACK : /usr/lib/liblapack.so.3.9.0
- Base packages : base, datasets, graphics, grDevices, methods, parallel, stats, utils
- Other packages : BiocGenerics 0.34.0, graph 1.66.0, igraph 1.2.6, knitr 1.30, network 1.16.1, qgraph 2.22.0, RBGL 1.64.0, sna 2.6, statnet.common 4.4.1, xtable 1.8-4

— Loaded via a namespace (and not attached) : annotate 1.66.0, AnnotationDbi 1.50.3, askpass 1.1, assertthat 0.2.1, Biobase 2.48.0, BiocFileCache 1.12.1, BiocParallel 1.22.0, biomaRt 2.44.4, Biostrings 2.56.0, bit 4.0.4, bit64 4.0.5, bitops 1.0-6, blob 1.2.1, coda 0.19-4, codetools 0.2-16, compiler 4.0.3, crayon 1.3.4, curl 4.3, DBI 1.1.0, dbplyr 2.0.0, DelayedArray 0.14.1, digest 0.6.27, dplyr 1.0.2, ellipsis 0.3.1, evaluate 0.14, generics 0.1.0, GenomeInfoDb 1.24.2, GenomeInfoDbData 1.2.3, GenomicAlignments 1.24.0, GenomicFeatures 1.40.1, GenomicRanges 1.40.0, glue 1.4.2, grid 4.0.3, highr 0.8, hms 0.5.3, httr 1.4.2, IRanges 2.22.2, lattice 0.20-41, lifecycle 0.2.0, magrittr 2.0.1, Matrix 1.2-18, matrixStats 0.57.0, memoise 1.1.0, mvtnorm 1.1-1, openssl 1.4.3, pillar 1.4.7, pkgconfig 2.0.3, prettyunits 1.1.1, progress 1.2.2, purrr 0.3.4, qtl 1.47-9, R6 2.5.0, rappdirs 0.3.1, Rcpp 1.0.5, RCurl 1.98-1.2, Rgraphviz 2.32.0, rlang 0.4.10, rle 0.9.2, Rsamtools 2.4.0, RSQLite 2.2.1, rtracklayer 1.48.0, S4Vectors 0.26.1, stats4 4.0.3, stringi 1.5.3, stringr 1.4.0, SummarizedExperiment 1.18.2, tibble 3.0.4, tidyselect 1.1.0, tools 4.0.3, vctrs 0.3.6, xfun 0.20, XML 3.99-0.5, xml2 1.3.2, XVector 0.28.0, zlibbioc 1.34.0

Références

- Bron, C., & Kerbosch, J. (1973). « Finding all Cliques of an Undirected Graph (algorithm 457) ». *Communications of the ACM*, 16(9), 575-576. <https://doi.org/10.1145/362342.362367>
- Butts, C. T. (2020). *sna: Tools for Social Network Analysis* [R package version 2.6]. <https://CRAN.R-project.org/package=sna>
- Carey, V., Long, L., & Gentleman, R. (2020). *RBGL: An interface to the BOOST graph library* [R package version 1.64.0]. <https://doi.org/10.18129/B9.bioc.RBGL>
- Castelo, R., & Roverato, A. (2009). « Reverse Engineering Molecular Regulatory Networks from Microarray data with qp-graphs ». *Journal of Computational Biology*, 16(2), 213-27. <https://doi.org/10.1089/cmb.2008.08TT>
- Csárdi, G., & Nepusz, T. (2006). « The igraph Software Package for Complex Network Research ». *InterJournal*, 1695(5), 1-9. <http://igraph.org>
- Eppstein, D., Löffler, M., & Strash, D. (2010). « Listing all Maximal Cliques in Sparse Graphs in Near-optimal Time ». *International Symposium on Algorithms and Computation*, 403-414. https://doi.org/10.1007/978-3-642-17517-6_36
- Gardin, J.-C., & Garelli, P. (1961). « Étude des établissements assyriens en Cappadoce par ordinateur ». *Annales : Économies, sociétés, civilisations*, 16(5), 837-876. <https://doi.org/10.3406/ahess.1961.420758>
- Harary, F., & Ross, I. C. (1957). « A Procedure for Clique Detection using the Group Matrix ». *Sociometry*, 20(3), 205-215. <https://doi.org/10.2307/2785673>
- Makino, K., & Uno, T. (2004). « New Algorithms for Enumerating All Maximal Cliques ». In T. Hagerup & J. Katajainen (Éd.), *Algorithm Theory*

– *SWAT 2004* (p. 260-272). Springer-Verlag. https://doi.org/10.1007/978-3-540-27810-8_23

Östergård, P. R. J. (2001). « A New Algorithm for the Maximum-weight Clique Problem ». *Nordic Journal of Computing*, 8, 424-436. <https://doi.org/10.5555/766502.766504>