

## Thème 3 : Applications à la sociologie

### Chapitre 7 : Analyse statistique implicative des transitions professionnelles dans la Genève du 19e siècle

Gilbert Ritschard, Matthias Studer et Michel Oris

Laboratoire de démographie et d'études familiales, Université de Genève  
gilbert.ritschard@unige.ch, matthias.studer@unige.ch, michel.oris@unige.ch,  
<http://www.unige.ch/ses/demog>

**Résumé.** Cet article reprend l'analyse statistique implicative de la dynamique socioprofessionnelle dans la première moitié du 19e à Genève que nous avons présentée aux rencontres ASI4 (Oris et Ritschard, 2007) et la compare à une analyse supervisée des dissimilarités entre transitions. Les données considérées résultent de l'appariement deux à deux de 6 recensements. Plus précisément, nous considérons le groupe socioprofessionnel (GSP) des individus retenus et son changement entre deux recensements successifs. Nous nous intéressons aux types de transition (stable, devenir actif, cesser l'activité, ...) ainsi qu'aux nouveaux venus (immigrés et naissances) et disparus (émigrés et décédés). L'analyse statistique implicative donne une vision synthétique des liens entre ces dynamiques et les GSP concernés, ainsi qu'avec un certain nombre de variables démographiques et culturelles (sexe, âge, état-civil, religion). Elle met en lumière notamment des polarisations autour de variables clé. L'analyse des dissimilarités permet quant à elle de segmenter la population en groupes homogènes en fonction des caractéristiques démographiques et culturelles. Le recours à l'intensité d'implication pour identifier les transitions typiques des groupes ainsi obtenus s'avère une précieuse aide à l'interprétation et donne les éléments nécessaires à la comparaison avec les résultats du graphe implicatif.

## 1 Introduction

Nous reprenons et étendons ici le travail (Oris et Ritschard, 2007) présenté aux 4èmes rencontres d'Analyse statistique implicative, ASI4, où nous rapportons une expérience d'analyse statistique implicative dans un contexte de démographie historique. L'étude porte sur le recrutement et le renouvellement des groupes socioprofessionnels à Genève (Oris et al., 2006) entre 1816 et 1843. Plus précisément, nous nous intéressons aux changements individuels de statut socioprofessionnel que l'on observe entre deux recensements successifs, l'objectif étant de mieux comprendre comment ces changements ou transitions sont liés aux caractéristiques démographiques des individus ainsi que leur impact sur la démographie de la Cité de Genève.

Dans Oris et Ritschard (2007), nous avons montré comment l'analyse statistique implicative, ASI, vient enrichir les connaissances obtenues par des méthodes démographiques et d'analyse de données classiques, l'intérêt de l'ASI résidant essentiellement dans sa force synthétique. Elle donne en effet une vision globale aisément compréhensible d'un ensemble d'associations et permet ainsi, par rapport notamment aux approches qui se focalisent sur l'explication d'une caractéristique particulière, de mettre en lumière des polarisations tout à fait éclairantes. Nous complétons ici cette comparaison en proposant une analyse originale des dissimilarités entre les couples d'états d'origine et de destination qui caractérisent les transitions. Cette étude exploite les principes de l'analyse de variance pour générer un arbre de segmentation des transitions observées en fonction de caractéristiques démographiques (âge, sexe, état-civil) et culturelles (religion). Elle n'est pas totalement déconnectée de la statistique implicative, puisque nous exploitons l'intensité d'implication pour identifier les transitions caractéristiques dans chaque nœud de l'arbre.

L'article est organisé comme suit. Nous commençons à la section 2 par rappeler brièvement le contexte historique de la Cité de Genève dans cette première moitié du 19e. La section 3 décrit ensuite l'origine et l'organisation des données utilisées. L'analyse statistique implicative fait l'objet de la section 4 tandis que la section 5 présente l'analyse des dissimilarités des transitions et démontre l'intérêt de l'intensité d'implication comme aide à l'interprétation des résultats obtenus. Enfin, nous concluons à la section 6 en soulignant les apports respectifs de l'ASI et de l'analyse des dissimilarités pour notre problématique, mais aussi quelques-unes de leurs limites.

## 2 Le contexte historique genevois

La Réforme protestante a fait de Genève la "Rome calviniste" qui, plus de deux siècles durant, a vécu telle une forteresse menacée au milieu d'un environnement essentiellement catholique. En 1816, au terme d'une annexion à la France et suite au Traité de Vienne de 1815, la république urbaine devient un canton doublement mixte grâce à la fusion avec des municipalités rurales catholiques qui appartenaient auparavant à la France ou au royaume de Piémont-Sardaigne. A Genève intra-muros, la proportion de catholiques au sein de la population passe de 11% en 1816 à 28% en 1843, puis 46,4% en 1900. Cela a profondément marqué les esprits des "vieux Genevois" qui percevaient leur identité menacée par l'afflux d'immigrés transformant subitement en un ensemble multiculturel un bloc monolithique séculaire (Herrmann, 2003). Le choc a été d'autant plus rude que l'essentiel de la croissance démographique a été assurée par le solde migratoire durant toute la première moitié du 19e siècle. Déjà au 18e, les pratiques de contrôle des naissances s'étaient diffusées des élites aux masses populaires. Les couples genevois mariés entre 1800 et 1850 n'ont eu en moyenne (Schumacher, 2002) que 2,32 enfants ! Parallèlement, les freins traditionnels restent serrés au maximum, avec un âge moyen des femmes au 1er mariage de 28 ans et un célibat définitif de presque 20%. Genève est à cette époque probablement le seul endroit au monde où cumulent leurs effets au sein du même régime démographique et le malthusianisme traditionnel et le néo-malthusianisme moderne. Dès lors, bien que la mortalité infantile ait chuté de 200 à 100/130 ‰ entre la deuxième moitié du 18e et la première moitié du 19e siècle, la balance naturelle présente un excès des naissances sur les décès qui n'atteint que 557 unités entre 1806 et 1850. La croissance de la population, qui passe de 21327 à 31200 au cours de cette période, est due à plus de 90% à

l'excédent des immigrations sur les émigrations. Les structures de la population genevoise ont donc été bien plus transformées par les migrations que ne le suggérerait une expansion somme toute modeste, dans le contexte global d'explosion urbaine qui a caractérisé le 19<sup>e</sup> siècle.

Il est vrai que l'économie ne pouvait guère soutenir une progression plus marquée. La ville s'est dotée de la structure d'activités que lui imposait son insularité religieuse, son absence de domination sur l'arrière-pays catholique. En particulier, elle a développé la Fabrique, qui "est l'ensemble des arts et des artistes qui concourent à la création des montres et des bijoux. Le travail en est réparti en une multitude de petits ateliers" (Babel, 1942b, p 13). Après une période faste entre 1750 et 1792, la ville a beaucoup souffert durant la période française (Guichonnet, 1974, p 273). Le marasme économique a duré jusqu'en 1819, voire 1822 (Babel, 1942a, p 44). Le textile (fabrication d'indiennes) n'ayant pas su résister au retour de la concurrence anglaise après 1816, la Fabrique s'est encore plus imposée comme l'activité dominante. Elle emploie 5000 personnes en 1828 (Guichonnet, 1974, p 287) ; 35,4% des hommes qui se marient à Genève entre 1822 et 1845 travaillent dans ce secteur. Cette domination est dangereuse car industrie d'exportation par excellence, l'horlogerie est sensible aux fluctuations politico-guerrières qui affectent ses marchés extérieurs. Ainsi, si la période qui s'étend entre 1830 et 1845 fut prospère pour la Fabrique et l'ensemble de l'économie urbaine (Guichonnet, 1974, p 288), la crise qui a traversé le continent (1845-1847) sera violemment ressentie. Avec l'arrivée des radicaux au pouvoir en 1846, la ville connaît de profonds changements dont la destruction des murailles et la modernisation, en particulier le développement d'industries performantes (Perroux, 2006).

La période que nous étudions est donc à la croisée des temps. Entre 1816 et 1843, Genève reste une ville tranquille dominée par un "conservatisme éclairé" (Dufour, 1997, p 99), soucieux de préserver le modèle social ancien alors même que les flux migratoires commencent à transformer les structures sociales en profondeur.

### 3 Les données

Les données étudiées ont été collectées dans le cadre d'une recherche visant à reconstruire les parcours de vie dans la cité de Genève au 19<sup>e</sup> et réalisée avec le soutien financier du Fonds national suisse pour la recherche scientifique. Elles sont tirées des archives genevoises et proviennent plus particulièrement de six recensements de la population genevoise de 1816, 1822, 1828, 1831, 1837 et 1843. Plus précisément, les données de ces recensements n'étant évidemment pas informatisées, nous n'avons extrait que les informations relatives aux individus dont le patronyme commence par la lettre "B", ce qui représente environ 12,5% de la population. Au total, en cumulant les 6 recensements, ce ne sont pas loin de 30'000 notices individuelles qui ont été relevées. La technique d'historiens consistant à échantillonner les "B" est neutre socialement et ethniquement, et s'apparente à un tirage aléatoire représentatif de la population (Bardet, 1983). Elle simplifie grandement la collecte à partir d'archives dans la mesure où elle permet de dépouiller plusieurs sources distinctes et d'y retrouver les mêmes individus.

La particularité de nos données est en effet que disposant et d'un échantillon alphabétique et du nom et de caractéristiques individuelles comme l'année de naissance, le sexe et même l'adresse, nous avons dès lors pu appairer les notices individuelles des recensements successifs. C'est cet appariement que nous exploitons pour analyser les entrées et sorties de la vie active et les mobilités socioprofessionnelles entre deux états de la population. Les cas qui n'ont pu être

Statuts sociaux		Groupes socioprofessionnels	
<i>ss_inc</i>	Inconnu	<i>gsp_inac</i>	Inactif
<i>ss_nqua</i>	Manuel sans qualification	<i>gsp_nqua</i>	Sans qualification
<i>ss_art</i>	Manuel qualifié	<i>gsp_art</i>	Artisan
<i>ss_colb</i>	Col blanc	<i>gsp_hor</i>	Horloger
<i>ss_pmb</i>	Petite et moyenne bourgeoisie	<i>gsp_com</i>	Commerçant
<i>ss_eli</i>	Elites	<i>gsp_serv</i>	Services privés et publiques

TAB. 1 – Liste des groupes socioprofessionnels et statuts sociaux

appariés fournissent également des informations précieuses. Un individu présent à un recensement qui ne l'est plus au recensement suivant a nécessairement quitté la cité, soit parce qu'il a émigré soit parce qu'il est décédé. De même, un individu qui n'était pas présent au recensement précédent et qui n'est pas né dans l'intervalle est nécessairement un immigré. Nonobstant la perte des "feux-follets", c'est-à-dire de ceux qui n'ont fait que passer par Genève entre deux opérations censitaires, c'est un tableau dynamique assez complet qui nous est ainsi dessiné.

L'analyse est centrée sur les recensements séparés de 6 ans. On considère ainsi les états au temps  $t$ , soit 1816, 1822, 1831 et 1837, et au temps  $t + 6$  (1822, 1828, 1837, 1843), ainsi que l'évolution de ces états entre  $t$  et  $t + 6$ .

Parmi les informations collectées, nous avons dénombré environ 1200 métiers qui ont été réorganisés d'une part en groupes socioprofessionnels, et d'autres part en statuts sociaux. Le tableau 1 liste les catégories retenues pour chacun de ces regroupements. Pour distinguer les états au début de l'intervalle de ceux à la fin nous préfixons les notations par ' $t_$ ' pour indiquer le début. Par exemple ' $t_gsp\_hor$ ' correspond à horloger en  $t$  et ' $gsp\_hor$ ' à horloger en  $t + 6$ . Le tableau 2 indique comment les cas retenus se répartissent selon ces catégories au temps  $t$ . Notons la distinction entre le statut social 'non qualifié' et le groupe professionnel 'non qualifié'. Un garçon boucher ou une fille de boutique par exemple ont un statut social 'non qualifié' mais appartiennent au GSP des artisans. Les transitions qui nous intéressent sont définies à partir des groupes socioprofessionnels et sont récapitulées au tableau 3 :

L'objectif de l'étude étant de comprendre comment le recrutement et la mobilité socio-professionnels sont liés aux facteurs démographiques nous avons retenus également comme variables pour notre analyse l'âge, le sexe et l'état civil. Par ailleurs, intéressé par l'impact

Groupes socioprofessionnels	Statuts	Inconnu	Manuel ss qual.	Manuel qualifié	Col blanc	P.M.B.	Elite	Total
Inactif		4467	23	0	79	1	344	4914
Sans qualification		274	1672	96	118	3	0	2163
Horlogerie		0	71	1330	0	213	0	1614
Artisan, manuel qualifié		0	173	1527	3	80	0	1783
Commerce		0	112	64	21	537	7	741
Services publics et privés		0	28	18	37	156	82	321
Total		4741	2079	3035	258	990	433	11536

TAB. 2 – Croisement des groupes socioprofessionnels avec les statuts sociaux (au temps  $t$ )

Transition	(désignation)	GSP en $t$	GSP en $t + 6$	autre condition
reste inactif	(inactif)	inactif	inactif	
devient actif	(nv_actif)	inactif	actif	
stable	(stable)	actif	actif	$GSP(t) = GSP(t + 6)$
mobile	(mobile)	actif	actif	$GSP(t) \neq GSP(t + 6)$
cesse l'activité	(retraite)	actif	inactif	
nouveau venu	(nv_venu)	non présent	présent	
disparu	(disparu)	présent	non présent	

TAB. 3 – *Les types de transitions*

possible de la montée du catholicisme qui, comme mentionné plus haut fait plus que doubler entre 1816 et 1843, nous avons également inclus la religion dont nous distinguons trois modalités : protestant, catholique et autre. S'agissant de l'état civil, le nombre de divorcés étant très faible ( $< 10$ ) nous ne considérons que les états célibataire, marié et veuf. Pour la religion comme pour l'état-civil nous retenons l'état au début  $t$  de l'intervalle censitaire à l'exception des nouveaux venus pour lesquels nous ne disposons que de l'état en  $t + 6$ . Pour l'âge, nous retenons celui du milieu de l'intervalle, soit l'âge en  $t + 3$ . Pour les analyses de statistique implicative, l'âge a été discrétisé en 3 classes (en minimisant la variance intra groupe) :  $age1$  à moins de 16 ans,  $age2$  de 16 à 41 ans et  $age3$  pour les plus de 41 ans. On distingue également deux périodes, l'une couvrant les transitions de 1816 à 1822 et de 1822 à 1828 et la deuxième celles de 1831 à 1837 et de 1837 à 1843.

On trouvera dans Oris et Ritschard (2007) l'analyse de la structure de la population et des transitions avec les outils classiques d'analyse démographique. Cette analyse renseigne sur les taux d'entrée et de sortie de la vie active, ainsi que sur les arrivées et départs de la Cité. Elle fait ressortir notamment l'importance des flux migratoires et la forte croissance du secteur de l'horlogerie, mais reste essentiellement limitée à la population dans son ensemble. Le calcul de tables, avec leurs probabilités, leurs courbes de survie, leurs âges moyens, serait a priori aussi voire plus profitable s'il était fait par groupe d'âges, sexe, état-civil et religion par exemple. Mais deux obstacles se présentent. L'un est la dispersion des effectifs, surtout lorsqu'ils sont distribués par groupes d'âges quinquennaux. L'autre est la multiplicité de résultats qui en résulterait, tant à cause de la division en six groupes socioprofessionnels que de la multiplicité des dynamiques. Il y a donc nécessité d'approches plus synthétiques.

Une première possibilité est la modélisation statistique. Les régressions logistiques présentées dans Oris et Ritschard (2007) permettent ainsi de nuancer l'information des tables par la prise en compte de variables démographiques et de la religion. Elles n'en produisent pas moins une avalanche de chiffres à peu près aussi déconcertante qu'avec les tables. L'analyse des correspondances multiple est une autre alternative considérée. Elle donne une vision globale des liens entre GSP, types de mobilité et facteurs socio-démographiques. Elle reste sujette cependant aux difficultés d'interprétation des méthodes factorielles.

Nous considérons ci-après deux approches qui nous semblent pouvoir apporter un éclairage synthétique novateur des liens entre démographie et dynamique socioprofessionnelle. La première s'appuie sur la statistique implicative comme outil d'analyse de données (Gras et al.,

1996). Ceci permettra d’illustrer comment la statistique implicative vient enrichir aussi bien la vision partielle des tables démographiques, que la vue globale résultant d’une analyse factorielle. La seconde qui est une contribution originale est une analyse par arbre des dissimilarités des transitions.

## 4 L’analyse statistique implicative

Nous nous proposons donc d’examiner nos données avec les outils de la statistique implicative, soit plus particulièrement ceux mis à disposition par le logiciel CHIC (Couturier et al., 2006), à savoir l’arbre de similarité fondé sur la vraisemblance du lien de Lerman et al. (1981), l’arbre cohésitif orienté (Gras et Kuntz, 2006) fondé sur l’intensité implicative de Gras et al. (1996), et le graphe implicatif.

La figure 1 montre un premier graphe d’implication obtenu en n’incluant dans l’analyse que les transitions et les groupes socioprofessionnels d’origine et de destination. Le graphe a été obtenu avec des seuils relativement bas. Toutefois, cela ne remet pas en cause sa pertinence statistique. En effet, comme nous avons utilisé la mesure entropique d’implication, les seuils ne doivent pas être interprétés comme des significations statistiques. Avec la mesure classique non entropique, les relations indiquées ici sont d’ailleurs toutes significatives à des seuils supérieurs à 95%, mais se trouvent être noyées dans une quantité d’autres relations dont l’interprétation est moins intéressante.

Démonstration de l’efficacité de la méthode, la figure 1 montre des relations triviales découlant des définitions même du tableau 3, à savoir que ceux qui restent inactifs sont inactifs en  $t$  et en  $t + 6$ , que les retraités deviennent inactifs en  $t + 6$  alors que les nouveaux actifs étaient inactifs en  $t$ . Les autres relations sont plus instructives. Elles nous indiquent d’une part que les manuels non qualifiés en  $t + 6$  et dans une moindre mesure les artisans sont essentiellement constitués de nouveaux venus à Genève, d’autre part que quand on appartient à ces mêmes groupes en  $t$ , on a de forte chances de quitter la cité dans les 6 ans suivants. L’importante rotation de la population semble donc concerner principalement ces groupes de manuels qualifiés et non qualifiés.

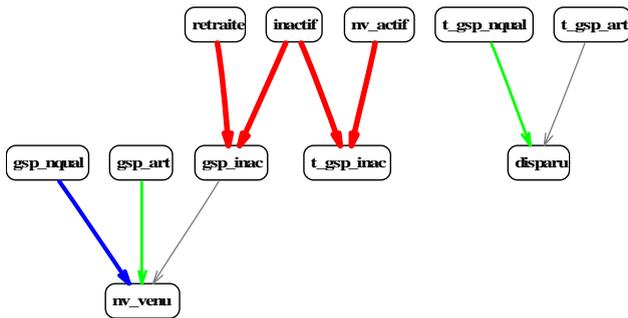


FIG. 1 – Transitions et groupes socioprofessionnels. Mesure entropique, seuils 99%, 81%, 63%, 58%.

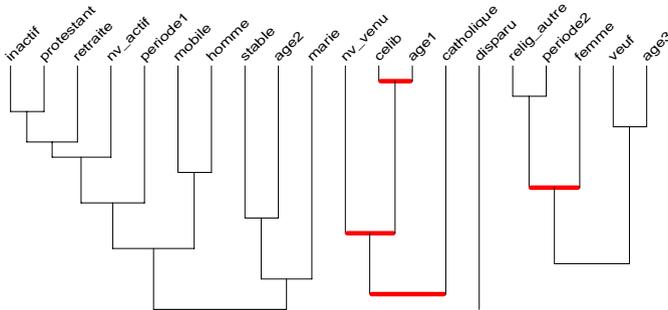


FIG. 2. *Transition et caractéristiques démographiques : arbre des similarités.*

Les figures 2, 3 et 4 montrent respectivement l’arbre des similarités, l’arbre cohésitif et le graphe implicatif obtenus en incluant les transitions, les variables démographiques et la religion. L’arbre des similarités fait ressortir clairement quatre groupes. En partant de la gauche, le premier correspond aux enracinés qui restent dans la cité ; il englobe essentiellement les familles locales. Leurs membres présentent le paradoxe apparent d’être professionnellement et stables et mobiles, mais la prévalence de la mobilité dans ce groupe s’explique essentiellement par l’entrée des jeunes sur le marché du travail (variable “nv\_aktif”). Ces types de transitions se recoupent essentiellement avec les caractéristiques protestant, homme et marié, ce qui est cohérent. Le second groupe comprend les nouveaux arrivants qui sont plutôt jeunes — et donc célibataires — et catholiques, et le troisième ceux qui quittent la cité. Ces disparus ne se regroupent avec aucune caractéristique démographique. Enfin, le dernier groupe n’est guère intéressant car il ne comporte aucune variable de transition et présente des caractéristiques hétérogènes.

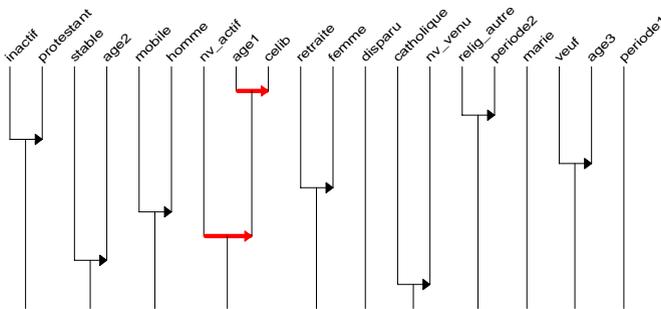
Le principal enseignement de l’arbre cohésitif est qu’il n’y a pas de méta-règle intéressante. La seule qui apparaît, à savoir ‘nouveau venu’  $\Rightarrow$  (‘jeune’  $\Rightarrow$  ‘célibataire’), n’est rien d’autre en effet que la règle simple ‘nouveau venu’  $\Rightarrow$  ‘jeune’, puisque les jeunes sont évidemment célibataires, en particulier dans le contexte genevois de mariage tardif décrit plus haut.

Le graphe implicatif 4, obtenu comme celui de la figure 1 avec le critère entropique et des seuils relativement bas, est plus instructif. On y observe quatre modalités qui polarisent les effets des types de transition. Ce sont ‘protestant’, ‘femme’, ‘homme’ et ‘célibataire’. Les nouveaux venus, les inactifs et ceux qui commencent une vie active sont principalement célibataires, les deux derniers groupes étant de jeunes célibataires.<sup>1</sup> L’inactivité, le début d’activité et la cessation d’activité sont plus le fait des femmes, tandis que la stabilité tout comme la mobilité socioprofessionnelle sont l’apanage des hommes. Quant à l’attribut ‘protestant’ il est associé à toutes les transitions internes, c’est-à-dire toutes sauf les nouveaux venus et les disparus. Ceci indique que si l’on trouve une majorité de protestants dans tous les groupes de transitions non migratoires, les fidèles de Calvin sont moins concernés par les mouvements

1. L’existence des fermetures transitives que nous commentons a bien entendu été vérifiée. Ainsi, les implications directes des inactifs et nouveaux venus sur célibataires ont toutes deux des intensités supérieures à 75%. Nous avons choisi de ne pas les représenter dans le graphe pour des raisons de lisibilité.

migratoires. On note, en particulier, que les catholiques sont plutôt constitués de nouveaux venus, le seuil de cette relation étant cependant inférieur à 65%. Les implications fortes de ‘autre religion’ sur les femmes et les célibataires sont plus difficiles à interpréter en raison de la nature résiduelle de cette modalité de la variable religion.

Pour tenter de voir comment les relations mises en évidence sont liées aux statuts sociaux, nous avons introduits ceux-ci comme variables supplémentaires ce qui permet d’obtenir avec CHIC des mesures de leur typicalité pour les implications qui nous intéressent. La typicalité fournie peut être vue comme la probabilité que l’implication soit indépendante de la variable supplémentaire. Ainsi, une typicalité inférieure à 5% indique que l’implication est caractéris-



Arbre cohésitif: C:\G\Project\CHIC\my examples\chic\_aveiro partition supp.csv  
 FIG. 3 – Transition et caractéristiques démographiques : arbre cohésitif, mesure entropique.

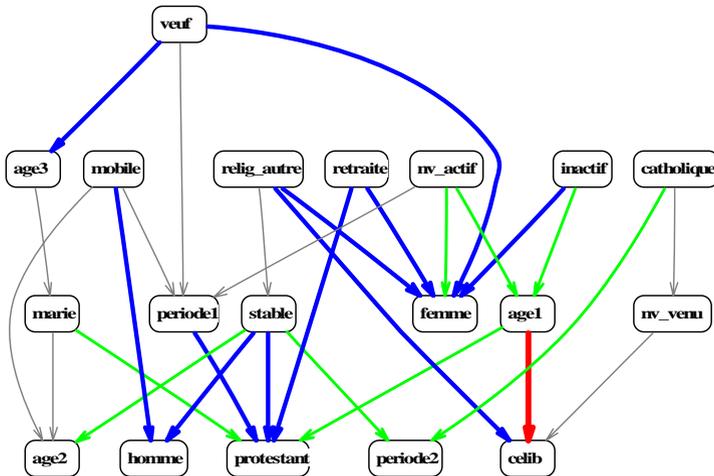


FIG. 4 – Transitions et variables démographiques : graphe implicatif, mesure entropique, seuils 99%, 75%, 65%, 55%.

Chemins	Statuts	Inconnu	Manuel ss qual.	Artisan, manuel qualifié	Col blanc	P.M.B.	Elite
stable $\Rightarrow$ protestant	.	.	x	x	x	x	.
stable $\Rightarrow$ homme	.	.	.	x	x	.	x
mobile $\Rightarrow$ homme	.	.	.	x	x	x	x
nv_actif $\Rightarrow$ protestant	.	.	x	x	x	x	.
nv_actif $\Rightarrow$ célibataire	.	.	x	x	x	x	.
nv_actif $\Rightarrow$ age1 $\Rightarrow$ célibataire	x	.	.	.	x	.	.
nv_actif $\Rightarrow$ femme	.	.	x	.	.	.	.

TAB. 4 – *Typicalité des statuts sociaux pour quelques chemins*

tique des individus ayant le statut social en question.

Par rapport à la figure 4, il est intéressant de noter que l'implication 'stable  $\Rightarrow$  protestant' n'est pas — avec des typicalités (Gras et al., 2006) supérieures à 0.75 — caractéristique des membres de l'élite, ni des individus de statut social inconnu. Toutes les autres typicalités étant quasi-nulles ( $< .001$ ), elle distingue cependant tous les autres statuts. L'implication 'stable  $\Rightarrow$  homme' est, quant à elle, typique des artisans, des cols blancs et de l'élite, mais ne l'est ni de la petite et moyenne bourgeoisie, ni des non qualifiés. Il en va de même du fait que la mobilité socioprofessionnelle soit l'apanage des hommes, si ce n'est que cette dernière relation est également typique parmi la petite et moyenne bourgeoisie. Si l'on considère maintenant les implications importantes de 'nouvel actif' sur 'célibataire' et 'protestant', les indices de typicalités montrent qu'ils caractérisent les non qualifiés, les artisans, les cols blancs et la petite et moyenne bourgeoisie, les statuts inconnus et les cols blancs étant typiques du chemin qui passe par la catégorie 'jeunes' (age1). Enfin, il est instructif, quoique attendu, de relever que la prédominance de femmes parmi la population qui débute (ou retourne à) la vie active est surtout propre au bas de l'échelle des statuts sociaux. Le tableau 4 récapitule les typicalités relevées.

Les figures 5 et 6 ont été obtenues en incluant comme variables régulières les groupes socioprofessionnels et les statuts sociaux respectivement du début  $t$  et de la fin  $t + 6$  de l'intervalle. On retrouve évidemment les quatre mêmes modalités de polarisation, avec toutefois une polarisation plus forte sur le nœud 'homme' où convergent les quatre GSP (tant en  $t$  qu'en  $t+6$ ) du haut de l'échelle, reflétant ainsi la dominance masculine dans les activités valorisantes. Par contraste, on voit que seul le GSP non qualifié converge sur 'femme'.

Il est intéressant de relever ici les différences entre les effets des statuts sociaux en tant que variables propres, et leurs typicalités discutées plus haut. Ici nous observons les effets sur les autres variables indicatrices, tandis que les typicalités rendent compte de l'effet sur les implications. En l'absence de méta-règles pertinentes, les typicalités et les effets comme variables régulières restent très similaires : quand une variable est typique d'une implication elle converge en général aussi sur la conclusion de la règle. Ainsi, par exemple, les statuts sociaux 'petite et moyenne bourgeoisie' et 'horlogers' qui sont typiques de l'implication 'mobile  $\Rightarrow$  homme', convergent sur 'homme'. Dès lors, comme à nouveau aucune méta-règle pertinente ne ressort des arbres cohésitifs (non montrés) correspondant aux graphes des figures 5 et 6, il ne nous a pas paru opportun de traiter par exemple les GSP en supplémentaires.

## Transitions socioprofessionnelles dans la Genève du 19e

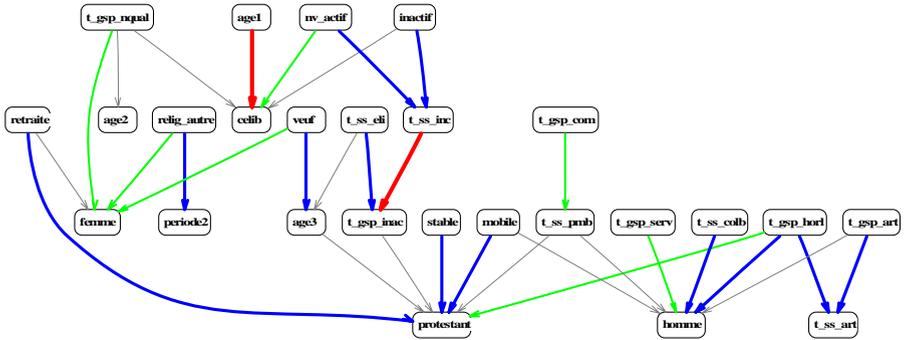


FIG. 5 – Graphe implicatif, groupe socioprofessionnel et statut social en  $t$ . Seuils 99%, 90%, 85% et 80%.

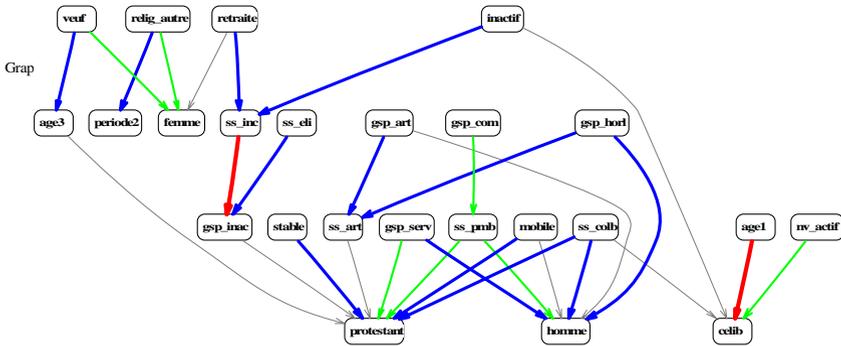


FIG. 6 – Graphe implicatif, groupe socioprofessionnel et statut social en  $t + 6$ . Seuils 99%, 90%, 85% et 80%.

## 5 Segmentation des transitions et intensité d'implication

L'analyse statistique implicative, tout comme l'analyse des correspondances menée dans Oris et Ritschard (2007), consiste essentiellement en une étude des liens entre modalités des variables. Elle contraste de ce point de vue avec les approches plus axées sur l'étude des liens entre unités statistiques comme les techniques visant à mettre en évidence des typologies par le biais de classification supervisées ou non des individus. Nous proposons de comparer ici les résultats obtenus avec ceux d'une analyse supervisée des dissimilarités des cas réalisée selon les techniques discutées dans Studer et al. (2009). Ce sera aussi l'occasion de démontrer l'intérêt de l'intensité d'application comme aide à l'interprétation des résultats de cette analyse.

Il s'agit dans un premier de temps de déterminer les dissimilarités entre chaque paire de cas, les cas étant dans notre contexte des transitions entre  $t$  et  $t + 6$ . Notons que comme la

base comprend plus de 18'000 transitions, cela représente plus de 162 millions de dissimilarités. Pour évaluer ces dissimilarités, le profil de chaque cas est considéré en termes des groupes socio-professionnels en  $t$  et  $t + 6$ . Chaque cas est ainsi représenté par le vecteur des 14 variables indicatrices de type  $t\_gsp$  et  $gsp$ . Les dissimilarités retenues sont les distances euclidiennes entre profils selon les scores de l'analyse des correspondances multiple de ces variables. Avec cette façon de procéder, deux cas qui diffèrent sur des caractéristiques rares sont plus dissemblables que deux cas qui diffèrent sur des caractéristiques fréquentes.

Le principe de l'analyse est ensuite celui des arbres de décisions. Il consiste à construire une arborescence en cherchant à chaque pas la meilleure façon d'éclater un nœud selon les modalités d'un prédicteur. La particularité est ici le critère utilisé, à savoir retenir l'éclatement qui explique la plus grande part de la diversité ou pseudo-variance des transitions. On exploite pour cela des relations qui permettent de calculer les pseudo-variances expliquées et résiduelles à partir des dissimilarités entre paires de transitions, et l'éclatement est sélectionné en fonction de la significativité de la part expliquée que l'on évalue par un test de permutation (voir Studer et al., 2009, pour les détails).

La figure 7 montre l'arbre obtenu. Pour l'interpréter, nous nous appuyons sur le concept d'intensité d'implication et visualisons dans les nœuds les transitions d'implication statistiquement significative. Plus précisément, l'implication considérée est celle de la règle dont la prémisse est définie par la branche menant au nœud et la conclusion la transition en question. Les transitions représentées sont ainsi les plus caractéristiques du nœud au sens où ce sont celles dont la fréquence relative des contre-exemples (total d'autres transitions) au sein du groupe est la plus significativement inférieure à la proportion d'autres transitions dans le nœud initial. Dans chaque nœud les transitions d'implication significative sont ordonnées de bas en haut par ordre décroissant de significativité. Au nœud initial, où toute les implications sont nulles par définition, sont reproduits les 10 transitions les plus fréquentes.

Cet arbre met en évidence trois groupes de transitions caractéristiques, à savoir les transitions entre les GSP inconnu et non actif qui, de façon attendue, sont essentiellement le fait de jeunes (moins de 16 ans en  $t + 3$ ), les transitions entre inconnu et sans qualification qui caractérisent les femmes adultes à l'exception notable des femmes protestantes de plus de 40 ans, et les transitions entre artisan, horloger et inconnu qui sont caractéristiques des hommes adultes. On peut relever que la stabilité dans la Fabrique (transition d'horlogerie vers horlogerie) concerne typiquement les hommes non catholiques, et que la stabilité dans le domaine du commerce est caractéristique des hommes non catholiques mariés. On peut relever également les importants mouvements de populations qu'illustrent les pavés bleus (inconnu) qui indiquent des immigrés lorsqu'ils sont à gauche (excepté pour les jeunes où il peut s'agir de naissances) et des émigrés (ou décès) lorsqu'ils sont à droite.

Si l'on compare à présent les enseignements de cette analyse des dissimilarités avec ceux de l'analyse de la section précédente, on peut relever leur similitude. Ainsi, la mobilité, c'est-à-dire la transition entre deux GSP actifs apparaît comme le fait d'hommes surtout tant dans le graphe de la figure 4 que dans l'arbre de la figure 7, où aucune des transitions typiques des femmes et des enfants ne se fait entre deux GSP actifs différents. Il en est de même de la stabilité, à savoir la transition d'un GSP actif à ce même GSP, qui au vu de la Figure 4 est l'apanage des hommes protestants, ce qui est confirmé par l'arbre puisque, si l'on excepte la stabilité dans l'état non-qualifié, les autres formes de stabilité sont typiques des hommes et essentiellement des hommes non catholiques. De façon générale, il ressort de l'arbre que



## 6 Conclusion

Nous avons présenté la mise en œuvre d'une analyse statistique implicative dans un contexte de démographie historique où il s'agissait d'appréhender l'articulation démographique des dynamiques socioprofessionnelles dans la Genève du 19<sup>e</sup> siècle. Plus que de tirer une conclusion sur le fonctionnement de la société genevoise de cette époque — que le lecteur intéressé trouvera dans Oris et al. (2006) —, nous aimerions ici faire le bilan de ce que nous ont effectivement apporté d'une part l'ASI et d'autre part l'arbre de dissimilarité des transitions couplé à une interprétation en termes d'intensité d'implication.

S'agissant de l'analyse réalisée avec CHIC, cet apport réside dans la vision synthétique et structurée des relations d'implication entre les multiples caractéristiques considérées. Les tables, taux et quotients utilisés en démographie permettent certes d'aller dans le détail et de quantifier les relations en termes de pourcentage de population concernée. Elles nécessitent cependant de se focaliser sur des aspects spécifiques et conduisent dès que l'on cherche à segmenter la population simultanément selon plusieurs variables de contrôle — sexe, état-civil, classe d'âge quinquennale, religion — à des effectifs souvent bien trop faibles, et sinon à une avalanche indigeste de chiffres. Les régressions logistiques peuvent palier en partie à ces faiblesses en permettant de mesurer les effets propres qui subsistent quand on contrôle pour les facteurs démographiques. L'inconvénient est qu'elles nous obligent à travailler sur des sous-populations, ceux qui restent actifs, ceux qui sont inactifs en début de période, ceux qui restent présents à Genève en fin de période, etc. Dans une optique plus synthétique nous avons également dans Oris et Ritschard (2007) procédé à une analyse factorielle des correspondances multiples. Le résultat s'est avéré éclairant, mais les tendances mises en évidence restent, bien que faisant sens, relativement floues. L'arbre de similarité des modalités offre une vision plus structurée des caractéristiques individuelles qui s'associent. Il serait intéressant, ce que nous n'avons ni eu le temps ni la place de faire ici, de comparer les regroupements proposés avec ceux résultant de méthodes traditionnelles de 'clustering' de variables. Une telle comparaison réalisée dans un autre contexte (Studer et al., 2007) semble curieusement indiquer que le 'clustering' classique produit une solution plus proche de l'arbre cohésitif que de celui de l'arbre de similarités. Finalement, la structuration sous forme de graphe d'implication nous est apparue plus enrichissante encore. Nous y avons retrouvé l'essentiel de ce que nous avons appris par des analyses détaillées menées avec les outils démographiques de base, et surtout elle nous a permis de mettre en lumière la forte polarisation des caractéristiques démographiques que sont le sexe et l'état-civil, et en particulier la dichotomie entre les enracinés protestants et les turbulences migratoires qui sont plus le fait de catholiques.

En toute équité il convient évidemment aussi de mentionner les limites de l'ASI. Un premier aspect auquel nous avons été confrontés est l'important travail de recodage de nos variables multinomiales qu'a nécessité la mise en œuvre de l'ASI. Un second point, plus fondamental, tient au fait que l'ASI se fonde sur les seuls liens bruts deux à deux entre caractéristiques. Les intensités de lien ne rendent ainsi compte que d'effets bruts, alors que l'on peut dans certaines circonstances être intéressé aux effets propres contrôlés pour les effets des autres caractéristiques. Par exemple, nous avons observé que les nouveaux venus sont en général célibataires, mais on sait aussi que surtout les jeunes sont célibataires. On peut alors se demander si le fait que les nouveaux venus soient célibataires ne reflète pas indirectement un effet d'âge. La relation est-elle valable autant pour les jeunes que pour les moins jeunes ? Une réflexion sur la possibilité de pouvoir distinguer et visualiser, en ASI, les effets propres des

effets bruts pourrait être une piste de développement. Un troisième élément qui peut perturber les adeptes de la modélisation statistique, est l'absence de critère permettant de juger de la pertinence statistique globale du modèle retenu. Quel pourrait être un équivalent de la déviance utilisée en modélisation statistique, ou de la part d'inertie reproduite en analyse factorielle ?

Pour ce qui est de l'étude des dissimilarités des transitions qui s'apparente à une analyse arborescente de la variance, elle s'avère utile pour identifier les facteurs les plus discriminants et renseigne sur leurs interactions. Contrairement au graphe d'implication qui rassemble des liens bruts mesurés entre paires de modalités sans distinguer a priori entre variables à prédire et prédicteurs, l'analyse arborescente des dissimilarités proposée est conçue pour décrire la diversité d'une variable réponse, le type de transition dans notre cas, en fonction de facteurs discriminants potentiels. Il s'agit d'une méthode supervisée au sens où la segmentation se fait en fonction des valeurs des prédicteurs. La nouveauté méthodologique dans ce que nous avons présenté est l'utilisation de l'intensité d'implication pour identifier les transitions caractéristiques des nœuds, ce qui donne une lecture intelligible de l'arbre. La méthode étant destinée à expliquer la variance, une mesure naturelle de l'information apportée et la part de variance expliquée. La faible valeur de 7.3% pour l'arbre de la figure 7 nous indique qu'il reste une part importante de variabilité résiduelle dans les nœuds.

Bien sûr cette méthode arborescente a également des limites. En particulier, comme pour toute méthode arborescente, se pose la question de la stabilité de la structure obtenue. Par ailleurs, nous n'avons pas tenu compte que plusieurs transitions pouvaient concerner un même individu ce qui nécessiterait une approche mult niveau pour laquelle il n'existe pas actuellement de solution dans le contexte des arbres de décision.

Au total, la corroboration mutuelle des enseignements des deux approches est un signe clair de leur intérêt et pertinence.

## Références

- Babel, A. (1942a). *La crise économique de Genève à l'époque révolutionnaire et les remèdes qu'on a tenté de lui opposer*. Genève : Georg.
- Babel, A. (1942b). *La Fabrique genevoise*. Neuchâtel-Paris : Attinger.
- Bardet, J.-P. (1983). *Rouen aux XVIIe et XVIIIe siècles : les mutations d'un espace social*. Paris : Société d'édition d'enseignement supérieur.
- Couturier, R., A. Bodin, et R. Gras (2006). CHIC v3.7 Classification Hiérarchique Implicative et Cohésitive. Guide d'utilisation, Ecole Polytechnique, Université, Nantes.
- Dufour, A. (1997). *Histoire de Genève*, Volume 3210 of *Que Sais-je ?* Paris : PUF.
- Gras, R., S. Ag Almouloud, M. Bailleul, A. Laher, M. Polo, H. Ratsimba-Rajohn, et A. Totahasina (1996). *L'implication statistique : Nouvelle méthode exploratoire de données*. Recherches en didactique des mathématiques. Grenoble : La pensée sauvage.
- Gras, R., J. David, J.-C. Régner, et F. Guillet (2006). Typicalité et contribution des sujets et des variables supplémentaires en analyse statistique implicative. In G. Ritschard et C. Djeraba (Eds.), *EGC'2006*, Volume RNTI-E-6 (2 volumes) of *Revue des Nouvelles Technologies de l'Information*, pp. 359–370. Cépaduès.
- Gras, R. et P. Kuntz (2006). Discovering R-rules with a directed hierarchy. *Soft Computing* 10(5), 453–460.

- Gras, R., P. Orús, B. Pinaud, et P. Gregori (Eds.) (2007). *Nouveaux apports théoriques à l'analyse statistique implicative et applications (actes des 4èmes rencontres ASI4)*, Castellò de la Plana, Espagne. Departament de Matemàtiques, Universitat Jaume I.
- Guichonnet, P. (1974). *Histoire de Genève*. Toulouse : Privat.
- Herrmann, I. (2003). *Genève entre république et canton. Les vicissitudes d'une intégration nationale (1814-1846)*. Québec-Genève : Presses de l'Université Laval-Éditions Passé Présent.
- Lerman, I. C., R. Gras, et H. Rostam (1981). Elaboration d'un indice d'implication pour données binaires I. *Mathématiques et sciences humaines* (74), 5–35.
- Oris, M. et G. Ritschard (2007). Dynamique socioprofessionnelle dans la Genève du 19e, enseignements d'une analyse de statistique implicative. In Gras et al. (2007), pp. 287–300.
- Oris, M., G. Ritschard, et G. Ryczkowska (2006). Recrutement et renouvellement des groupes socioprofessionnels à Genève, 1816-1843. In *14e Colloque de l'Association Internationale des Démographes de Langue Française AIDELF, Aveiro, 18-22 septembre 2006*, Aveiro, pp. 791–805. Université d'Aveiro et AIDELF.
- Perroux, O. (2006). *Tradition, vocation et progrès. Les élites bourgeoises de Genève (1814-1914)*. Genève : Slatkine.
- Schumacher, R. (2002). De l'analyse classique à l'analyse différentielle : nuptialité, fécondité et mortalité à Genève pendant la première moitié du 19e siècle. mémoire de DEA en Histoire économique et sociale, Université de Genève.
- Studer, M., G. Ritschard, L. Baccaro, N. S. Müller, et D. A. Zighed (2007). Relations entre types de violation des libertés syndicales garanties par les conventions de l'OIT : une analyse de statistique implicative des résultats d'une fouille de texte. In Gras et al. (2007), pp. 111–122.
- Studer, M., G. Ritschard, A. Gabadinho, et N. S. Müller (2009). Analyse de dissimilarités par arbre d'induction. *Revue des nouvelles technologies de l'information RNTI E-15*, 7–18.

## Summary

This paper reconsiders the analysis of the socioprofessional dynamic in the first half of the 19th century Geneva that we presented at the ASI4 Conference. It complements the study with a supervised analysis of the dissimilarities between transitions. Data are two by two matches between 6 censuses. More specifically, we consider the socioprofessional group (SPG) of each considered individual and its change between two successive censuses. Our interest is in the types of transition (stable, becoming active, becoming inactive, ...) as well as in new comers (immigrants and births) and lost cases (emigrants and deaths). The statistical implicative analysis of our data gives an insightful image of the relationships between these dynamics and the concerned SPG, as well as with some demographic and cultural variables (sex, age, religion). It highlights for instance polarizations around key variables. The dissimilarity analysis permits to segment the population into homogeneous groups using demographic (age, sex, marital status) and cultural (religion) predictors. The interpretation relies on typical transitions identified for each group by means of the intensity of implication. By so doing, we get also the required elements for comparing the outcome with the results of the implicative graph.

