

Version auteur du texte publié (p.91-110) dans

RÉSEAUX D'INNOVATION

Enjeux de la communication au sein d'une technopole, la cas Sophia Antipolis

Sous la direction de Paul Rasse, Céline Masoni et Jacques Araszkievitz

Coll. Communication et Civilisation

ISBN : 978-2-296-06448-5 - octobre 2008 - 192 pages

**Cartographies des flux virtuels
entre les acteurs de la technopole de Sophia Antipolis**

Eric Boutin, Gabriel Gallezot, Peggy Cadel, Pei Liu, Guillaume Perrin

CONTEXTE

La dynamique d'un territoire est fonction de l'articulation qui existe entre ses membres, de la capacité de ses acteurs institutionnels, de l'entreprise et de l'éducation à se mobiliser et à travailler en synergie. Différentes mesures de cette dynamique collaborative sont envisageables. Dans ce texte, nous nous sommes attachés aux relations hypertextuelles existantes entre les sites web des acteurs du territoire présents sur le site de Sophia Antipolis. Nous considérons comme Gérard Dubey¹ que : « *Parler du virtuel dans l'absolu revient (...) à admettre la possibilité d'une réalité sans référence à ce qui existe ou a existé. Or une telle réalité ex-nihilo est naturellement introuvable. Les relations qui se tissent sur le Web ont toutes pour modèle ou antécédent les relations sociales réelles* ». Ainsi, le lien hypertexte entre deux sites web est la trace virtuelle d'une relation réelle particulière. Ce sont, bien sûr, ces traces que nous travaillerons, pour déterminer une certaine vue de l'organisation sophilopolitaine ; il s'agira, notamment de les isoler et de les qualifier. Sur cet aspect du traitement des traces, le lien hypertexte recouvre plusieurs significations. Au début du Web, la

¹ Dubey G., *Le lien social à l'ère du virtuel*, Paris, PUF, 2001.

relation « citant/cité » légitimait la position forte, la centralité, de certains sites web. A cette signification originelle sont venues s'ajouter d'autres : dans une logique de plus en plus marchande (liens commerciaux, liens sponsorisés, achat de mots-clés, ...), le lien hypertexte est un outil au service du référencement, du marketing du site web dans les premières réponses des moteurs de recherche. Ces nouveaux sens du lien hypertexte viennent brouiller la trace directe (citant/cité) qui pourrait s'établir entre espace réel et espace virtuel. Certaines études conduites par Thelwall dans le domaine des sites web des universités² se sont attachées à caractériser des liens hypertextes par l'observation microscopique du contexte textuel du lien, permettant ainsi de se détacher d'une logique uniquement consumériste, pour focaliser sur la sémantique de la trace. Dans cet esprit, nous chercherons ici, à travers l'analyse des liens hypertextuels des sites web d'organismes sopolitains, les facteurs de caractérisation de ces « traces organisationnelles » entre les acteurs du territoire de Sophia Antipolis. Nous rendons compte de ces situations sous forme de cartes.

QUESTIONS DE RECHERCHE

Les hypothèses formulées ci-dessous permettent d'explicitier les critères retenus et sont susceptibles de qualifier les flux virtuels entre les acteurs de la technopole. D'autres hypothèses pourraient être formulées, d'autres éléments de détermination en découleraient. Les auteurs ont fait le choix de circonscrire leur champ d'analyse en raisonnant sur un corpus de données librement disponible, obtenu de manière non intrusive, ne nécessitant pas un processus de traitement trop lourd. Il s'agit ici d'une première phase d'analyse qui pourrait être prolongée par le recueil de données complémentaires, si cette étude trouvait un écho favorable.

La question nodale ici posée interroge en les appréhendant les facteurs qui vont déterminer la capacité d'un site web à émettre

² M. Thelwall, « What is this link doing here ? Beginning a fine-grained process of identifying reasons for academic hyperlink creation », *Information Research* 8, n° 3, avril 2003.

ou recevoir des liens hypertextes vers (ou en provenance) d'autres sites web. Plusieurs catégories d'hypothèses peuvent être formulées, la logique du lien hypertexte pouvant dépendre du statut du site web, de son envergure, de son ancienneté ou de son niveau territorial. Nous reprenons ci-après, successivement, chacune de ces familles d'hypothèses.

La première famille d'hypothèses suggère que les sites web peuvent être regroupés en différentes catégories : entreprise, association, éducation-formation, institution. Le choix de cette catégorisation obéit à l'idée selon laquelle chaque catégorie de cette typologie possède une logique propre en terme de création de liens hypertextes. Deux exemples caractéristiques peuvent être considérés. Certains acteurs institutionnels ont pour objectif de valoriser le territoire : ils jouent le rôle d'aiguilleur vers d'autres ressources. Cette fonction d'aiguilleur se retrouve-t-elle sur leur site web ? Les sites non-marchands qui ont pour objectif de valoriser les acteurs du territoire se caractérisent-ils par l'émission d'un nombre de liens sortant significativement supérieurs aux autres ? A l'inverse, certains acteurs obéissant à des logiques marchandes peuvent avoir des stratégies de rétention d'information pour ne pas orienter, par exemple, le client vers un concurrent. Transposé au web, cela se traduirait par le fait que certains acteurs aux logiques marchandes auraient peu de relations avec les autres acteurs marchands du même domaine. Cette observation peut-elle être faite dans le cadre de notre corpus ? De manière plus générale, ce découpage par type d'acteur amène à mesurer l'"endogamie" entre les différentes catégories, c'est-à-dire la capacité des acteurs à émettre ou à recevoir plus volontiers des liens hypertextes vers ou en provenance d'acteurs appartenant à la même catégorie. Cette analyse de l'endogamie peut être effectuée à différents niveaux de granularité pour conduire, par exemple, à s'interroger sur les interactions hypertextuelles entre des sites web d'entreprise de même code APE.

La deuxième famille d'hypothèses conduit à s'interroger sur l'influence de la taille d'un site web sur sa capacité à générer ou recevoir des liens hypertextes. On peut émettre l'hypothèse selon laquelle l'existence d'un gros site multiplie les occasions de liens

entrants ou sortants entre ce site et les autres sites. Cette relation entre taille du site et nombre de liens entrants ou sortants peut-elle être validée empiriquement sur le corpus constitué ?

Enfin on peut formuler une troisième série d'hypothèses qui consistent à privilégier les facteurs spatio-temporels comme déterminants de la logique relationnelle. Au titre de la logique temporelle, on peut formuler l'hypothèse selon laquelle les relations qu'une entité crée avec son environnement prennent du temps. Il en découle que plus une entité est ancienne et plus, *ceteris paribus*, elle aura de relations avec d'autres entités. En transposant cette logique au Web, on peut émettre l'hypothèse selon laquelle les sites web anciens ont une probabilité plus forte que les plus récents de recevoir de la part des autres des liens hypertextes. Cette hypothèse est-elle validée sur le corpus ? En suivant la logique spatiale, on peut s'interroger sur l'influence du niveau de rattachement territorial du site web sur sa capacité à recevoir ou émettre des liens vers ou en provenance de sites web d'un territoire donné. Chaque site peut se rattacher à une dimension communale, intercommunale, départementale, régionale, nationale ou internationale. On pourrait émettre l'hypothèse selon laquelle les sites locaux vont plus valoriser des relations locales que les sites nationaux, ou à plus forte raison internationaux. Cette hypothèse est elle validée dans notre cas de figure ?

PROTOCOLE MIS EN ŒUVRE

Nous avons obtenu par le biais du SAEM Sophia Antipolis Côte d'Azur³ une liste de 800 organisations implantées sur le site de Sophia Antipolis. A l'aide de requêtes sur les moteurs de recherche du Web, nous avons recueilli 550 URL de sites web, puis collecté les 3000 premières pages de chacun de ces sites web en partant de la page d'accueil et en explorant le site avec un niveau

³ SAEM Sophia Antipolis Côte d'Azur, Aménagement, commercialisation et gestion du Parc de Sophia Antipolis et des autres zones d'activités départementales, <http://www.sophia-antipolis.net>

de profondeur de trois ⁴. Nous avons ensuite extrait de ce corpus la liste des liens hypertextes sortants (du site vers d'autres sites)⁵, pour ne garder qu'une liste fermée de liens qui font référence aux 550 sites web. Ainsi, on dispose à l'issue de ces étapes d'un réseau constitué par les interactions entre ces 550 sites. Ce réseau peut être représenté comme un ensemble de nœuds et de flèches. Chaque nœud correspond à un des 550 sites web de notre corpus. Une flèche entre deux nœuds signifie qu'il existe au moins un lien hypertexte d'une page du site de départ vers une page du site d'arrivée. Ce réseau peut être également analysé par les indicateurs de l'analyse réseau présentés, par exemple, dans l'ouvrage de Wasserman et Faust⁶.

Ces données relationnelles sont enrichies pour chaque site du corpus par des informations identitaires sur chaque site, comme sa catégorie, son nombre de pages, son âge. Chacun des 550 sites est classé dans une des catégories (entreprise, association, éducation-formation, institution). La catégorie « Entreprise » désigne une organisation ayant une activité marchande hors éducation et recherche. La catégorie « Education » désigne une organisation ayant une activité d'éducation et de recherche dans le domaine public ou privé. La catégorie « Association » désigne une organisation qui revendique un statut associatif. Enfin, on entend par « Institution » une organisation qui joue, dans son domaine, le rôle de liant entre d'autres acteurs : organisation professionnelle, collectivité territoriale. Le nombre de pages du site est estimé par le nombre de pages du site dans la base de données de Google. Pour chaque site, cette information est le résultat de la commande `inurl :www.nomdusite` sous google. L'âge du site est obtenu par interrogation d'un *registrar*. La dimension communale,

⁴ Cette collecte s'est déroulée du 6 au 11 octobre 2007 en utilisant le logiciel *webpipe pro* (<http://www.datamystic.com/>).

⁵ Cette opération a été réalisée par le logiciel *text pipe pro* (<http://www.datamystic.com/>)

⁶ Wasserman S., Faust K., *Social Network Analysis methods and Applications*, Cambridge, England and New York, Cambridge University Press, 1994.

intercommunale, départementale, régionale, nationale ou internationale du site est appréciée par les informations relatives à l'adresse fournie par le propriétaire du nom de domaine lors du dépôt du nom de domaine. Le code activité de l'entreprise est obtenu grâce aux informations gratuites délivrées par certaines bases de données en ligne (www.societe.com par exemple).

Nous obtenons donc, à l'issue de ce processus de collecte, un corpus constitué des liens des organisations sophilippines, caractérisées par leurs inter-relations et une qualification de leurs identités sur le Web, qui nous permet de cartographier les acteurs.

RESULTATS OBSERVES

Sur les 550 sites étudiés, 490 ont des interactions entre eux et il y a 807 interactions entre ces 490 sites. Le réseau global qui représenterait les 807 interactions entre ces 490 sites est donc inextricable et peu exploitable. Nous allons donc en proposer différentes vues de synthèse qui constitueront chacune une grille de lecture spécifique de la réalité que nous cherchons à décrire. On observe par exemple *figure 1*, le noyau de taille maximale composé de 5 noeuds. Il s'agit d'un groupe d'organisations qui ont entre elles les relations les plus denses.

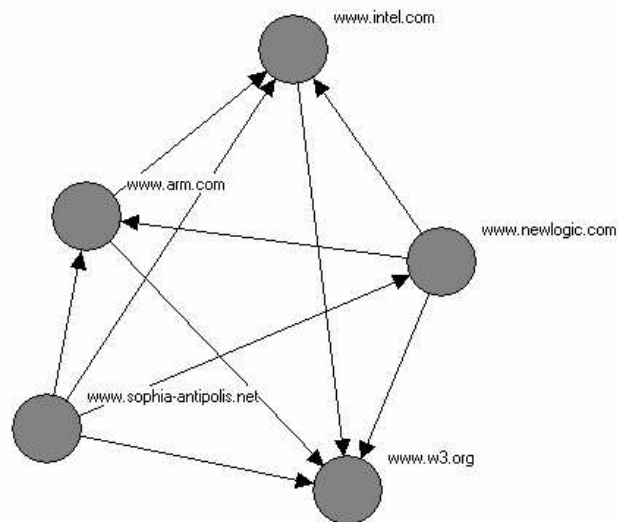


figure 1 : Noyau de taille maximale

On a souhaité *figure 2* recontextualiser les membres du noyau de taille maximale au sein du réseau global. Les 5 sommets (organisations) constituant le noyau de taille maximale sont alors représentés par des points de diamètre supérieur aux autres. Ce graphe permet de mettre visuellement en évidence la différence qu'il peut y avoir entre *w3.org*, qui reçoit des liens, et *sophia-antipolis.org*, qui en émet, une organisation qui cristallise l'activité sophilopolitaine, une autre qui la déploie. Cette représentation est connexe à la réalité.

Pour donner à lire autre chose, nous avons supprimé les deux sommets *www.w3.org* et *www.sophia-antipolis.net* ainsi que tous les sommets qui n'avaient qu'un lien avec les autres. Le résultat est présenté *figure 3*. On voit apparaître un réseau assez lisible qui comporte plusieurs composantes fortement connexes, des grappes, qu'il conviendrait de confronter à la réalité, ou d'autres analyses de ces mêmes acteurs pour révéler une situation, exposer des éléments ignorés jusqu'alors et capables d'impulser des changements organisationnels.

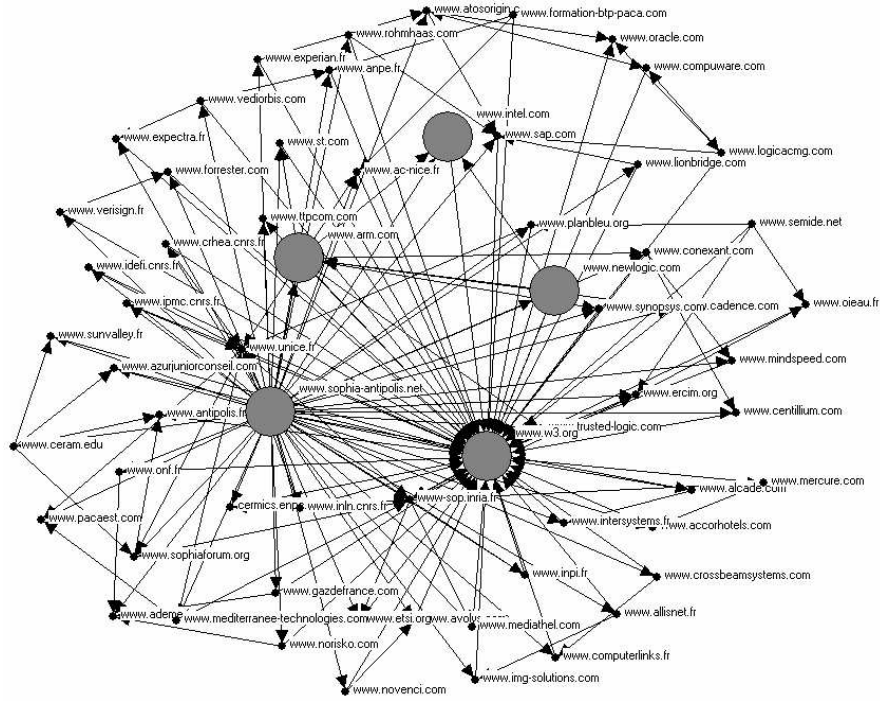


figure 2 : Interactions entre les sites du réseau global

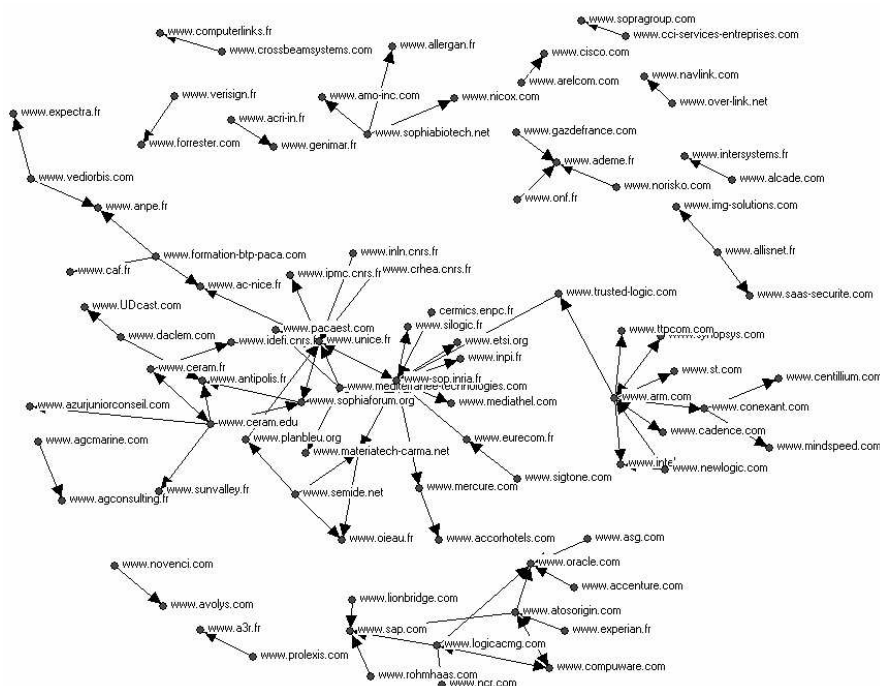


figure 3 : Réseau filtré par la suppression de deux sommets

Les précédentes cartes indiquaient les liens « bruts », non qualifiés entre les organisations. Nous reprenons l'hypothèse selon laquelle il existerait un enclin relationnel spécifique propre à certaines catégories d'acteurs. Ainsi, nous avons réparti les acteurs en 4 grandes familles : Entreprise (ENT), Education (EDU), Association (ASSO), Institution (INSTI). Les résultats sont présentés *tableau 1*. Lorsque cette affectation n'était pas possible, le site n'a pas été classé : ainsi on observe que 454 des 485 sites ont été affectés :

type	Nombre de sites web
INSTI	19
ENT	399
EDU	19
ASSO	17

tableau 1 : Répartition des sites web par famille d'acteurs

Nous avons ensuite cherché à qualifier les 808 interactions qui existent entre ces 485 sites web, en pointant chaque catégorie ou famille au regard de l'autre. On obtient ainsi une information agrégée qui peut être présentée de façon équivalente sous forme matricielle ou cartographique. Cette agrégation nous a permis de qualifier 758 interactions (quand le statut du site de départ et du site d'arrivée étaient définis). Les résultats sont présentés de façon équivalente *tableau 2* et *figure 4*.

	INSTI	ENT	EDU	ASSO
INSTI	30	349	17	20
ENT	244	41	3	0
EDU	16	4	16	6
ASSO	6	0	4	2

tableau 2 : Interactions réelles ou observées

Dans la représentation cartographique, chaque catégorie est représentée par un sommet et chaque épaisseur de lien est proportionnelle à l'interaction définie dans le *tableau 2*. La relation réflexive au niveau de chaque sommet est appréciée par l'épaisseur du sommet.

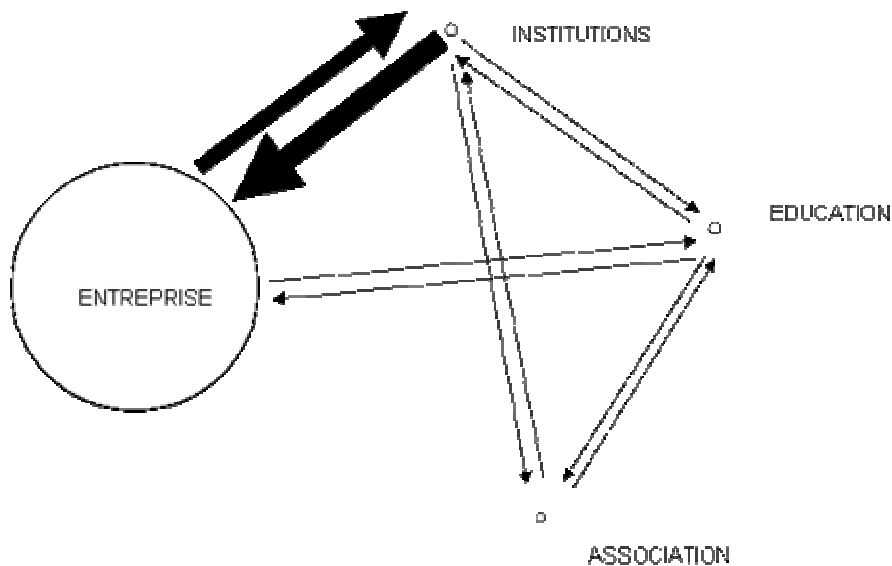


figure 4 : Interactions réelles ou observées : vision cartographique

Cette cartographie et la matrice associée révèlent que les interactions sont polarisées par deux familles d'acteurs (institution et entreprise). Les autres interactions sont assez marginales au regard de ces deux principales. Le nombre de relations entre sites d'entreprises est assez faible. Là encore, ces résultats pourraient être superposés à d'autres analyses et permettre ainsi une lecture approfondie de l'ensemble.

Pour donner une portée statistique à ces observations, nous avons effectué un test du *chi deux* qui cherche à invalider l'hypothèse selon laquelle chaque acteur a une probabilité identique d'être en relation avec un autre acteur. Le test du *chi deux* est effectué en comparant des valeurs observées (figurant dans le *tableau 2*) à des valeurs théoriques correspondant à une répartition équiprobable des associations entre les acteurs. Pour raisonner sur

des effectifs par cellule supérieurs à 5, nous avons supprimé de l'analyse les sites de type associatif.

Les données du *tableau 3* précisent quel devrait être le volume des associations théoriques entre les acteurs.

	INSTI	ENT	EDU	ASSO
INSTI	1,33	27,88	1,33	1,19
ENT	27,88	585,47	27,88	24,94
EDU	1,33	27,88	1,33	1,19
ASSO	1,19	24,94	1,19	1,06

tableau 3 : Interactions théoriques entre catégories

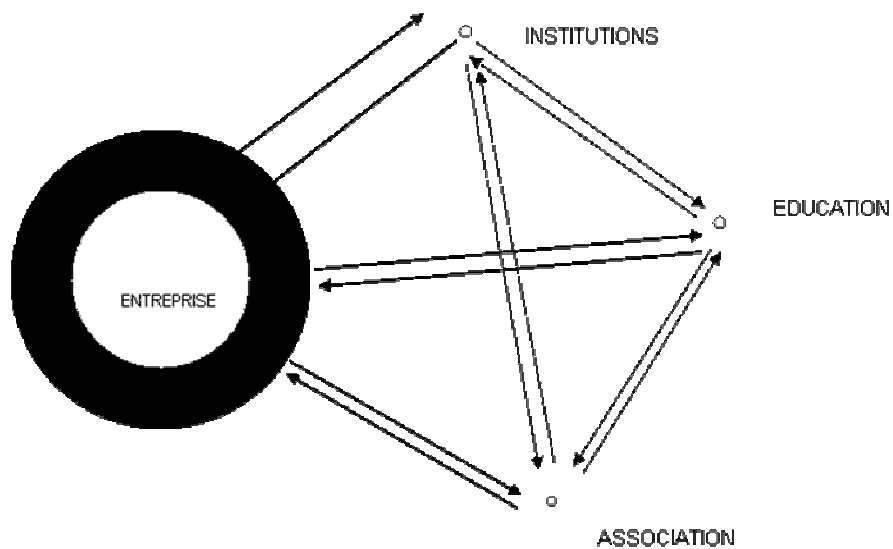


figure 5 : Interactions théoriques : vision cartographique

Le test du *chi deux* permet de valider avec une certitude de 95% que la nature de l'acteur a une influence sur les relations qu'il aura avec d'autres familles d'acteurs.

Si l'on rapproche la *figure 4* de la *figure 5*, on observe que les relations entre les entreprises sont statistiquement et significativement plus faibles que ce qu'elles devraient être. Les relations entre le pôle entreprise et le pôle institution sont statistiquement significativement plus fortes que ce qu'elles devraient être.

Cette observation peut être réexaminée à la lumière du réseau global. En effet, on a déjà remarqué que parmi les acteurs, deux sites classés parmi les institutions avaient un rôle prédominant en terme de relations avec les autres acteurs. Sophia-antipolis.net est un site qui a de nombreux liens hypertextes vers d'autres acteurs et w3.org qui reçoit de nombreux liens de la part de nombreux acteurs (entreprises notamment). On observe que sur les 808 interactions, 706 (soit 87%) font intervenir l'un de ces deux sites. 87% des liens de ces deux sites vont vers des sites d'entreprises (ceci est normal puisque 87% des acteurs sont des entreprises). 89% des liens qui pointent sur l'un de ces deux sites émanent d'un site de type entreprise. Ces deux sites n'ont donc pas de relations privilégiées avec certaines catégories d'acteurs. Si on supprime ces deux sites, on obtient des résultats qui ne permettent plus, faute d'effectifs, de réaliser un test du *chi deux*.

A l'issue de cette comparaison, il apparaît que l'analyse des relations entre catégories d'acteurs est perturbée par le fait que 87% des relations pointent vers ou émanent de deux acteurs. La relation forte entre Entreprise et Institution est due au poids particulier de ces deux acteurs. En revanche, les résultats confirment le nombre faible d'interactions entre les sites de type Entreprise. Il serait alors intéressant d'effectuer un focus sur les sites d'entreprises pour identifier l'activité des entreprises qui interagissent.

Sur les 41 interactions entre entreprises, pour un certain nombre, on ne connaît pas le code APE de l'entreprise citante ou

citée. Un travail d'affinage serait sans doute utile pour qualifier à la main les APE qui manquent. En l'état, on doit se contenter de 22 associations, présentées *figure 6*, sur un total de 41. Le réseau ci-dessous montre le niveau d'interaction entre code APE des acteurs. Aucun test ne peut être effectué car les effectifs ne sont pas suffisamment importants.

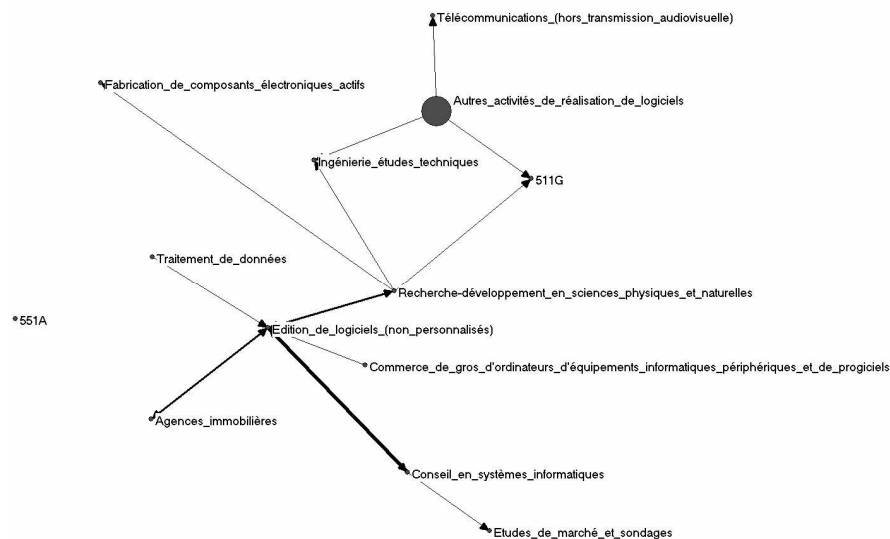


figure 6 : Interactions entre codes APE des entreprises

Le cercle « autres activités de réalisation de logiciels » plus épais que les autres correspond à un code pour lequel il y a deux entreprises du même secteur d'activité qui pointent l'une vers l'autre.

On peut également s'intéresser au nombre de liens entrants ou sortants par catégorie. On a supprimé de l'étude les deux sites www.w3.org et www.sophia-antipolis.net pour ne pas biaiser les résultats.

Si on s'intéresse, pour chaque site, au nombre de liens moyens qu'il a vers d'autres sites et au nombre de liens qu'il reçoit des autres sites, on peut observer les différences suivantes, précisées dans le *tableau 4*.

catégories	Nombre moyen de liens sortants vers d'autres sites	Nombre moyen de liens entrants sur les sites de la catégorie
Entreprises	1,46	1,14
Institutions	2,2	1,66
Sites associatifs	2,33	1,58
Sites Educatifs	2,63	2,29

tableau 4 : Nombre moyen de liens entrants et sortants par catégorie

Le *tableau 4* confirme que les sites d'entreprises ont proportionnellement moins de liens sortants vers d'autres sites et moins de liens entrants que les sites des autres catégories. Soulignant peut-être ainsi, que la fertilisation croisée sopolitaine, s'effectue moins par les entreprises que par d'autres catégories, ou qu'elle conserve encore le statut de mythe.

Pour valider une de nos hypothèses de départ, nous proposons d'analyser la correspondance entre la logique relationnelle du site et son niveau territorial. Pour ce faire, nous avons attribué à chaque site l'adresse du propriétaire du nom de domaine. Ces adresses ont fait l'objet d'une ventilation au niveau de plusieurs échelons territoriaux.

échelon territorial	nombre de sites
communal	157
départemental	70
régionale	11
national	117
international	102
total	457

tableau 5 : Niveau territorial du déposant du nom de domaine

Selon le même principe de catégorisation par famille, nous observons ici les interactions entre ces différents échelons territoriaux. Pour chaque lien hypertexte entre paire de sites, on établit une relation entre le niveau territorial du site de départ et le niveau territorial du site d'arrivée. On obtient le résultat suivant :

	communal	départemental	régional	national	international
communal	154	67	8	99	151
départemental	4	3		5	41
régional	2	1		1	6
national	7	5		10	88
international	4			8	97

tableau 6 : Interactions entre niveaux territoriaux

Nous avons représenté en grisé les cellules correspondant aux valeurs significativement supérieures à ce qu'elles devraient être si les interactions s'effectuaient uniquement en proportion du nombre d'acteurs par catégorie. Comme dans le cas précédent, ces observations sont biaisées par l'existence de deux acteurs principaux (sophia-antipolis.net qui est un site déposé dans la

commune et w3.org qui est un site dont le propriétaire est international). Cela se traduit dans la matrice par une première ligne significativement supérieure à la moyenne (beaucoup de liens sortants partent de sophia-antipolis.net) et une dernière colonne significativement supérieure à la moyenne (w3.org reçoit un grand nombre de liens). Pour gommer l'effet de ces deux sites, nous proposons de les enlever, en sachant que l'on passe ainsi, de l'étude de 761 interactions à l'étude de 96 interactions (soit 12% du total). Le résultat est fourni *tableau 7*. On peut observer le reste de la quantité d'information, mais on ne dispose pas d'une base statistique solide pour en tirer des observations.

	communal	départemental	national	international
communal	8	3	7	6
départemental	6	4	6	
national	7	5	10	4
international	3		7	20

tableau 7 : Observations réelles après avoir enlevé ces deux sites

Il est très difficile de parler d'interaction entre les acteurs de Sophia. Les 475 acteurs ont entre eux moins de 100 associations, ce qui correspond à une densité très faible. 80% des interactions sont le fait de deux sites.

Une des autres hypothèses que nous avons formulées est que la densification de certains liens pourrait être expliquée par l'ancienneté du site, selon le principe parfois observé, où le nouvel arrivant observe le fonctionnement de l'ancien et s'agrège au réseau pré-existant. La *figure 7* montre sur une échelle diachronique l'apparition des sites web des acteurs sur la toile et donc leur « visibilité » virtuelle pour les autres acteurs.

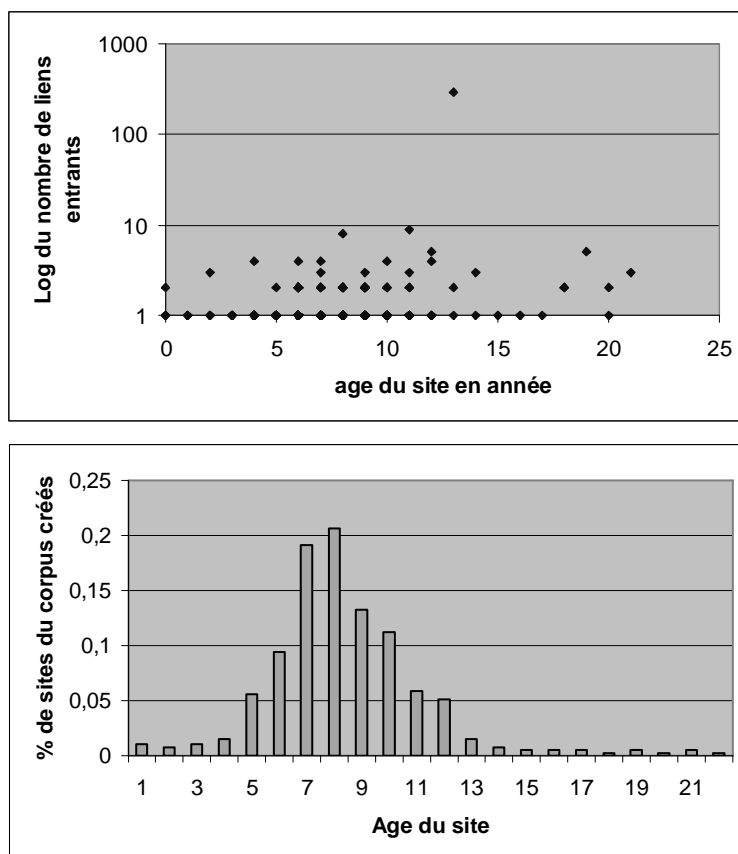


figure 7 : Age d'un site web et nombre de liens entrants

On observe ainsi qu'il n'y a pas de relations significatives entre l'ancienneté du site et son nombre de liens entrants, et ce malgré une pyramide des âges plutôt concentrée entre 5 et 12 ans. On remarquera aussi que le territoire sophilopolitain, en terme de date de création de site web, comporte certes des précurseurs (dès 1986), mais la masse se situe autour des années 2000, comme beaucoup d'autres sites au niveau international.

Un autre facteur explicatif du nombre de liens pourrait être la taille du site apprécié par son nombre de pages. Le principe sous-jacent est que plus le nombre de pages est élevé plus le site comporte d'informations et plus il touche potentiellement un

nombre des lecteurs importants, augmentant ainsi la probabilité d'être cité plus souvent. Le graphe présenté *figure 8* montre qu'il n'en est rien (sauf pour un site, celui du W3C certainement). Il y a une grande homogénéité des liens entrants quelle que soit la quantité de pages des différents sites.

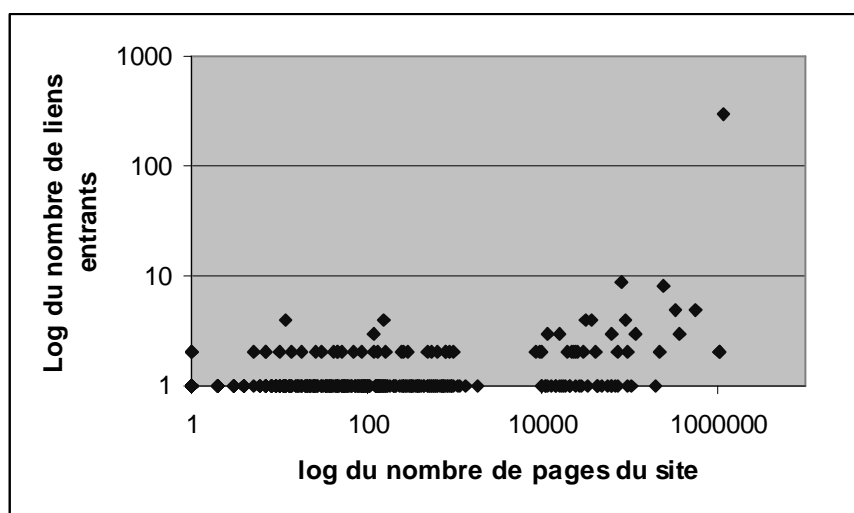


figure 8 : Relation entre nombre de pages du site et nombre de liens entrants

CONCLUSION

Les relations hypertextuelles mises en évidence à l'aide des trois catégories d'hypothèses indiquent des « traces organisationnelles », qui sont assez éloignées d'une figure rhizomatique, au sens où des tiges souterraines ramifiées (les noeuds des cartes *figure 1* et *figure 2*) émettraient chaque année des racines et des bourgeons (des liens entre acteurs), qui donneraient naissance à des tiges aériennes affleurant la surface du sol (des nouveaux acteurs). Au contraire, l'analyse des flux virtuels entre les acteurs du territoire sophilopolitain, le dispositif Sophia Antipolis montre deux actants majeurs, l'un fonctionnant comme attracteur (le w3c) l'autre fonctionnant comme liant (sophia-antipolis.net).

La représentation cartographique statistiquement calculée à partir des liens que s'attribuent les acteurs eux-mêmes propose finalement la métonymie de la technopole : un duo (le w3c et sophia-antipolis.net) homéostasique, aidé de trois éléments internationaux externes (Intel, Newlogic, AM) qui régule le dispositif.

Aussi, si on annihile ce duo pour gagner en granularité dans l'analyse du dispositif, on révèle d'autres éléments constitutifs du dispositif (*figure 3*), notamment un pôle éducation-recherche assez fortement représenté et détaché des acteurs commerciaux.

Parmi les résultats, on notera aussi l'homogénéité thématique du dispositif autour du « logiciel » et ce, quel que soit le type d'organisation... Le duo régule-t-il à plein l'entropie ?

Enfin, nous avons aussi mis en évidence qu'en dehors du duo, la densification du réseau est faible et majoritairement bilatérale. Il n'y a donc pas de véritable maillage, pas de rhizome ; les résultats, indiqués par les cartes réseaux, montrent encore un dispositif où la fertilisation croisée est en chemin.