



Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in: <https://oatao.univ-toulouse.fr/22144>

To cite this version:

Barros-Sampaio, Ricardo and Dousset, Bernard and Bahsoun, Wahiba *Les maladies négligées : approche bibliométrique et présentation des résultats sur le Chagas*. (2015) In: VSST 2015 : Veille Stratégique Scientifique & Technologique, 11 May 2015 - 13 May 2015 (Grenade, Spain). (Unpublished)

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

Les maladies négligées : approche bibliométrique et présentation des résultats sur le Chagas

Ricardo BARROS-SAMPAIO *

rsampaio.br@gmail.com

Bernard DOUSSET **, Wahiba BAHSOUN **

bernard.dousset@irit.fr, Wahiba.Bahsoun@irit.fr

* Fiocruz Brasilia, L3 Avenue North, S / N Campus Darcy Ribeiro, Brasilia (Brésil)

** Université de Toulouse, IRIT (UMR 5505) 118, route de Narbonne, 31062 Toulouse (France)

Mots clefs :

extraction de connaissances à partir de données, changement, nouvelles technologies de l'information et de communication (NTIC), représentation

Keywords:

knowledge extraction in databases, change, new information and communication technologies (NICT), representations

Palabras clave:

extracción de información a partir de los datos, cambio, nuevas tecnología de la informacion y la comunicacion (NTIC), representaciones

Résumé

Dans le cadre d'une étude transversale sur l'origine des recherches menées dans le monde sur les maladies négligées, nous avons systématiquement étudié les publications scientifiques issues de la base de donnée PubMed. Cette étude fait apparaître l'effort consenti par chaque pays pour endiguer ces maladies insuffisamment étudiées mais qui représentent une menace très forte dans les pays peu développés ou émergents. Le manque d'intérêt aussi bien pour la recherche amont que pour la mise au point de traitements efficaces étant certainement dû aux faibles retombées financières attendues. Les maladies que nous avons étudiées sont les suivantes : Chagas, tuberculose, malaria, leishmaniose, dengue, lèpre, schistosomiase, ... Nous ne présentons ici que des résultats partiels sur la maladie de Chagas qui sévit notamment au Brésil, pays qui, au fil du temps, est devenu leader dans les recherches sur cette maladie parasitaire.

1 Introduction

La maladie de Chagas ou trypanosomiase brésilienne est une forme de trypanosomiase (~ maladie du sommeil), une maladie parasitaire des régions tropicales, localisée en Amérique du Sud et Centrale, provoquée par un trypanosome (*Trypanosoma cruzi*), induite par les réduves (punaises hématophages).

Le Chagas en quelques chiffres :

- 100 millions de personnes à risque d'infection (25% Amérique latine),
- 8 à 10 million de personnes infectées,
- 13 000 décès par an (OMS) voire 50 000 (Yves Carlier 2003),
- 300 000 nouveaux cas par an,
- 13% de mortalité périnatale au Brésil.

Ci-contre le mode de développement de la maladie de Chagas et sa transmission à l'homme (Source Wikipedia).

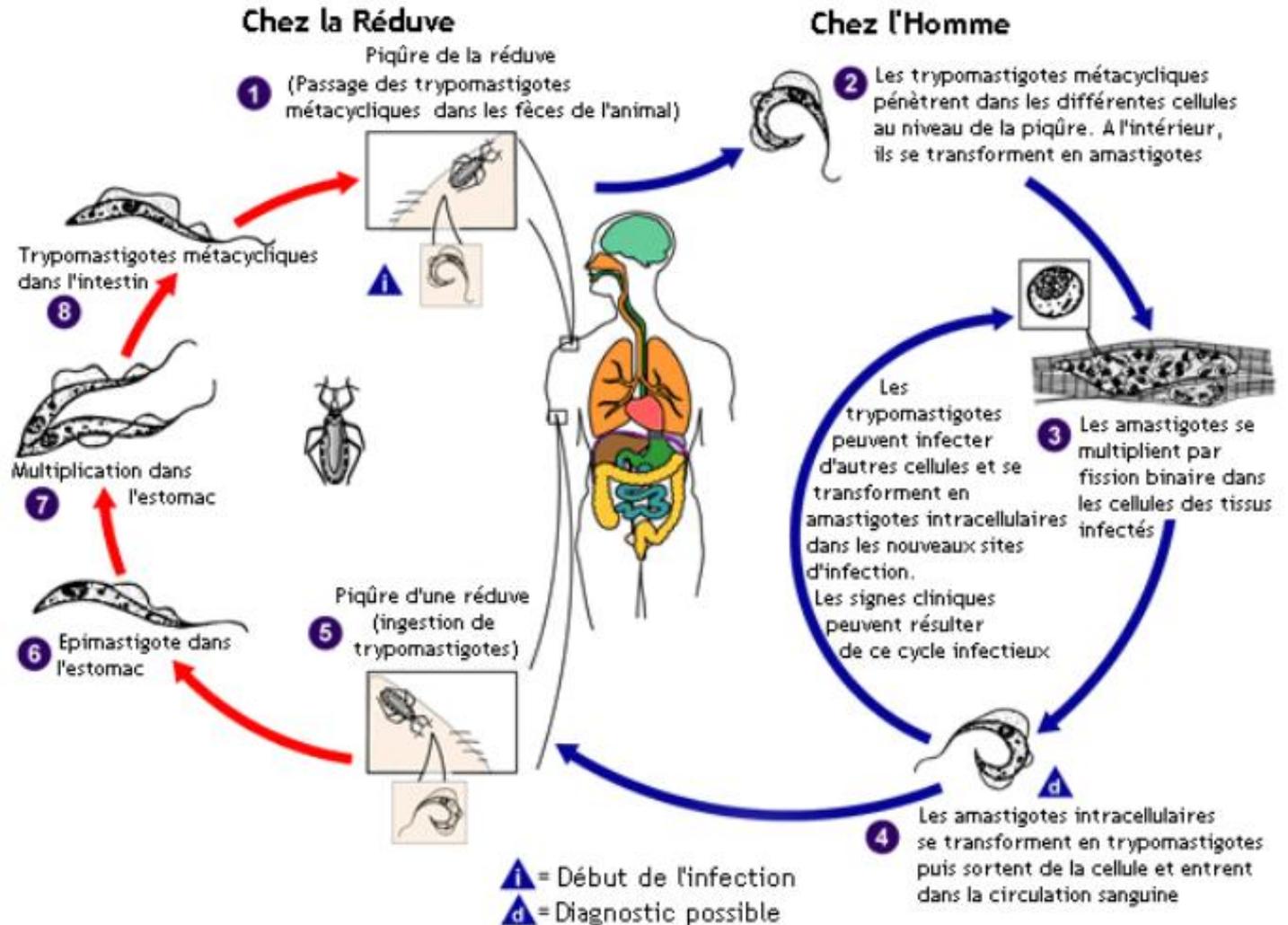


Figure 1 : Mode de transmission de la maladie de Chagas.

2 Méthodologie de l'étude

2.1 Constitution du corpus

Nous avons donc utilisé la base de données PubMed, le corpus déchargé en 2013 est constitué de 15500 notices bibliographiques, le nombre d'auteurs recensés est de 25100 après synonymies (9100 à 2 publications et plus), le nombre de journaux est de 1517, 81 pays ont participé à ces recherches ainsi que 26 des états du Brésil.

2.2 Collecte des données et constitution du corpus

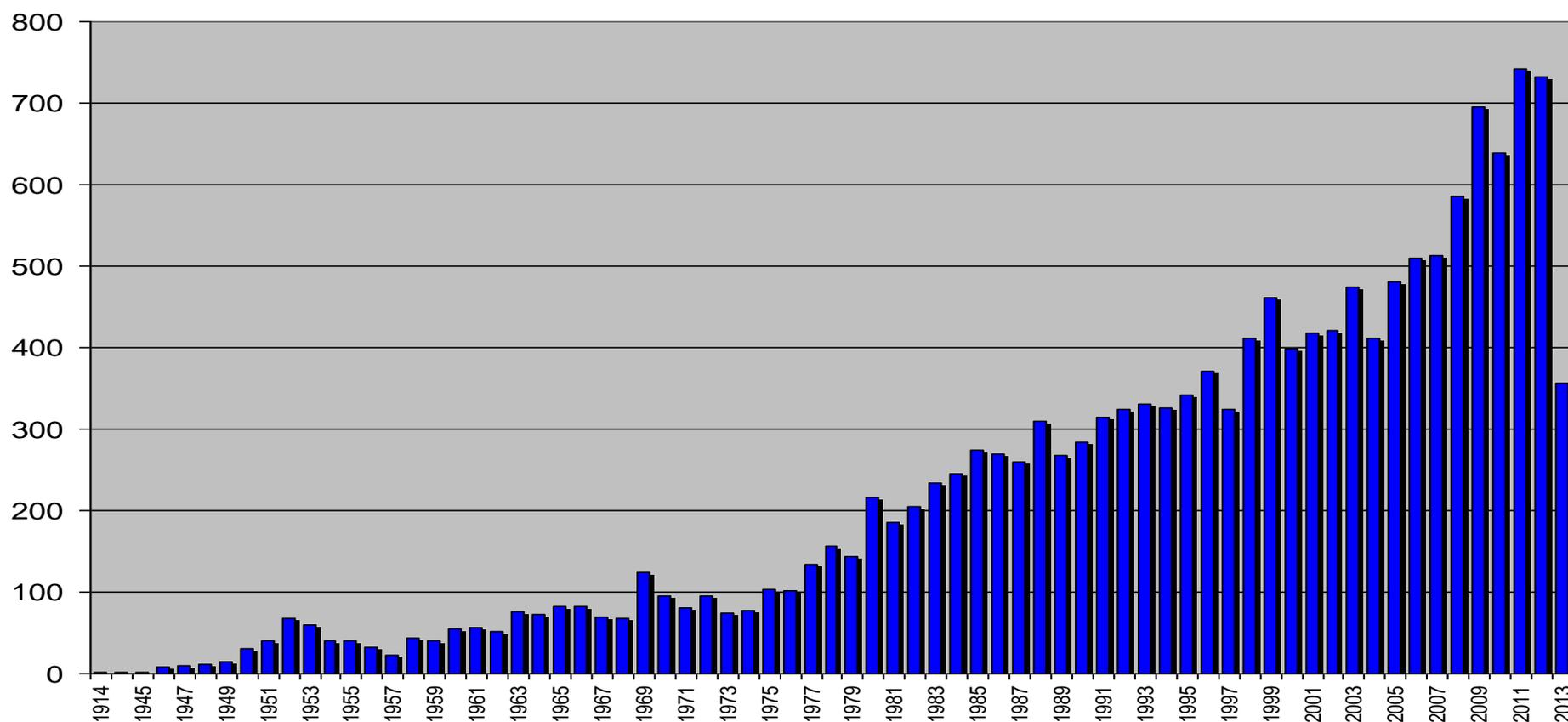


Figure 2 : Nombre d'article parus par année dans PubMed

La montée est très progressive depuis 1975, facteur qui est à relativiser par le nombre global de publications indexées dans PubMed pour chaque année. Deux pics sont à signaler : début des années 50 (de 1950 à 1956) et ensuite de 1969 à 1972.

3 Les acteurs du domaine

3.1 Les journaux

Sur la base PubMed, nous trouvons 1517 journaux qui ont abordé la maladie de Chagas. Le premier avec 714 publication émane de l'institut **Oswaldo Cruz** créée en 1900 dans le district de Rio de Janeiro et dont la vocation est la recherche et le traitement des maladies tropicales. Vient ensuite le journal *Molecular and Biochemical Parasitology* avec 489 publications suivi d'une revue à nouveau brésilienne : *The Revista do Instituto de Medicina Tropical de São Paulo* (The Journal of the Institute of Tropical Medicine of São Paulo) avec 469 articles, puis de *The American Journal of Tropical Medicine and Hygiene* avec le même nombre d'articles et d'une troisième revue brésilienne *Revista da Sociedade Brasileira de Medicina Tropical* (The Journal of the Brazilian Society of Tropical Medicine). Les revue brésiliennes sont donc fortement présentes sur le sujet est restent actives jusqu'à la période actuelle.

En effet, le croisement des journaux avec les périodes fait ressortir une assez bonne stabilité de la production scientifique dans les principaux journaux, sont à noter la disparition de Bol-Chil-Parasitol, Hospital-Rio-J et J-Protozool ainsi que l'émergence de Plos-Negl-Trop-Dis avec 153 articles depuis 2008 et Plos-One avec 97 articles pour la même période.

Une analyse factorielle sur ce croisement permet de mettre en évidence d'un seul coup d'œil tous les profils d'évolution des 1517 journaux recensés. En haut à droite de la figure 3 se trouvent les journaux émergents comme Plos-Negl-Trop-Dis, au centre du nuage ceux qui ont régulièrement publié des articles sur la maladie de Chagas sur l'ensemble des périodes prises en compte comme Mem-Inst-Oswaldo-Cruz.

		1903-13	1991-20	2001-20	2008-20
1	714 MEM INST OSWALDO CRUZ				
2	489 MOL BIOCHEM PARASITOL				
3	469 REV INST MED TROP SAO PAULO				
4	469 AM J TROP MED HYG				
5	467 REV SOC BRAS MED TROP				
6	459 EXP PARASITOL				
7	306 ARQ BRAS CARDIOL				
8	295 J PARASITOL				
9	285 TRANS R SOC TROP MED HYG				
10	283 ACTA TROP				
11	282 INFECT IMMUN				
12	248 MEDICINA (B AIRES)				
13	236 PARASITOL RES				
14	181 J IMMUNOL				
15	176 BOL CHIL PARASITOL				
16	162 INT J PARASITOL				
17	158 PARASITOLOGY				
18	157 PLOS NEGL TROP DIS				
19	154 J BIOL CHEM				
20	141 BRAZ J MED BIOL RES				
21	121 INT J CARDIOL				
22	118 HOSPITAL (RIO J)				
23	110 J PROTOZOOLOG				
24	107 ANN TROP MED PARASITOL				
25	100 PLOS ONE				
26	95 BIOORG MED CHEM				
27	94 J MED ENTOMOL				
28	94 ANTIMICROB AGENTS CHEMOTHER				
29	91 REV BRAS MALARIOL DOENCAS TROP				
30	87 J MED CHEM				
31	83 BIOCHEM BIOPHYS RES COMMUN				
32	81 PARASITE IMMUNOL				
33	77 INFECT GENET EVOL				
34	76 CAD SAUDE PUBLICA				
35	75 MICROBES INFECT				
36	75 J INFECT DIS				
37	74 REV BRAS BIOL				
38	71 PROC NATL ACAD SCI U S A				
39	70 PARASITOL TODAY				
40	70 BIOCHEM J				
		1903-13	1991-20	2001-20	2008-20
1	mem_inst1	149	252	171	142
2	mol_bio	149	190	108	42
3	rev_inst1	337	77	35	20
4	am_j_tr1	183	121	70	95
5	rev_soc1	54	156	148	109
6	exp_par	153	111	84	111
7	arq_bra	154	68	43	41
8	j_paras	170	66	39	20
9	trans_r	201	38	21	25
10	acta_tr	56	65	65	97
11	infect_6	79	96	75	32
12	medicin	161	61	17	9
13	parasit4	20	57	78	81
14	j_immun	86	44	32	19
15	bol_chi	132	44		
16	int_j_p	37	30	56	39
17	parasit6	35	34	42	47
18	plos_ne			4	153
19	j_biol_	16	55	57	26
20	braz_j_2	50	61	21	9
21	int_j_c1	10	21	32	58
22	hospita	118			
23	j_proto	104	6		
24	ann_tro	55	21	20	11
25	plos_on			3	97
26	bioorg_2		8	32	55
27	j_med_e	19	16	32	27
28	antimic	15	18	30	31
29	rev_bra6	91			
30	j_med_c	8	15	31	33
31	biochem	16	19	30	18
32	parasit2	22	35	9	15
33	infect_5			20	57
34	cad_sau		29	30	17
35	microbe		5	38	32
36	j_infec2	24	17	14	20
37	rev_bra	72	2		
38	proc_na	22	23	18	8
39	biochem3	9	32	18	11
40	parasit5	25	45		

Figure 2 : Nombre d'articles par journal et distribution sur 4 périodes homogènes

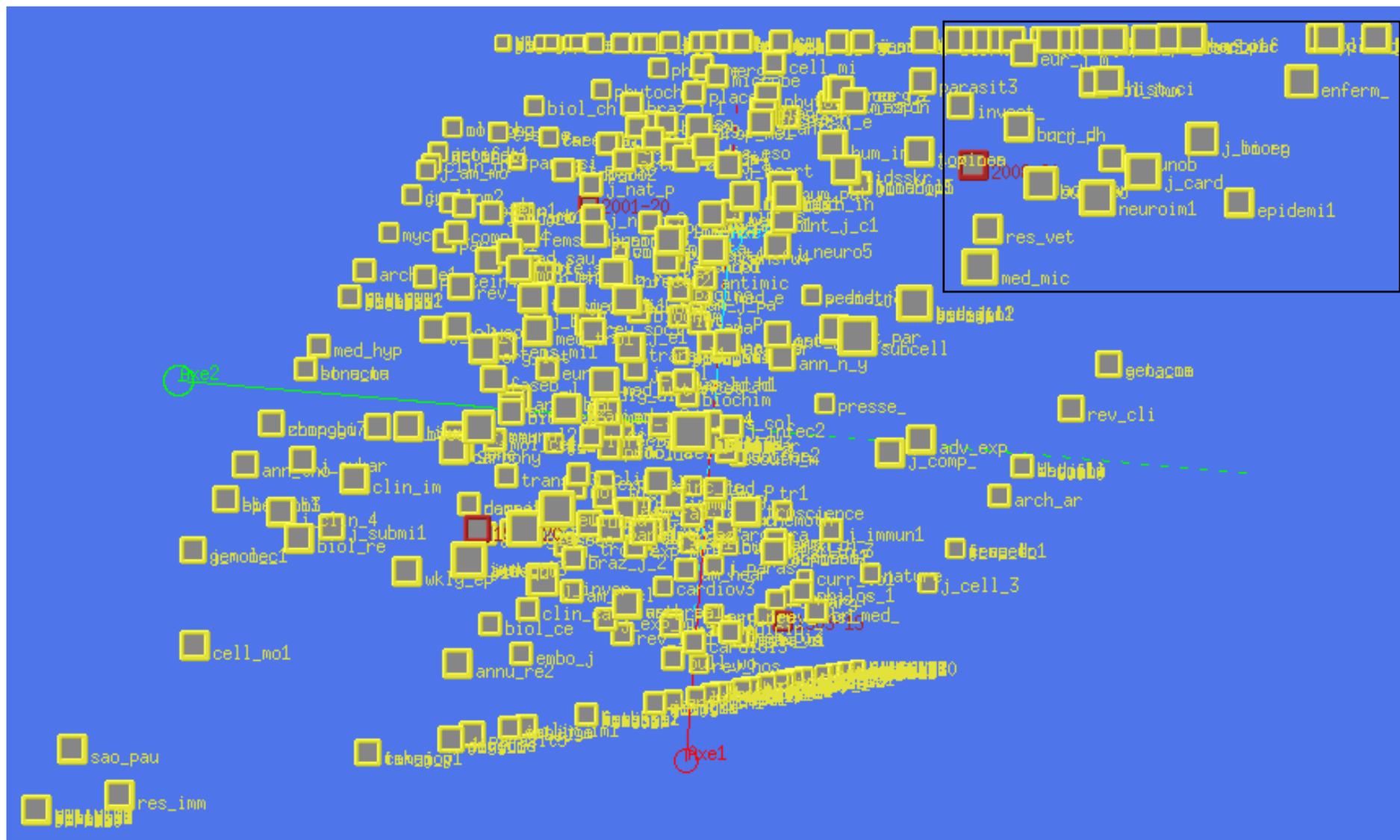


Figure 3 : Analyse factorielle des correspondances [10] des journaux sur les 4 périodes de référence

3.2 Les auteurs

3.2.1 Le réseau social des auteurs

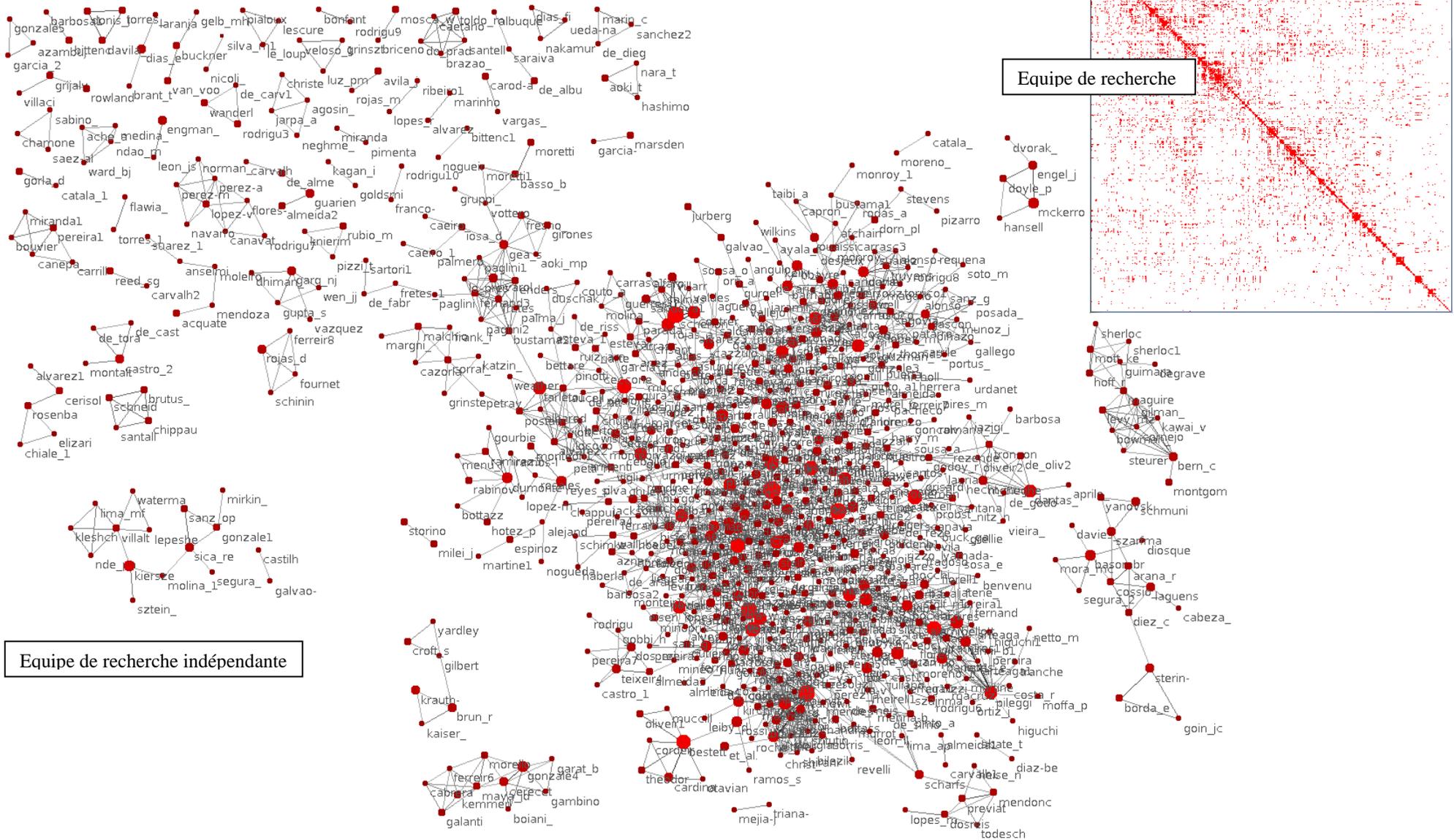


Figure 4 : Graphe du réseau social des auteurs (sans les liens 1) et tri par blocs diagonaux de la matrice initiale

3.2.2 L'évolution du réseau social des auteurs

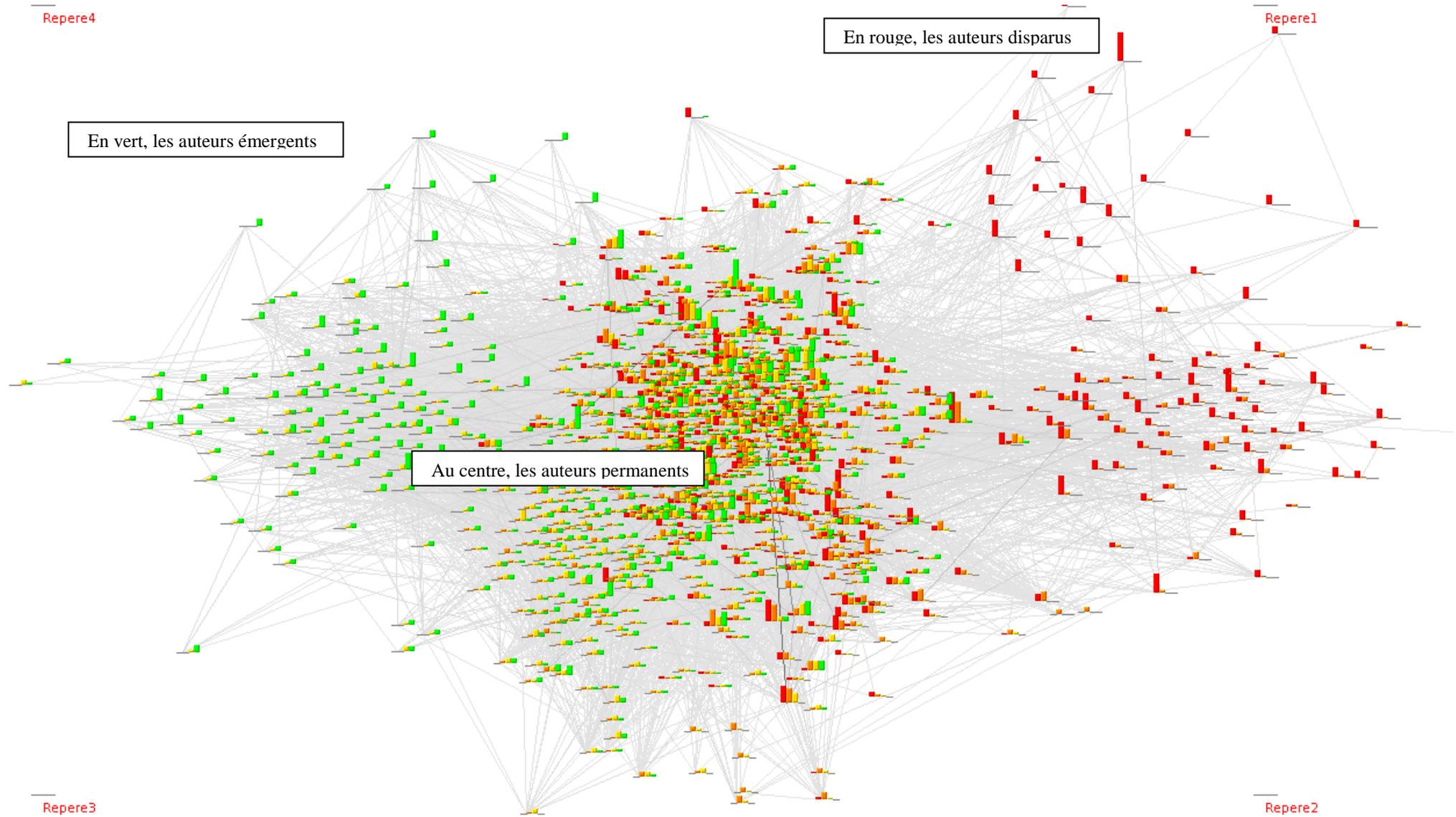


Figure 5 : Evolution du réseau social des auteurs : les repères représentent les 4 périodes de référence

Dans la figure 5, les auteurs se répartissent en fonction des périodes (repères) et de leurs liens (co-publications). Certaines équipes très anciennes ont disparu, d'autres n'ont été actives que durant les périodes intermédiaires (en jaune et orange), enfin les équipes émergentes sont en vert et se concentrent vers le repère 4, soit la période 2008-2013. La majorité de ces équipes ont en fait démarré au cours de la 3^e période donc avant 2008.

Ci-dessous, nous donnons le détail pour deux équipes, celle de Dias J. et celle de Docampo R. et de Cazzulo J. Les auteurs principaux de ces deux équipes se maintiennent depuis la première période, par contre on peut constater un turnover assez important au niveau de leurs collaborateurs. Pour la première équipe, Cardinalli Neto A. semble le plus émergent, pour la seconde il s'agit de Gonzalez M. et de Cerecetto H. Par contre nous pouvons constater que l'équipe de gauche est moins cohésive que celle de droite (moins de liens par auteur dans la plupart des cas).

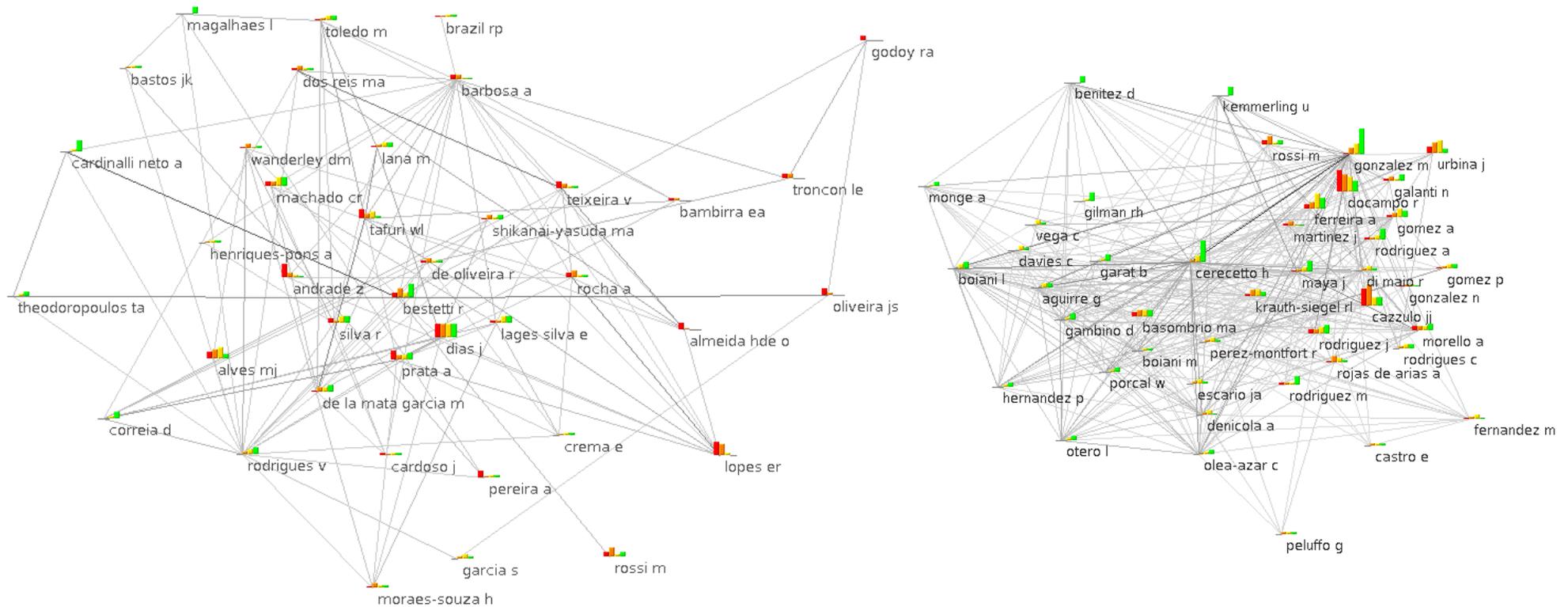


Figure 6 : Détail de l'évolution de deux équipes (à gauche DIAZ J., à droite DOCAMPO R.)

4 Localisation des recherches

4.1 Répartition en volume par pays et évolution

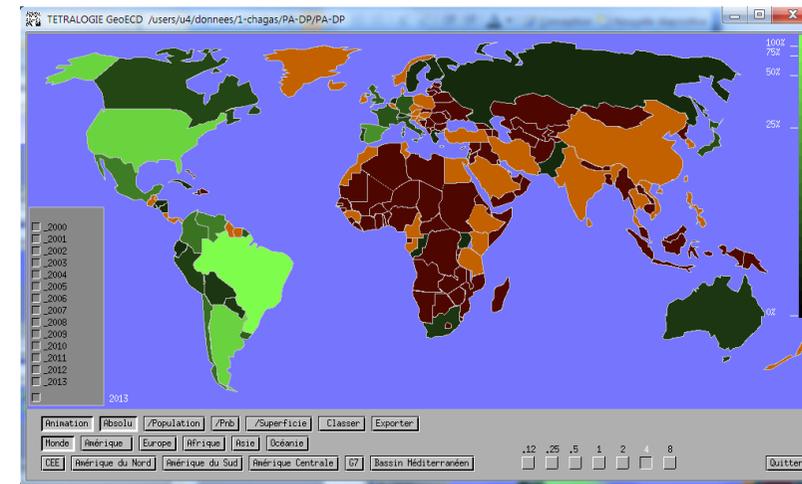
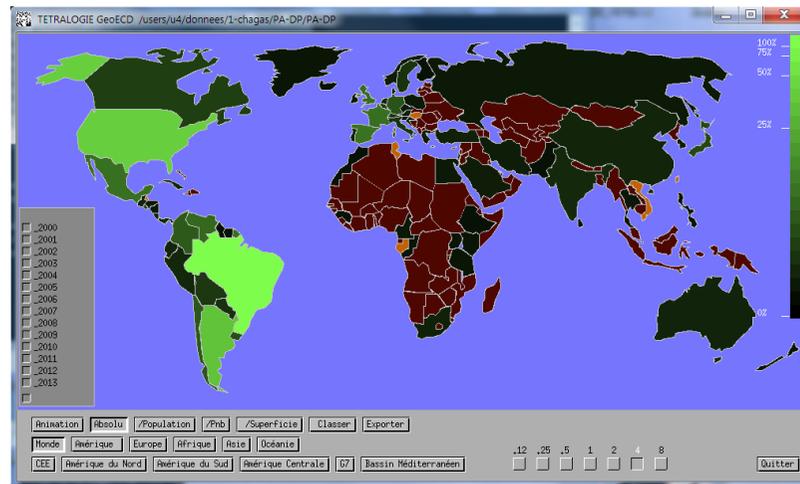


Figure 7 : à gauche répartition toutes périodes confondues, à droite production scientifique sur la dernière année.

4.2 Répartition en fonction du PNB de chaque pays

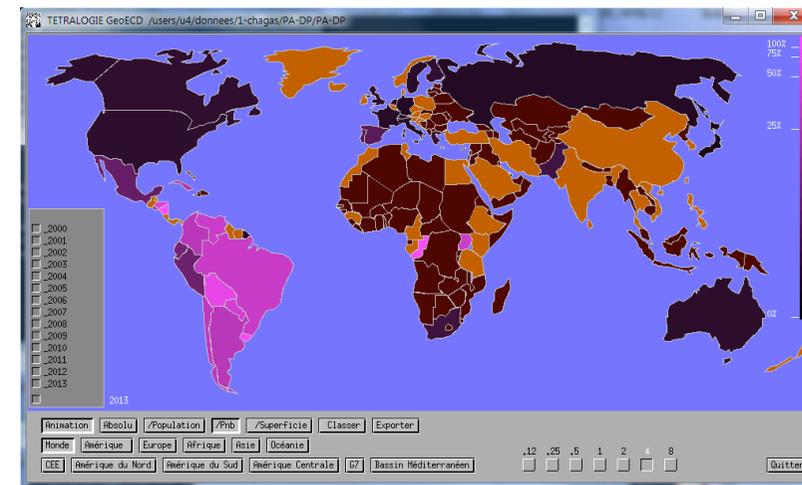
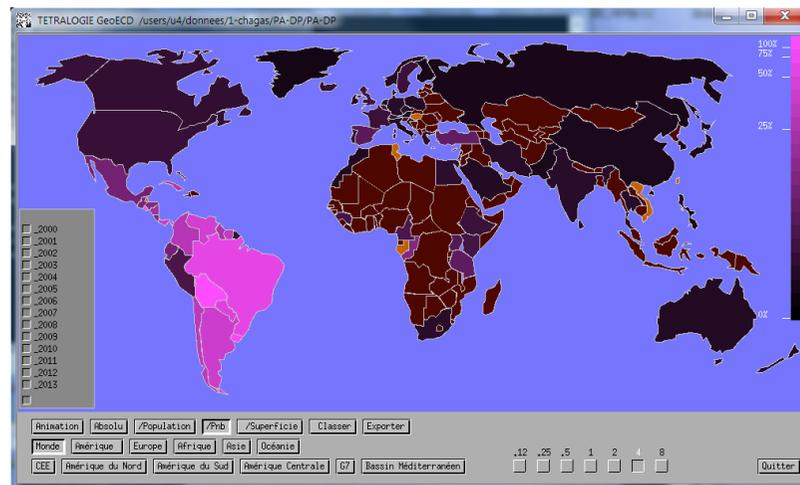


Figure 8 : à gauche répartition toutes périodes confondues, à droite production scientifique sur la dernière année.

4.3 Zoom sur la recherche au Brésil

- 1 540 SAO_PAULO
- 2 458 RIO_DE_JANEIRO
- 3 432 MINAS_GERAIS
- 4 49 DISTRICT_FEDERAL
- 5 47 GOIAS
- 6 40 PARANA
- 7 34 PERNAMBUCO
- 8 27 AMAZONAS
- 9 25 PARA
- 10 23 TOCANTIN
- 11 21 BAHIA
- 12 14 CEARA
- 13 13 SANTA_CATARINA
- 14 10 RIO_SUL
- 15 6 RIO_Grande_Do_Norte
- 16 5 MATO_SUL
- 17 4 RORAIMA
- 18 4 MATO_GROSSO
- 19 4 ESPIRITO_SANTO
- 20 2 AMAPA
- 21 1 SERGIPE
- 22 1 RONDONIA
- 23 1 PIAUI
- 24 1 PARAIBA
- 25 1 MARANHAO

Cette AFC représente la répartition des auteurs suivant leur localisation ou leurs collaborations avec les différents états du Brésil.

Ressortent les auteurs qui ne travaillent qu'avec un seul état ou un groupe d'états et ceux qui animent des collaborations plus larges entre au moins deux groupes d'états comme Barros ou Abad. Les collaborations fortes entre états sont aussi mises en évidence comme Rio Sul et Roraima en haut à droite.

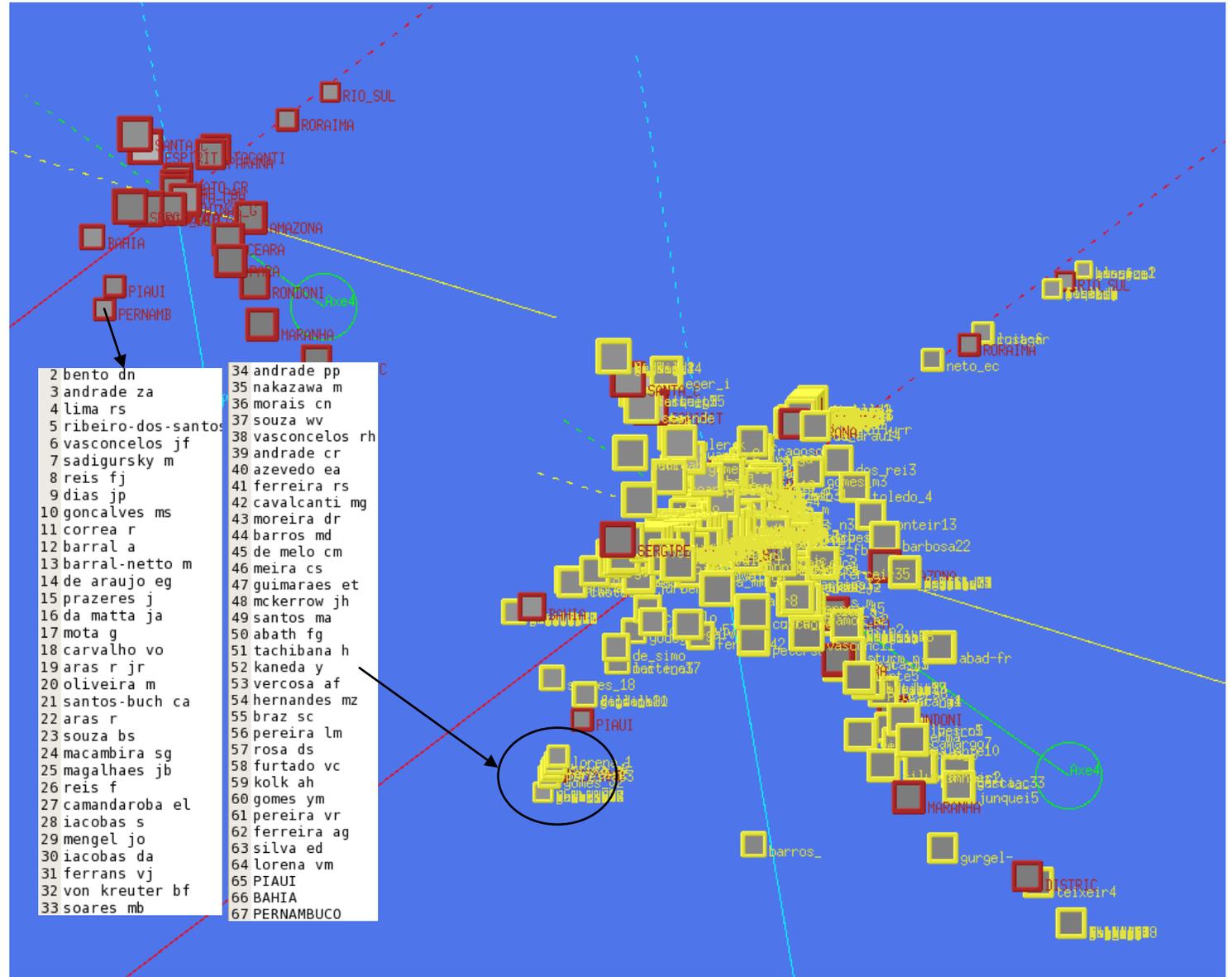


Figure 9 : AFC visualisant les correspondance entre auteurs et états du Brésil : détail sur un état le Permanbuco.

Nous pouvons aussi dresser les cartes de répartition de la recherche pour le Brésil sur l'ensemble des années ou uniquement sur la dernière 2013.



Figure 10 : à gauche répartition toutes périodes confondues, à droite production scientifique sur la dernière année.

5 L'analyse sémantique

5.1 Evolution du domaine et de sa terminologie

Nous pouvons étudier l'évolution du domaine et de sa terminologie en distinguant les termes entièrement nouveaux, qui ne sont présents que dans la dernière période et ceux dont l'utilisation augmente fortement dans cette même période mais qui sont déjà présents dans les périodes précédentes. Les premiers sont concentrés en un seul point sur la carte factorielle de la figure 11 en haut à droite, les autres sont fortement attirés par ce point en fonction de leur taux de présence dans cette période (la plus récente). Via une sélection par lasso, il est possible d'extraire la liste de ces termes et de créer un dictionnaire permettant de détecter l'innovation (cluster sémantiques cohérents) et de l'attribuer à des équipes de recherche, des laboratoires, des journaux, des pays ou des états brésiliens. Cette carte d'identité d'une innovation va nous permettre de la valider et de la contextualiser notamment en la croisant avec tous les autres termes ciblant le domaine.

La même méthode peut être appliquée au texte libre (ici titre et résumé) après détection des termes composés utilisés dans le domaine étudié. Ces termes composés ou multi-termes correspondent non seulement à des mots clés mais aussi se rencontrent dans le texte libre sans qu'ils soient présents dans les thésaurus (Mesh, key words, free terms, index de toutes sortes). Ils représentent bien souvent l'émergence terminologique avant qu'elle soit reconnue comme telle est qu'elle passe dans les thésaurus d'indexation validés par les sages. Il est possible ainsi d'avoir un coup d'avance pour la détection de l'innovation même si elle est à l'état de trace dans le corpus, deux ou trois document suffisent du moment qu'il utilisent un vocabulaire commun et entièrement nouveau.

1 Il y a 68 éléments dans cette liste
2 coinfection
3 real-time polymerase chain reaction
4 eicosanoids
5 phylogeography
6 immune evasion
7 parasite load
8 multilocus sequence typing
9 immunochromatography
10 th17 cells
11 disease resistance
12 euglenozoa infections
13 aptamers, nucleotide
14 receptor cross-talk
15 small molecule libraries
16 thymocytes
17 medical tourism
18 diagnostic tests, routine
19 adaptive immunity
20 nursing assessment
21 hypoglycemic agents
22 phthalazines
23 lysosomal-associated membrane protein 2
24 biosynthetic pathways
25 phosphatidate phosphatase
26 bibliometrics
27 acetylglucosamine
28 myeloid cells
29 respiratory tract infections
30 14-alpha demethylase inhibitors
31 pleural effusion
32 chemistry techniques, synthetic
33 parasitic diseases, animal
34 organophosphonates

35 translational medical research
36 sumoylation
37 nuclear proteins
38 animal experimentation
39 calcium-binding protein, vitamin d-dependent
40 hypertension, pulmonary
41 sedentary lifestyle
42 click chemistry
43 eugenia
44 phosphodiesterase inhibitors
45 insecticide-treated bednets
46 carbamoyl-phosphate synthase (glutamine-hydrolyzing)
47 platelet membrane glycoproteins
48 sequence analysis, protein
49 poverty areas
50 dna methylation
51 chromatography, reverse-phase
52 small-area analysis
53 asymptomatic diseases
54 boron compounds
55 prothrombin
56 receptor, endothelin b
57 microarray analysis
58 oxidative phosphorylation
59 comparative genomic hybridization
60 cervical intraepithelial neoplasia
61 colposcopy
62 trans-splicing
63 hypocreales
64 mite infestations
65 genetic association studies
66 pyrrolidinones
67 blood safety
68 energy transfer
69 soil microbiology

5.3 Thèmes émergents

- 1 il y a 89 éléments dans cette liste
- 2 adenosine deaminase
- 3 antibody affinity
- 4 homeostasis
- 5 host-pathogen interactions
- 6 primate diseases
- 7 diabetes mellitus
- 8 granzymes
- 9 potassium channels
- 10 ofloxacin
- 11 hydroxymethylglutaryl-coa reductase inhibitors
- 12 candida
- 13 3-iodobenzylguanidine
- 14 tablets
- 15 forkhead transcription factors
- 16 thromboxane a2
- 17 oils, volatile
- 18 paracoccidioidomycosis
- 19 neglected diseases
- 20 vaccines, subunit
- 21 membrane potential, mitochondrial
- 22 cytotoxins
- 23 drug monitoring
- 24 protein multimerization
- 25 gene knockout techniques
- 26 family characteristics
- 27 health services accessibility
- 28 lysosomal-associated membrane protein 1
- 29 pemphigus
- 30 adrenergic beta-1 receptor antagonists
- 31 simvastatin
- 32 suppressor of cytokine signaling proteins
- 33 culture media, conditioned
- 34 magnetic resonance angiography
- 35 human activities
- 36 epidemiologic studies
- 37 integumentary system
- 38 insect repellents
- 39 stress, psychological
- 40 patch-clamp techniques
- 41 copper
- 42 cicatrix
- 43 vasoactive intestinal peptide
- 44 transplantation, autologous
- 45 rosaniline dyes
- 46 imines
- 47 hyperthermia, induced
- 48 fetal resorption
- 49 glycerides
- 50 blood component removal
- 51 ascites
- 52 physical conditioning, animal
- 53 paraparesis, tropical spastic
- 54 suburban population
- 55 receptors, complement 3b
- 56 snakes
- 57 catechols
- 58 arthritis
- 59 pulmonary artery
- 60 leucyl aminopeptidase
- 61 weight gain
- 62 electric stimulation therapy
- 63 information services
- 64 chlorides
- 65 atp synthetase complexes
- 66 cyanides
- 67 mouth mucosa
- 68 cholinesterase inhibitors
- 69 leprosy, lepromatous
- 70 thalidomide
- 71 prodigiosin
- 72 massage
- 73 perchloric acid
- 74 hyponatremia
- 75 poultry diseases
- 76 maternal exposure
- 77 chemical precipitation
- 78 ceramides
- 79 enema
- 80 video recording
- 81 megacolon, toxic
- 82 histological techniques
- 83 adenosine triphosphatases
- 84 choroid plexus
- 85 p-glycoprotein
- 86 plasminogen activator inhibitor 1
- 87 fourier analysis
- 88 gastrointestinal tract
- 89 malpighian tubules
- 90 esophageal sphincter. lower

6 Conclusions

Les recherches sur la maladie de Chagas sont en augmentation régulière depuis 1975.

- Principal acteur: le Brésil (Sao Paulo, Rio, Minas Gerais)
- En 2013 il ne reste que 36 pays sur 80
- Nouveaux axes : coinfection, eicosanoids, phylogeography, immune evasion, ...
- Axes en émergence : adenosine deaminas, antibody affinity, ...
- Journaux actuels: mem inst O. Cruz, rev soc bras med trop, mol biochem parasitol, ...
- Les principaux auteurs sont des anciens : présents depuis la 1^o période.
- D'autres émergent bien : Gascon J., Cardinalli-Neto A., ...
- Il y a un certain renouvellement dans les équipes existantes.
- Il reste dans cette étude à travailler sur le texte libre (titre + résumé) et d'extraire les multitermes.
- Il faut détecter le vocabulaire émergent non présent dans le Mesh.
- Et enfin, générer les signaux faibles via les réseaux sémantiques qui les caractérisent.
- Une base de données comme le WoS peut aussi être analysée, avec l'avantage
 - d'une indexation plus homogène : keywords, free terms, ...
 - de l'ensemble des adresses,
 - des citations,
 - des financeurs.

7 Bibliographie

- [1] R. Barros Sampaio, B. Dousset, "Etude comparative entre publications scientifiques et brevets : cas d'une maladie négligée la 'Leshmaniasis'", Colloque VSST'2013, Nancy France, oct 2013. <http://www.xploorew.com/VSSST/Colloque/13-Nancy/Session-8/BARROS-SAMPAIO.pdf>
- [2] J. M. Ramos, G. Gonzalez-Alcaide, M. Bolanos-Pizarro, "Bibliometric analysis of leishmaniasis research in Medline (1945-2010)", Parasit. Vectors, vol. 6, p. 55, mar. 2013.
- [3] A. G. Vasconcellos, C. M. Morel, "Enabling Policy Planning and Innovation Management through Patent Information and Co-Authorship Network Analyses: A Study of Tuberculosis in Brazil", PLoS ONE, vol. 7, no10, p. e45569, out. 2012.
- [4] World Health Organisation (WHO) Consultative Expert Working Group on Research and Development. 2012. "Defining Disease Types I, II and III". http://www.who.int/phi/3-background_cewg_agenda_item5_disease_types_final.pdf
- [5] C. M. Morel, S. J. Serruya, G. O. Penna, R. Guimarães, "Co-authorship Network Analysis: A Powerful Tool for Strategic Planning of Research, Development and Capacity Building Programs on Neglected Diseases", PLoS Negl Trop Dis, vol. 3, no8, p. e501, ago. 2009.
- [6] B. Gay, B. Dousset, "Cartographie de réseaux d'alliances et analyse stratégique", Ingénierie Systèmes Inf., vol. 11, no2, p. 37–51, abr. 2006.
- [7] M. E. J. Newman, "Coauthorship networks and patterns of scientific collaboration", Proc. Natl. Acad. Sci., vol. 101, nosuppl_1, p. 5200–5205, jan. 2004.
- [8] M. E. J. Newman, "The structure of scientific collaboration networks", Proc. Natl. Acad. Sci. U. S. A., vol. 98, no2, p. 404–409, jan. 2001.
- [9] D. J. Watts, S. H. Strogatz, "Collective dynamics of 'small-world' networks", Nature, vol. 393, p. 440–442, jun. 1998.
- [10] J.-P. Benzécri, "Correspondence analysis handbook", New York, Marcel Dekker, 1992.