

AFSCET

Res-Systemica

Revue Française de Systémique
Fondée par Evelyne Andreewsky

Volume 25, Rochebrune 2022

Systèmes complexes ; théorie et pratiques

Res-Systemica, volume 25, article 05

Cartographier le paysage sémantique
de la science des systèmes complexes

Quentin Lobbé, Alexandre Delanoë

10 pages



Creative Commons

CARTOGRAPHIER LE PAYSAGE SÉMANTIQUE DE LA SCIENCE DES SYSTÈMES COMPLEXE

Compte rendu d'atelier cartographique

Quentin Lobbé & Alexandre Delanoë

CNRS- ISC-PIF

Lors des journées scientifiques de Rochebrune 2022, nous avons proposé aux participants de suivre un atelier de cartographie collective centré sur l'exploration d'un corpus de publications liées à la science des Systèmes Complexes. Nous nous sommes servi du logiciel libre d'analyse du langage naturel GarganText³² – un projet logiciel mené par Alexandre Delanoë, et développé à l'Institut des Systèmes Complexes Paris Île-de-France et soutenu par le CNRS.

GarganText permet de reconstruire – suivant diverses dimensions – le paysage sémantique d'un corpus de plusieurs milliers de documents textuels. Le logiciel analyse les interactions locales entre des mots ou des groupes de mots extraits des documents

³²<https://gargantext.org/>

d'origine afin de reconstruire à échelle globale la structure et la dynamique de l'ensemble du corpus. En outre, GarganText dispose de mécanismes collaboratifs permettant d'étudier un corpus en équipe et de manière décentralisée.

Le corpus utilisé pour cet atelier était un corpus de publications scientifiques³³ extraites des bases de données *Web of Science*, *Pubmed*, *Scopus*, *Jstor* et répondant à la requête suivante : ("complex system*" OR "complex adaptive system*" OR "complex network*" OR "complexity science" OR "complex social system*") AND ("network dynamic*" OR "swarm intelligence" OR "multi agent" OR "multiagent" OR "morphogenes*" OR "network analysis" OR "agent based" OR "biological network*" OR "self organi*" OR "econophysic*" OR "sociophysic*" OR "social network*" OR "collective intelligence" OR "collective behavior*" OR "evolutionary comput*" OR "opinion dynamic*" OR "pattern formation" OR "social simulation" OR "distributed system*" OR "distributed cognition" OR "multiscale" OR "multi scale" OR "multifractal" OR "multi fractal" OR "cell* automat*" OR "artificial societ*" OR "morphodynam*" OR "socio-semantic network*" OR "sociosemantic network*" OR "interaction* netw*" OR "nature-

³³Nous avons utilisé les méta-données de ces publications : titre, résumé, auteurs, date de publications, etc.

inspired comput*" OR "open-ended evolution" OR "evolutionary robot*" OR "hypernetw*" OR "nonequilibrium statistical physic*"). Ce faisant, nous avons récolté plus de 32000 publications liées à la science des Systèmes Complexes et publiées entre 1970 et 2022.

Lors de l'atelier, nous avons demandé à l'ensemble des participants d'annoter certaines publications de ce corpus suivant leurs domaines de recherche respectifs. Ce travail d'annotation a permis d'enrichir une première liste de mots / groupes de mots extraits automatiquement par GarganText en amont de l'atelier. Nous nous sommes ensuite servi de ce vocabulaire nouvellement créé afin de générer deux visualisations interactives : une carte sémantique et une phylométrie.

Ces deux représentations annotées par des experts du domaine, participent aujourd'hui d'une meilleure compréhension de la science des Systèmes Complexes dans son ensemble. Cet atelier n'est pourtant qu'une première étape, le travail de cartographie est amené à se poursuivre par l'enrichissement du corpus et l'approfondissement de l'étape d'annotation en ouvrant la carte à l'ensemble de la communauté scientifique Systèmes Complexes.

1) Carte sémantique

Dans GarganText la carte sémantique n'est pas un nuage de points dans un espace métrique a priori comme peuvent l'être des projections issues de réductions de dimensions du type «Analyses factorielles» de correspondances ou «analyse en composantes principales». Dans notre cas, la carte sémantique est représentée par un graphe, seules les relations de liens sont significantes et non pas les distances dans l'espace de visualisation du graphe. Les nœuds du graphe sont les termes sélectionnés par les analystes de manière collaborative. Les relations entre les termes sont construites à partir de leurs co-présences contextuelles dans les unités élémentaires, c'est-à-dire les méta-données (titre et résumé des publications scientifiques du corpus considéré).

Deux types de graphes sont proposés dans l'interface GarganText selon le choix de la mesure de similarité choisie. Chacune des mesures possède des propriétés interprétatives différentes que nous allons illustrer avec le corpus Systèmes Complexes.

Le graphe avec une mesure de similarité d'ordre 1 est dite conditionnelle et probabiliste étant donné le contexte d'apparition d'un terme en sachant l'autre. *La mesure de similarité d'ordre 2 dite distributionnelle évalue les profils d'interchangeabilité d'un terme étant donné l'ensemble de ses voisins dans les contextes élémentaires d'apparition.*

SYSTÈMES COMPLEXES : THÉORIE & PRATIQUES

- «complex networks», adaptation, évolution
- les modèles agents

Ces thèmes apparaissent récurrents dans l'état de la littérature systèmes complexes mais ils représentent aussi surtout les termes choisis par l'équipe d'analystes dans l'interface collaborative GarganText : la carte représente ainsi l'objet d'intérêt des participants.



Figure 2 : carte sémantique des méta-données de 32000 publications liées à la science des systèmes complexes. Carte générée par GarganText à partir d'une mesure de similarité d'ordre 2

Par contraste, le graphe d'ordre 2 (voir *Figure 2*) fait apparaître les concepts mobilisés par discipline, disciplines qui ne sont pas pourtant étiquetées comme telles a priori ; elles émergent donc des propriétés du graphe :

- Les systèmes complexes en biologie par leurs interactions cellulaires ;
- La dynamique des réseaux appliqués aux éco-systèmes et à l'économie ;
- Les applications en sciences sociales et leurs disciplines comme la psychologie et la prévalence des réseaux temporels comme outil d'analyse ;
- Les Systèmes sociaux, apprenants ou morphogénétiques avec les interactions des agents et l'usage de l'intelligence artificielle ;
- Enfin les effets d'échelles ou l'auto-organisation sont aussi des thématiques transverses que l'on voit émerger de cette carte.

Ainsi, les 2 types de graphes font émerger par contraste différents niveaux d'interprétation : soit au niveaux des sujets d'étude en tant que tels soit les disciplines avec leurs concepts privilégiés associés.

2) Phylométrie

Une phylométrie est un objet scientifique utilisé pour reconstruire l'évolution de la structure sémantique d'un corpus de documents textuels [1]. C'est un réseau d'héritage d'éléments de savoir qui prend graphiquement la forme d'un arbre généalogique d'idées ou de concepts extraits du corpus étudié [2].

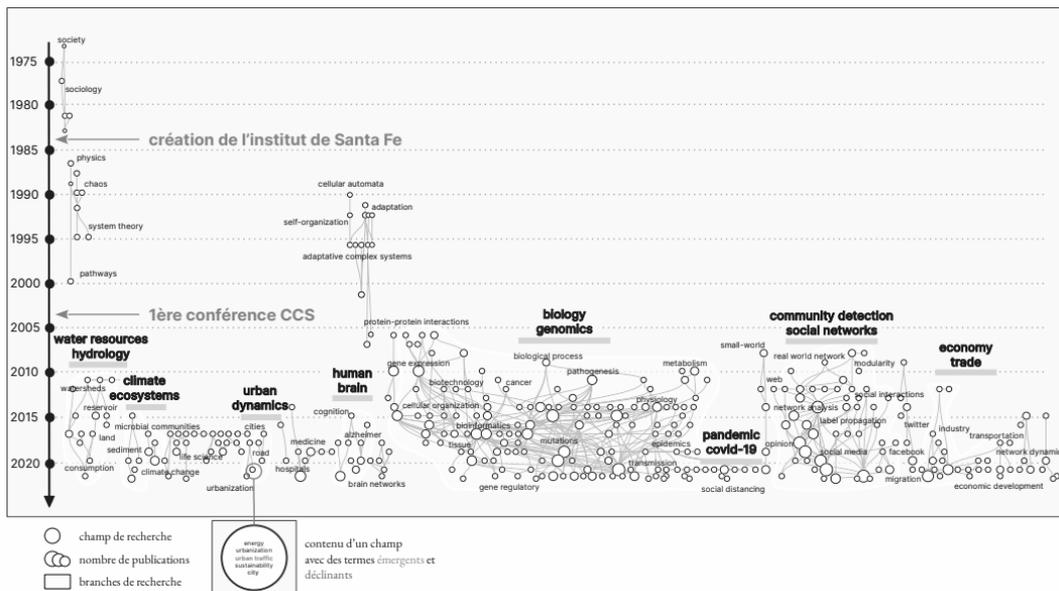


Figure 3 : phylométrie des méta-données de 32000 publications liées à la science des systèmes complexes. Phylométrie générée par GarganText.

Appliquée à une collection de publications scientifiques, une phylométrie révèle l'évolution du domaine ciblé sous la forme d'une multitude de branches de connaissances explorées au cours du temps par des générations de chercheurs. Chaque branche regroupe un nombre fini de champs de recherche eux mêmes

constitués de mots / groupes de mots fréquemment utilisés ensemble au sein d'une ou plusieurs publications et sur une période précise. Ces champs sont liés entre eux par proximité sémantique.

La Figure 3 est un export de la phylométrie interactive créée lors de notre atelier. Nous avons ici choisi de mettre en valeur (aplats jaunes) certaines branches qui nous semblent importantes pour comprendre l'évolution de la science des Systèmes Complexes. De la même manière, nous avons mis en avant certains des termes émergents les plus fréquemment utilisés dans ce corpus.

La Figure 3 retrace ainsi l'évolution de la Science des Systèmes Complexes à travers différents moments clefs de son histoire : les travaux pionniers dans les années 80 et 90 portés principalement par des physiciens et des mathématiciens, l'importance de la création de l'institut de Santa Fe en 1984 et plus encore de la structuration de l'ensemble de la communauté Systèmes Complexes autour des conférences *CCS* dès 2004, l'ouverture de cette science à d'autres domaines (biologie, géographie, économie, etc.) dans le courant des années 2000 et par là même son caractère intrinsèquement inter-disciplinaire.

Références

- [1] Chavalarias, D., Lobbé, Q. & Delanoë, A. Draw me Science - multi-level and multi-scale reconstruction of knowledge dynamics with phylomemories. *Scientometrics* (2021)
- [2] Lobbé, Q., Delanoë, A. and Chavalarias, D. Exploring, browsing and interacting with multi-level and multi-scale dynamics of knowledge, *Information Visualization*. (2021)