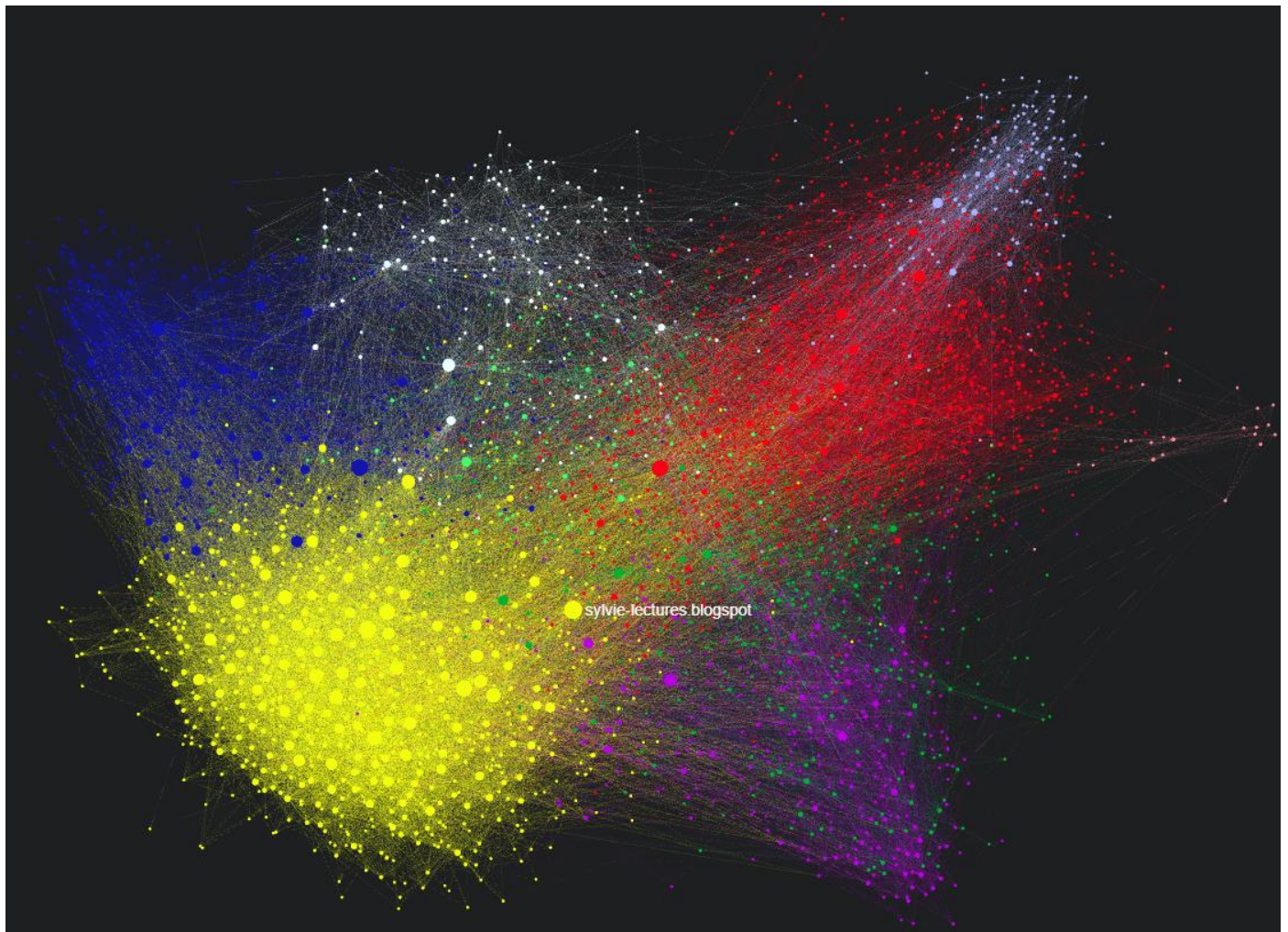


Auteurs : Crépel, Maxime (Médialab (Sciences Po)); Boullier, Dominique (CEE (Sciences Po)); Le Béhec, Mariannig (CEREGE (Université de Poitiers))

Publié: 2010

# Le Web du livre en France en 2010

A partir de cette cartographie, on peut observer que les formes de conversation autour du livre occupent un espace important sur le web. Cette cartographie permet de rendre compte du positionnement des réseaux de recommandation et de circulation des livres, ainsi que les formes d'association des sites par domaines littéraires. La cartographie est constituée de trois régions principales dans le graphe : la « blogosphère littéraire », les « communautés thématiques hétérogènes », et les « métiers du livre et l'infocom » qui se composent de types d'acteurs, de genres littéraires différents, et présentent des structures distinctes.



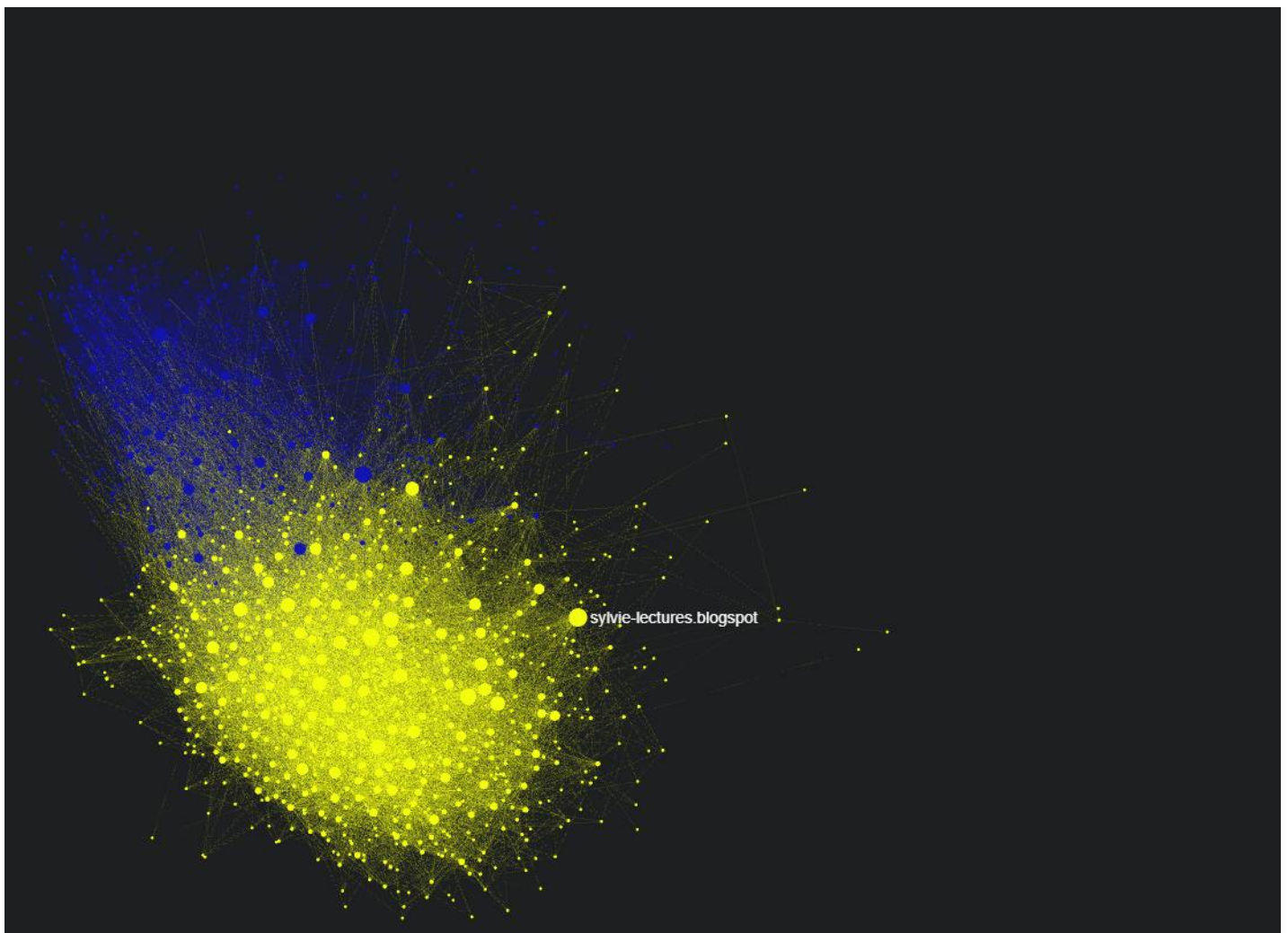
## Méthodologie

La cartographie du Web du livre en France a été constituée par une méthode d'extraction de données (crawl) à partir d'un corpus de départ constitué durant l'étude qualitative des réseaux de circulation des livres et de critique littéraire. Le corpus de départ ciblait un ensemble de 783 sites (Des sites d'acteurs grand public : cercles de lecture, blogs, réseaux sociaux, forums ; Des sites web des métiers du livre papier et numérique : auteurs, éditeurs, libraires, événements littéraires, critiques professionnels, revues littéraires ; Des acteurs institutionnels : Associations, syndicats, bibliothèques, organismes d'état).

L'extraction de données a permis de récupérer un ensemble de 7187 sites reliés par un lien hypertexte à ce corpus de départ (n+1) et possédant au minimum deux liens avec celui-ci. Après un travail de nettoyage et de filtrage du graphe initial (sites web hors de la thématique, sites non français, sites inactifs ou faiblement connectés), le graphe final est constitué de 3289 sites web reliés entre eux par un total de 52884 liens hypertextes. La taille des nœuds dépend du degré, c'est à dire du nombre de liens que possède un nœud, de sorte que les nœuds les plus grands sont ceux qui possèdent le plus de liens.

Deux opérations conjointes ont été réalisées afin d'analyser le graphe :

- Une première a consisté à visualiser le graphe (logiciel Gephi, algorithme de spatialisation force-atlas / [www.gephi.org](http://www.gephi.org)) et à opérer une détection automatique de communauté (clustering) qui repère différentes zones dans lesquelles les sites web sont plus fortement connectés entre eux.
- La seconde tâche réalisée a été une opération de catégorisation « à la main » de tous les sites du graphe et une catégorisation systématique à partir de mot-clés. Le travail de catégorisation des sites web permet de qualifier les clusters détectés automatiquement de manière algorithmique en sous-ensembles thématiquement cohérents.



## Blogosphère littéraire

Le premier sous-ensemble qu'est la « blogosphère littéraire », à gauche sur le graphe, est constituée des deux groupes « Littérature Générale & Etrangère » (en jaune) et « Littérature et Fantasy » (en bleu). Il a comme caractéristiques principales d'être composé quasi exclusivement de blogs (80%) de réseaux sociaux littéraires et de ne compter aucun site web d'acteur en charge de la production ou la commercialisation de livres. Sur le plan de la structure, il existe une très forte densité de liens (beaucoup d'interconnexion) entre sites, un diamètre faible (7 liens hypertextes maximum pour

aller d'un site web à un autre) et le degré moyen (moyenne des liens hypertextes qui relient un site à ces voisins) de 37,5 y est élevé par rapport au reste du graphe.

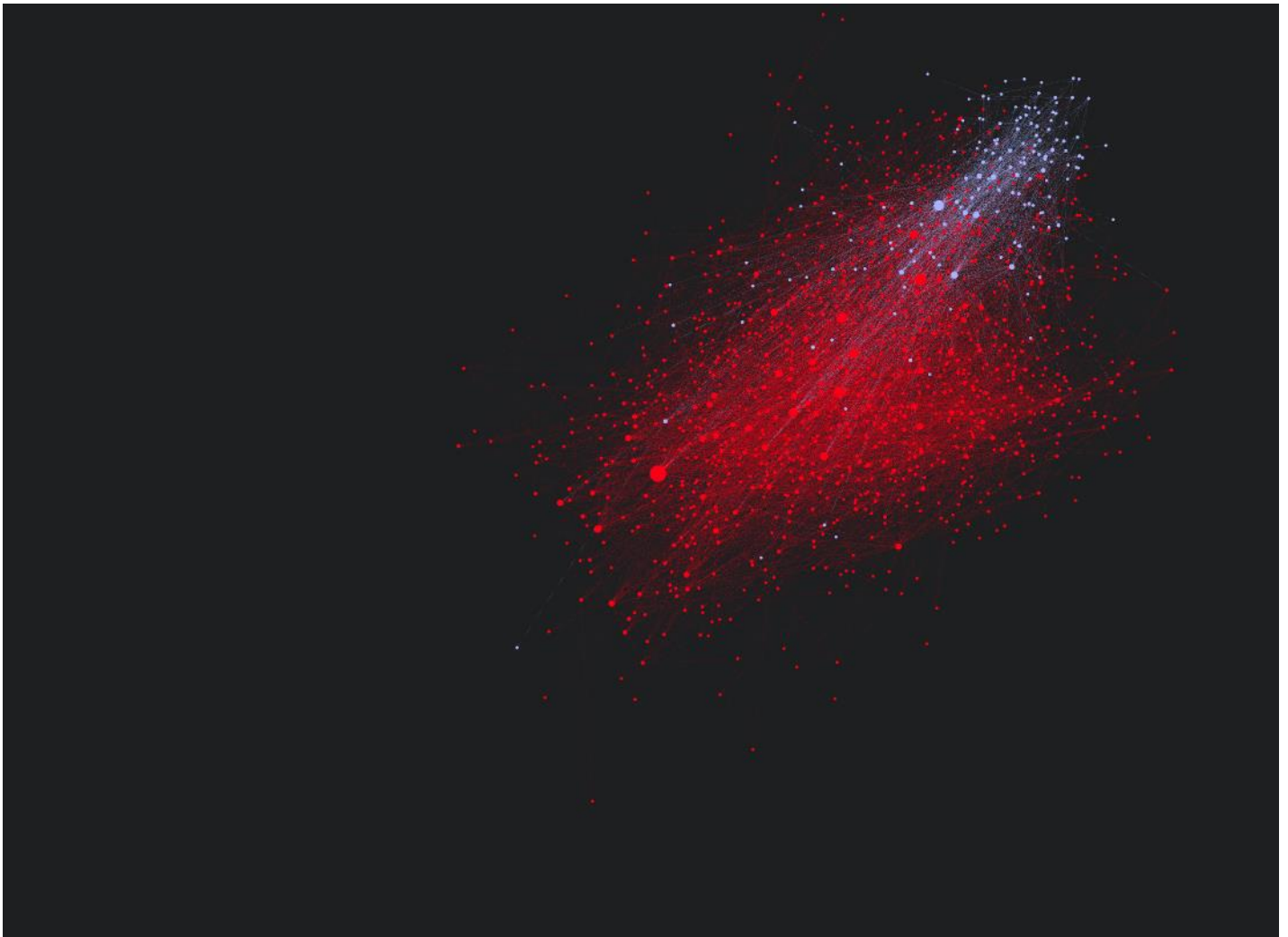


## Communautés thématiques hétérogènes

Le second sous-ensemble, au centre de la cartographie, qualifié de « communautés thématiques hétérogènes » est constitué des groupes « BD » (en vert), « Jeunesse » (en violet), « SFFF » (en blanc), « Polar » (en vert clair), et « Comics » (en rose). La particularité de ces communautés est qu'elles sont constituées d'un ensemble d'acteurs plus hétérogène que les deux autres régions du graphe. En effet, elles comptent entre 15% à 35% de blogs, de 15% à 20% sites de maison d'édition, et de 5% libraires. D'autres acteurs comme les illustrateurs ou les auteurs y sont également très présents. Il s'agit donc de collectifs moins importants en terme de nombre de sites web, moins connectés entre eux (degré moyen de 12 liens) mais qui sont très thématiques par genres littéraires et qui se composent de l'ensemble de la chaîne des acteurs du livre et des lecteurs.

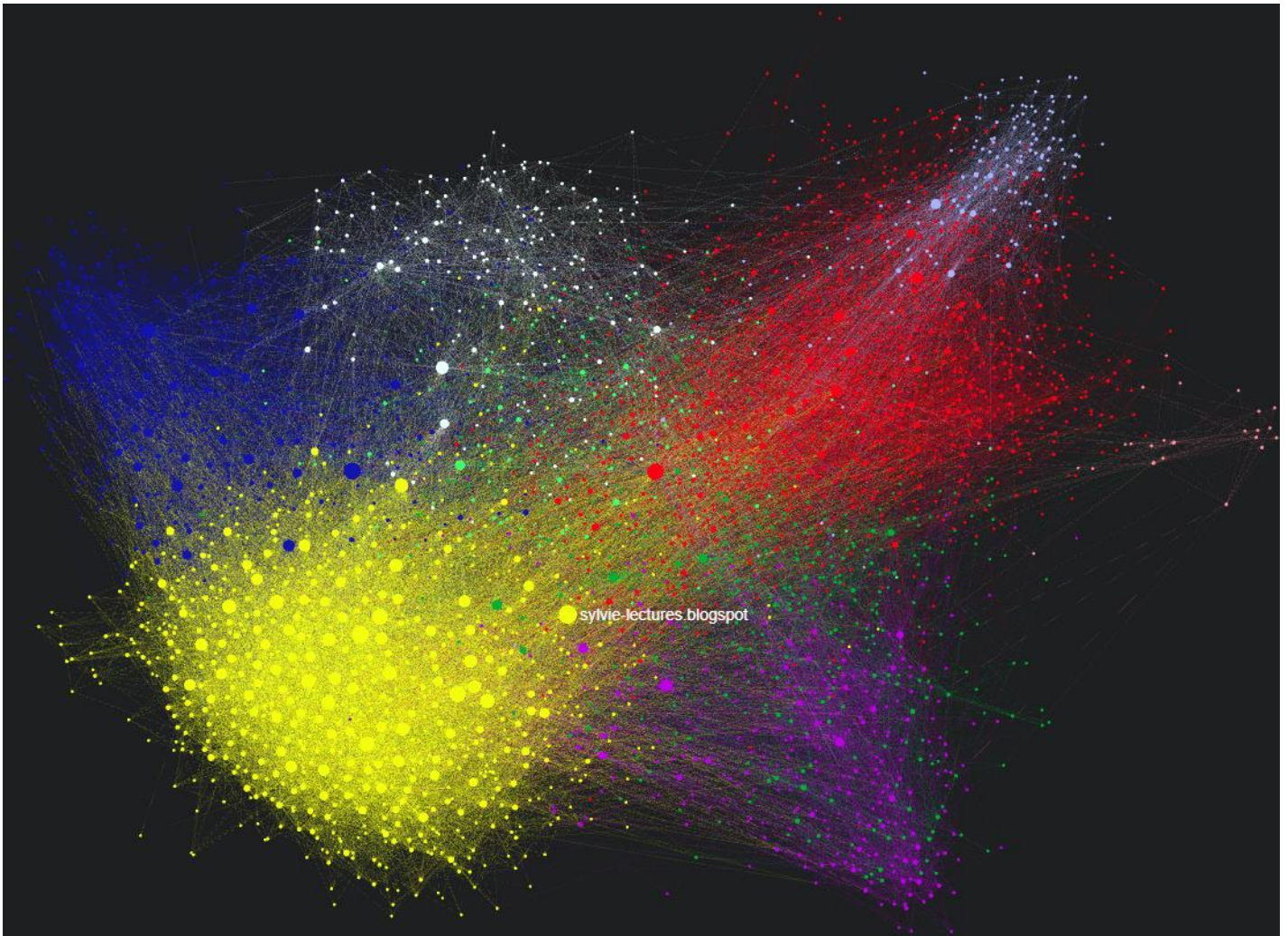
## Métiers du livre et Infocom

Le troisième sous-ensemble, « métiers du livres et l'infocom », à droite sur le graphe est composé des « Métiers du livre & Médias » (en rouge) et de « l'Infocom » (en bleu clair). Il est constitué de deux fois plus d'éditeurs (21%) et de libraires (9,5%) que dans le reste du graphe, on y constate aussi une forte concentration des acteurs du livre numérique (trois fois plus que dans le reste du graphe), ainsi qu'une forte présence de la presse littéraire et généraliste, des syndicats et des institutions liés au livre. Le diamètre y est plus élevé (12 liens hypertextes maximum pour aller d'un site web à un autre) et une densité beaucoup plus faible (0,0019) qui montre une moins forte interconnexion entre les sites que dans le reste du graphe.



## Séparation entre critique et circulation

Cette cartographie permet de rendre compte que l'espace numérique que représente le web est largement investi par des réseaux blogueurs littéraires qui constituent des « communautés » autour de genres particuliers de lecture. Ces réseaux sont des lieux de conversation autour des livres qui prennent la forme de commentaires et de notes de lecture plus ou moins formalisées. Elle nous permet aussi de voir que si le web se structure autour de réseaux d'intérêts distincts, ces réseaux ne présentent pas les mêmes caractéristiques sur le plan de leur structure et que leur positionnement nous indique une importante séparation entre les réseaux de blogueurs d'une part, et les réseaux des métiers du livre d'autre part. Cette séparation ne veut pas dire qu'il n'existe pas de liens entre ces différents sous-ensembles mais que les liens qui les unissent sont assez limités. On pourrait en conclure qu'à l'exception des « réseaux thématiques hétérogènes » qui se constituent comme des écosystèmes riches en terme de variété d'acteurs autour de certains genres, il existe une séparation marquée entre les lieux qui entretiennent les formes conversation-livre et les lieux de création, de production, de diffusion de l'objet-livre. Enfin les sites web exclusivement dédiés au livre numérique, tels qu'ils apparaissent sur cette cartographie de 2010, ne représentent que 2% des sites qui composent l'ensemble du graphe.



## Credits

Cartographie réalisée au médialab de Sciences-po par Dominique Boullier, Maxime Crépel, Mariannig Le Béhec dans le cadre dans le cadre du projet SOLEN. © 2011

Interface web par Guillaume Plique.

Site optimisé pour *Chrome*.

Pour citer cet article: Crépel M., Boullier D., Le Béhec M., "Le Web du livre en France en 2010", CEE, médialab, Sciences-po, SOLEN - projet FUI, 2010, <http://www.medialab.sciences-po.fr/publications/webdulivre/>