

# Gephi et l'analyse de données pour le SEO (2ème partie)



Par Daniel Roch

<b>Domaine :</b>	<b>Recherche</b>	<b>Référencement</b>
<b>Niveau :</b>	Pour tous	<b>Avancé</b>

*Gephi est un logiciel dont on entend parler régulièrement pour analyser la structure d'un site web. En réalité, Gephi est un logiciel d'analyse capable de traiter et de visualiser n'importe quel type de données, permettant ainsi d'approfondir notre analyse en référencement naturel. Nous allons donc voir dans cet article en deux parties comment ce logiciel fonctionne, ses qualités (mais également ses défauts) et comment on peut en tirer profit dans le cadre d'une stratégie SEO. Le mois dernier, nous avons abordé l'utilisation du logiciel Gephi. Puissant mais complexe, ce dernier permet d'analyser et surtout de visualiser des données. Dans cette seconde partie, nous allons donc aborder des cas concrets d'utilisation du logiciel dans une optique de référencement naturel.*

## **Analyser les silos de son site**

Il s'agit sûrement ici de la première utilisation pertinente que l'on peut faire avec Gephi. L'idée est de pouvoir visualiser si, oui ou non, le maillage interne est optimisé et si l'on a réussi à créer des sections cohérentes.

Pour faire cela, le premier moyen est de regarder le rendu initial que l'on a réalisé pour voir si des grappes de nœuds se démarquent ou non. La problématique est que cela est souvent assez difficile à visualiser.

Nous utiliserons donc la méthode dite de la modularité. Cet algorithme sert à déterminer et mettre en avant des groupes dans un tableau de données. Théoriquement, si les silos sont bien conçus, l'algorithme va les détecter automatiquement.

1<sup>ère</sup> étape : calculez la modularité. Pour cela, utilisez dans la colonne de droite l'onglet « Statistiques », puis cliquez sur « Exécuter » pour la ligne « Modularité ».

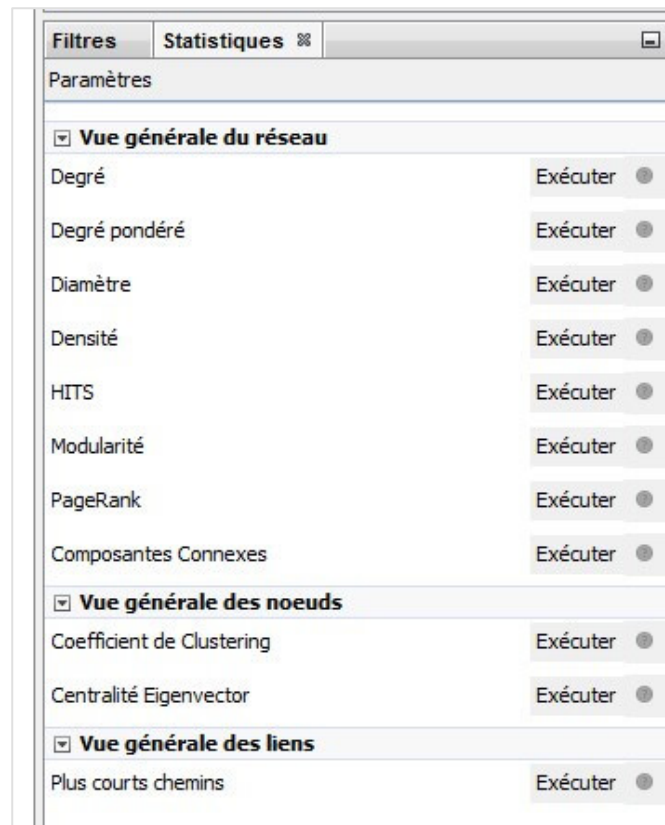


Fig. 1. L'onglet de statistiques

Dans la fenêtre qui s'ouvre, cochez « Aléatoire » et « Utiliser le poids ».

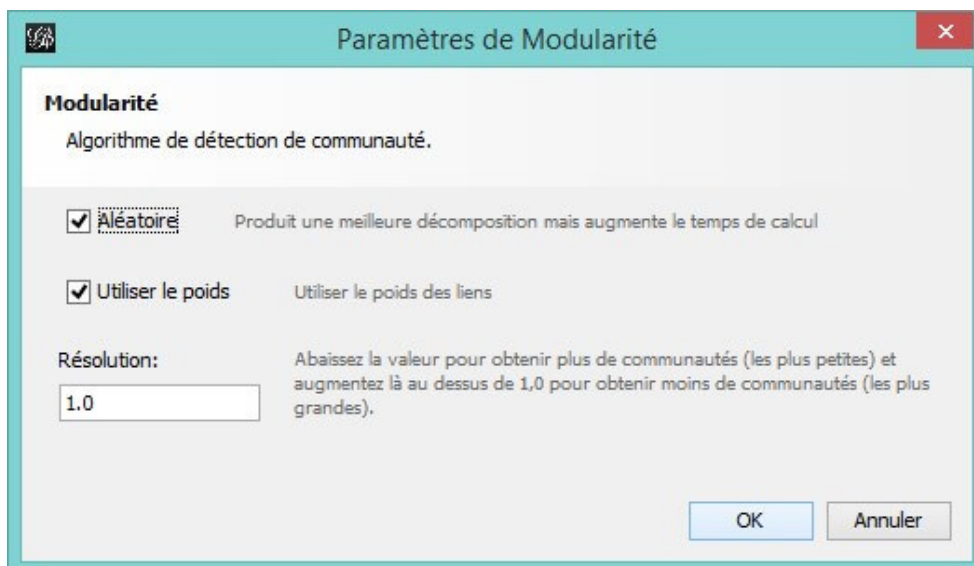


Fig. 2. Le paramétrage de la modularité

Une fois cette étape réalisée, la donnée « Modularité » est prête à l'emploi. Dans la colonne de gauche, sélectionnez dans « Aspect » :

1. l'onglet « Nœuds » ;
2. L'icône de couleur ;

3. Dans la liste déroulante, l'attribut « Modularité ».

Cliquez enfin sur « Appliquer ». Si les silos sont mal conçus, le graphique ressemblera à un sapin de Noël. Si à l'inverse ils sont bien créés, le logiciel Gephi va colorer plus ou moins parfaitement vos différentes sections de manière homogène, comme sur la figure 3.

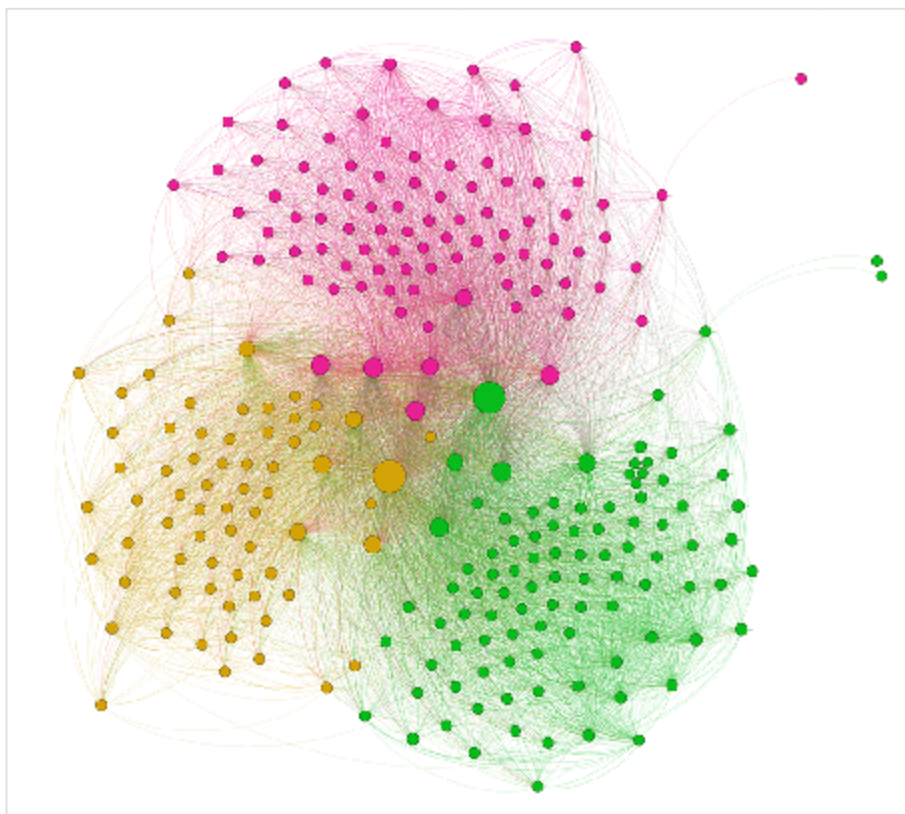


Fig. 3. Un exemple de rendu d'un site ayant 3 silos.

## ***Silos et liens réciproques pertinents***

Tout silo est améliorable. Si on prend l'exemple précédent, le site est bien conçu car trois sections sont visibles. Le souci est que dans ces sections, l'information n'est pas hiérarchisée (page principale, pages secondaires, pages de troisième niveau...). On peut donc améliorer notre travail.

Pour cela, on peut partir du dernier graphique (fig.3) pour chercher les endroits où notre hiérarchie est mauvaise, mais ce n'est pas forcément le plus efficace. Nous allons donc utiliser une autre analyse : les liens réciproques entre les pages (A faisant un lien vers B qui fait également un lien vers A).

Normalement, chaque silo doit contenir un grand nombre de liens réciproques internes, et peu de liens réciproques avec les autres silos. Et dans chaque silo, le nœud principal doit être celui faisant le plus de liens réciproques. Pour cette analyse, nous utiliserons cette fois-ci les filtres de la colonne de droite. On peut ainsi filtrer visuellement tous les nœuds

et les liens selon de très nombreux paramètres dans la partie « Bibliothèque », et les filtres actuellement appliqués seront visibles dans la section « Requêtes ».

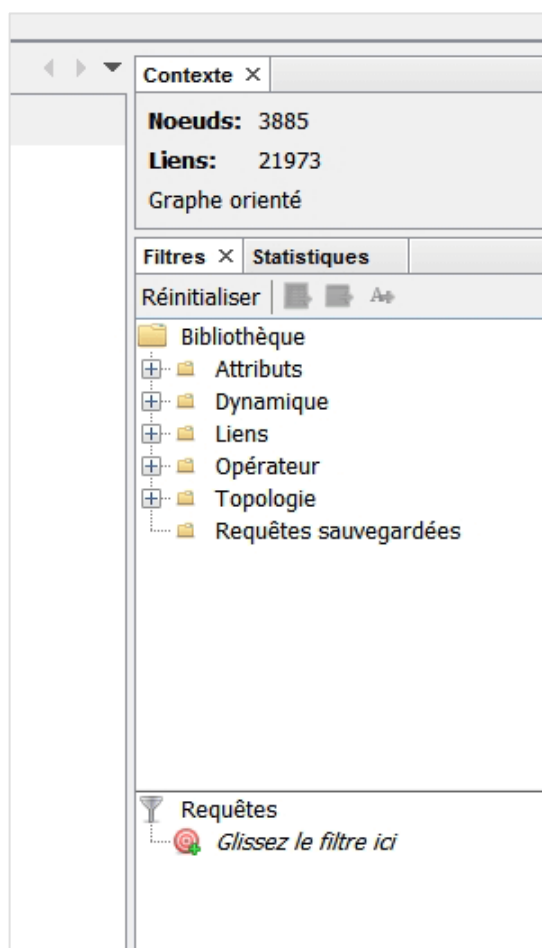


Fig 4. Les filtres dans Gephi

Dans « Liens », faites un double clic sur « Mutual Edge », et cliquez sur « Filtrer » en bas à droite. Cela peut ainsi vous donner ce type de rendu où l'on peut voir tous les liens réciproques, dont :

- Ceux dans un silo (que l'on doit conserver) ;
- Ceux entre les silos (que l'on doit réduire).



Fig. 5. La visualisation des liens mutuels

On voit d'ailleurs que la page située au centre est la page en couleur est celle avec le plus de liens réciproques. Il s'agit de la page « Plan du site », preuve d'un potentiel maillage trop important vers cette URL.

### **Analyser la répartition des pages populaires**

Il est également possible de vouloir faire ressortir le nombre d'impressions de chaque nœud dans les moteurs de recherche. Pour rappel, cette donnée a été obtenue via la connexion à l'API Search Console dans Screaming Frog Spider SEO.

Dans cet exemple, nous avons coloré les nœuds selon le nombre d'impressions, tandis que la taille du nœud correspond à son degré interne. Ainsi, si le rond n'est pas coloré (peu d'impressions) mais que ce dernier est grand, cela veut dire que cette URL a une place importante dans la structure du site, mais qu'elle n'est pas visible dans les moteurs de recherche. Ce serait donc une URL importante à travailler en priorité au niveau de son contenu, ou alors on devrait réduire le nombre de liens vers cette URL.

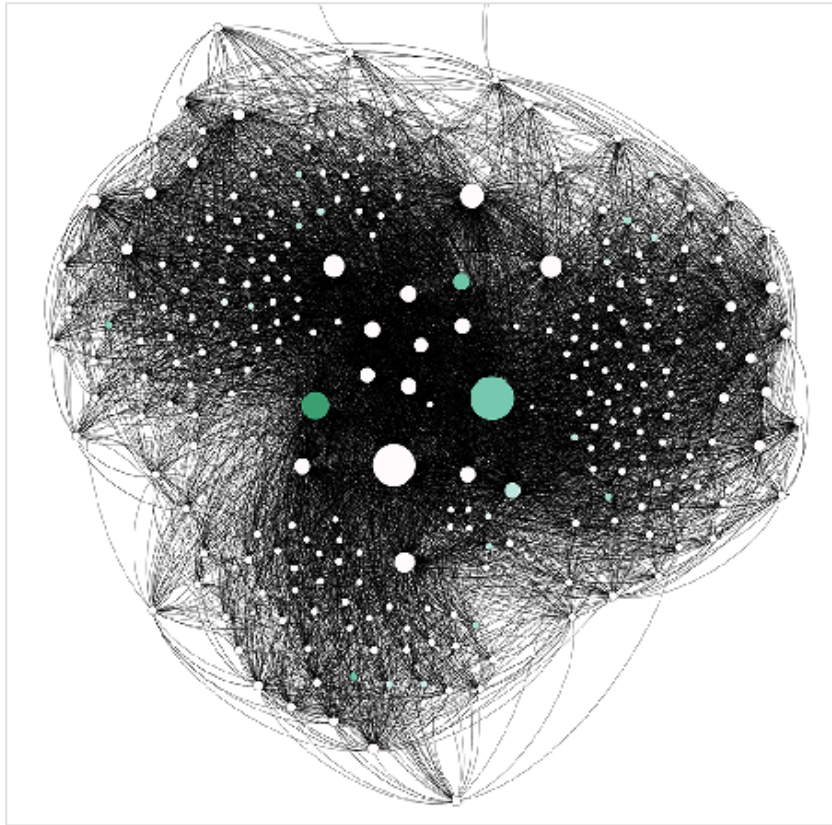


Fig. 6. Le croisement entre l'importance du nœud dans la structure et sa visibilité

D'ailleurs, concernant la taille des nœuds, sachez que vous avez plusieurs solutions possibles. Gephi vous propose en effet d'utiliser différents attributs :

- Le degré, c'est-à-dire un calcul brut du nombre de liens entre les nœuds ;
- Le PageRank, qui reprend le même concept que le degré en donnant plus de poids à la place du nœud dans la structure ;
- Eventuellement le Trust Flow et le Citation Flow de l'outil Majestic.

Pour les deux derniers, l'information vient de l'import du fichier CSV. Pour le PageRank, faites attention car il ne s'agit pas du PageRank de Google, mais d'une formule algorithmique différente. D'ailleurs, pour l'utiliser, vous devez d'abord calculer ce PageRank dans la colonne de droite.

### ***Faire ressortir les contenus inutiles ou en erreur***

Autre exemple dans la figure 7 : faire ressortir les pages en erreur ou inutiles (erreur 404, redirections 301...). On peut ainsi voir où se situent les URL qui polluent le référencement global du site analysé :

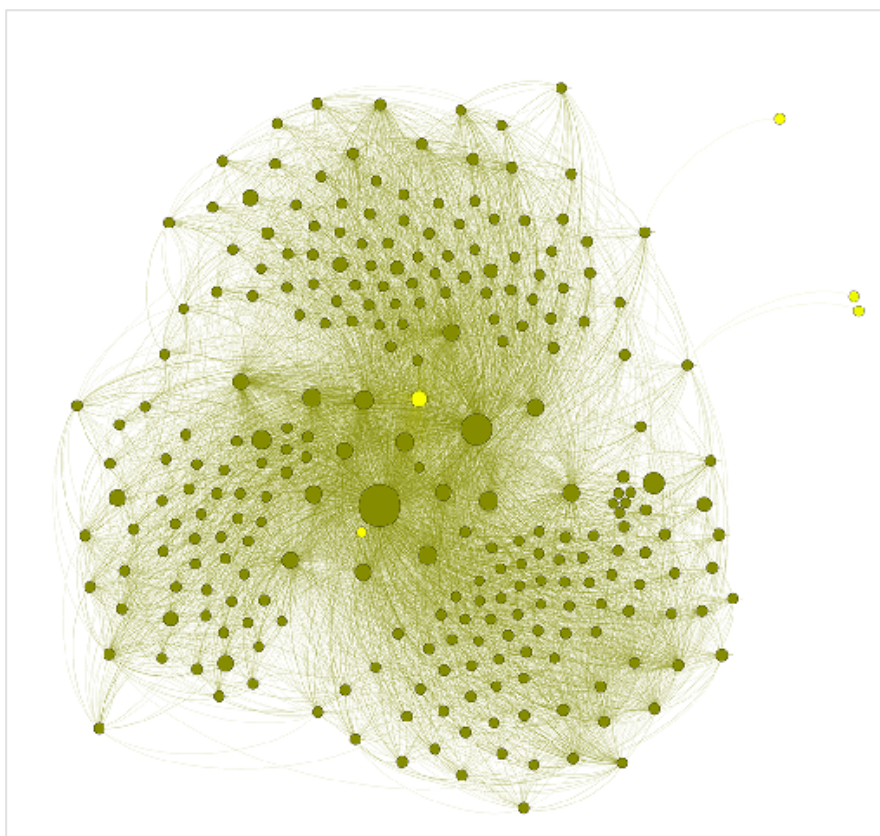


Fig. 7. Les URL malsaines apparaissent en jaune.

## ***Faire des test SEO***

Gephi peut également être utile pour faire des tests de référencement naturel, et notamment pour voir des corrélations ou non entre certaines données. Prenons un exemple :

*Le nombre de mots de mes contenus a-t-il une influence sur les backlinks obtenus, et donc sur le Trust Flow ?*

Pour répondre à la question, nous donnons donc à chaque rond une taille proportionnelle au Trust Flow, et nous les colorons de plus en plus foncé en fonction du nombre de mots présents :

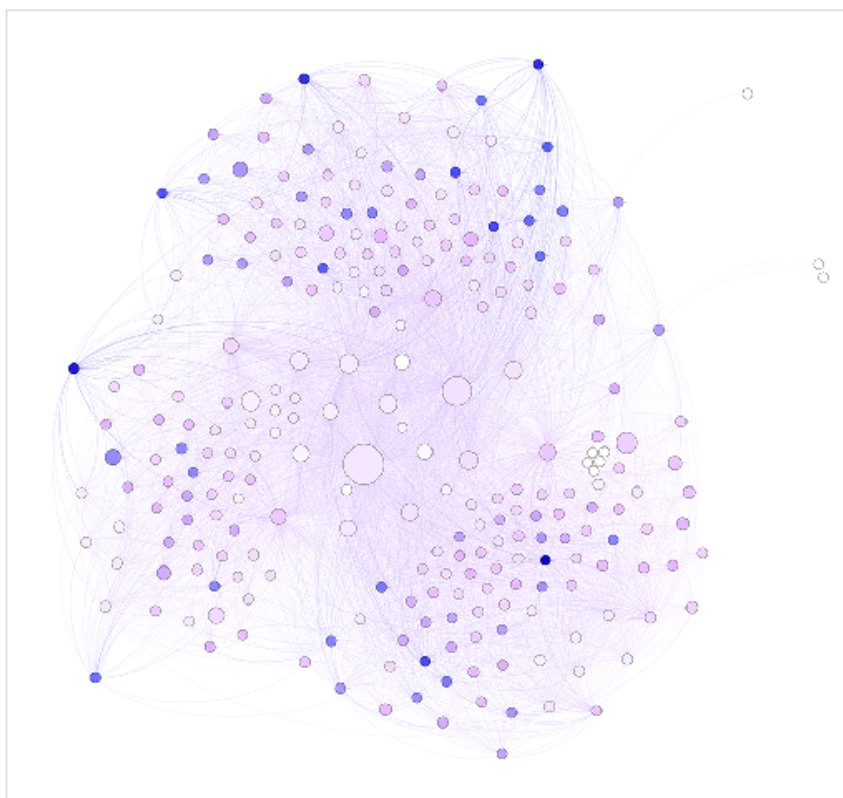


Fig. 8. Y a-t-il corrélation entre Trust Flow et nombre de mots dans le contenu ?

Le résultat est simple : il ne semble pas y avoir de corrélation entre ces deux données car dans le cas contraire, nous aurions des ronds foncés de grande taille.

### **Visualiser les domaines référents de chaque URL de mon site**

Autre exemple : nous pourrions vouloir visualiser les pages de notre site ayant peu de domaines référents afin de mieux rediriger notre travail de netlinking. Pour cette analyse, nous devons utiliser une fois de plus nos filtres, avec cette fois-ci les « Partitions ».

Une fois la partition cochée sur le paramètre « Nombre de domaines référents » (ou « RefDomains » en fonction de comment vous l'avez appelé dans le fichier CSV), nous allons pouvoir cocher ou décocher les chiffres qui nous intéressent. En bas à droite, juste avant le bouton « filtrer », vous verrez en effet des cases pour n'afficher que les nœuds ayant un nombre déterminé de RefDomain. Dans notre exemple, nous avons coché toutes les cases ayant un RefDomain supérieur ou égal à 10.



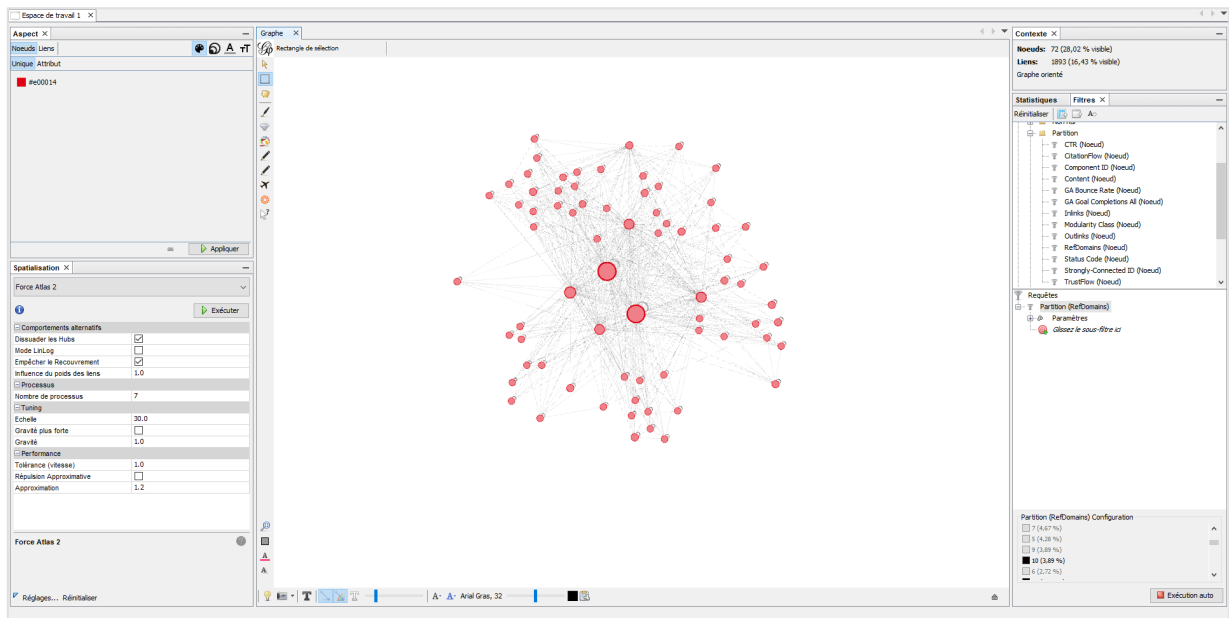


Fig. 9. Un exemple d'utilisation de filtres

Vous pouvez alors sélectionner tous les nœuds dans la fenêtre centrale, puis à gauche dans « Aspect » vous pouvez mettre en place une coloration unique de tous les nœuds. Vous pourrez ensuite cocher « Exécution auto » en bas à droite pour désactiver le filtre, tout en ayant mis en couleur les ronds qui correspondaient à votre filtre.

Dans la figure 10, nous colorons ainsi en rouge foncé les URL ayant un grand nombre de domaines référents.

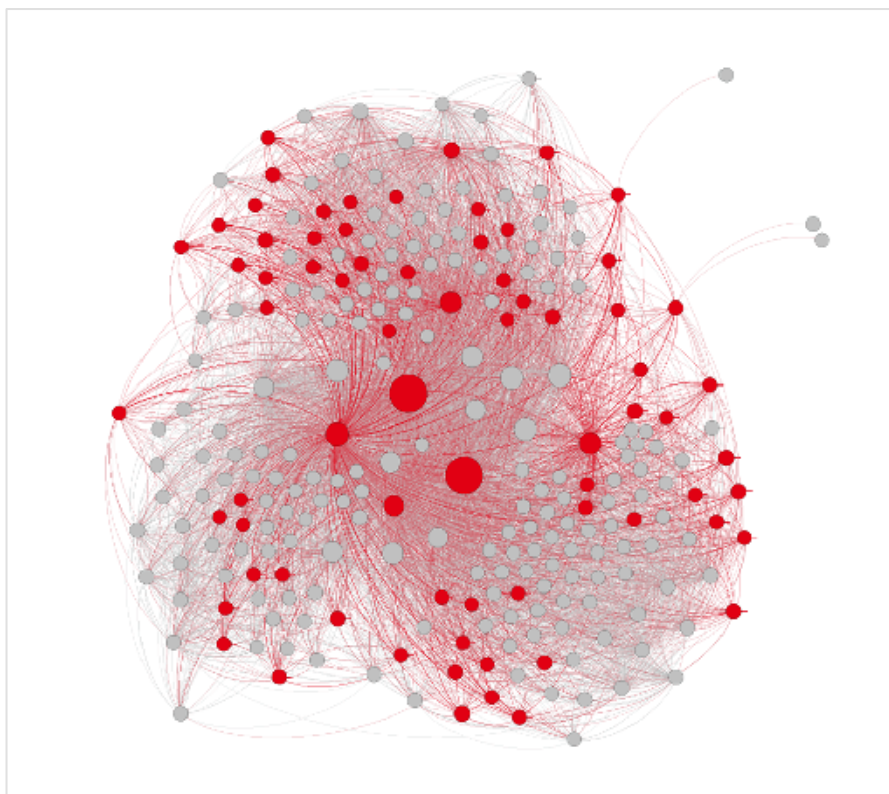


Figure 10. Les nœuds en rouge foncé reçoivent suffisamment de domaines référents

## **Visualiser les liens de chaque URL de mon site**

Il est également possible d'analyser et de mieux visualiser le nombre de domaines référents pour chaque URL, et ainsi savoir vers quelles pages nous devrions créer des liens depuis de nouveaux domaines.

Pour cela, une nouvelle étape d'importation a été nécessaire. Sur Majestic SEO, nous allons exporter la liste des liens vers le domaine analysé. A partir de ce fichier, nous allons donc créer deux nouveaux fichiers :

- Un fichier de nœuds avec les sources de chaque backlink (ID, Label) ;
- Un fichier de liens (Source, Target, Label).

On va également garder notre premier fichier de nœuds, celui de toutes nos URL de notre site. Commencez un nouveau projet, puis :

- Importez le fichier de nœuds initial, c'est-à-dire tous vos URL internes ;
- Importez le fichier de nœuds de vos backlinks ;
- Importez le fichier de liens de vos backlinks.

Une fois l'importation réalisée, on peut alors réutiliser nos outils de spatialisation comme au départ. La problématique que vous allez rencontrer est que l'on ne peut différencier nœud interne (URL de mon site) et nœud externe (source des liens). Vous pouvez donc soit procéder avec un filtre comme dans l'exemple précédent, soit utiliser une autre méthode de coloration plus simple. Pour cette deuxième solution :

- Rendez-vous dans le tableau de données ;
- Classez le tableau par ordre alphabétique d'URL ;
- Sélectionnez toutes vos URL internes ;
- Clic-droit puis « Sélectionner dans la vue d'ensemble » ;
- Puis coloration manuelle unique, ici en rouge dans notre exemple.

Dans la figure 11, le nombre de domaines référents de chacune de mes URL en rouge est bien plus parlant.

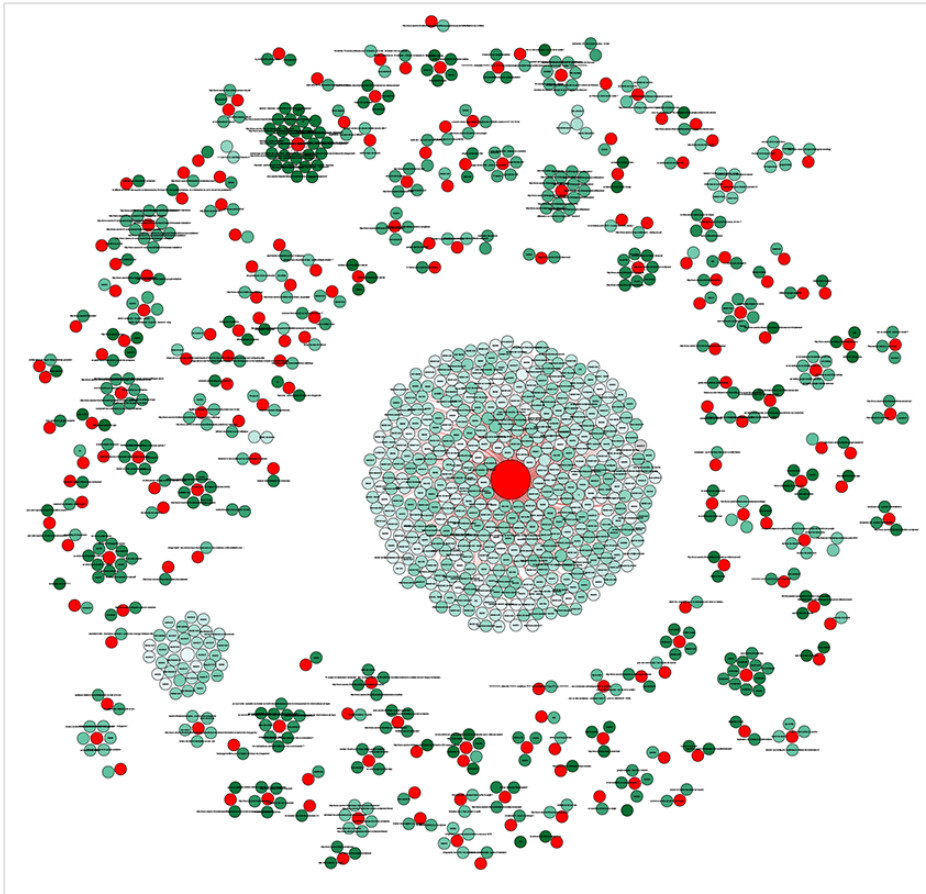


Fig. 11. Les nœuds en rouge sont les pages de mon site. Tous les autres sont les domaines référents associés.

## Conclusion

Les exemples donnés ici ne sont justement que des exemples. On pourrait faire bien plus. Il est possible d'utiliser Gephi dans une analyse Webmarketing au sens large, par exemple avec l'analyse des partages sociaux ou des conversions réalisées.

Bien entendu, il n'est pas conseillé de systématiquement recourir à ce type de logiciel, car l'analyse peut prendre du temps. Et surtout, on peut parfois avoir des données tout aussi lisibles dans un fichier Excel.

Malgré cela, Gephi est un logiciel très utile et puissant, notamment lors d'audits complexes de référencement naturel. Seul réel défaut, et non des moindres : Gephi reste un logiciel complexe et peu stable, ce qui peut parfois nuire à votre travail d'analyse.



**Daniel Roch**, Consultant WordPress, Référencement et Webmarketing chez SeoMix (<http://www.seomix.fr/>).