

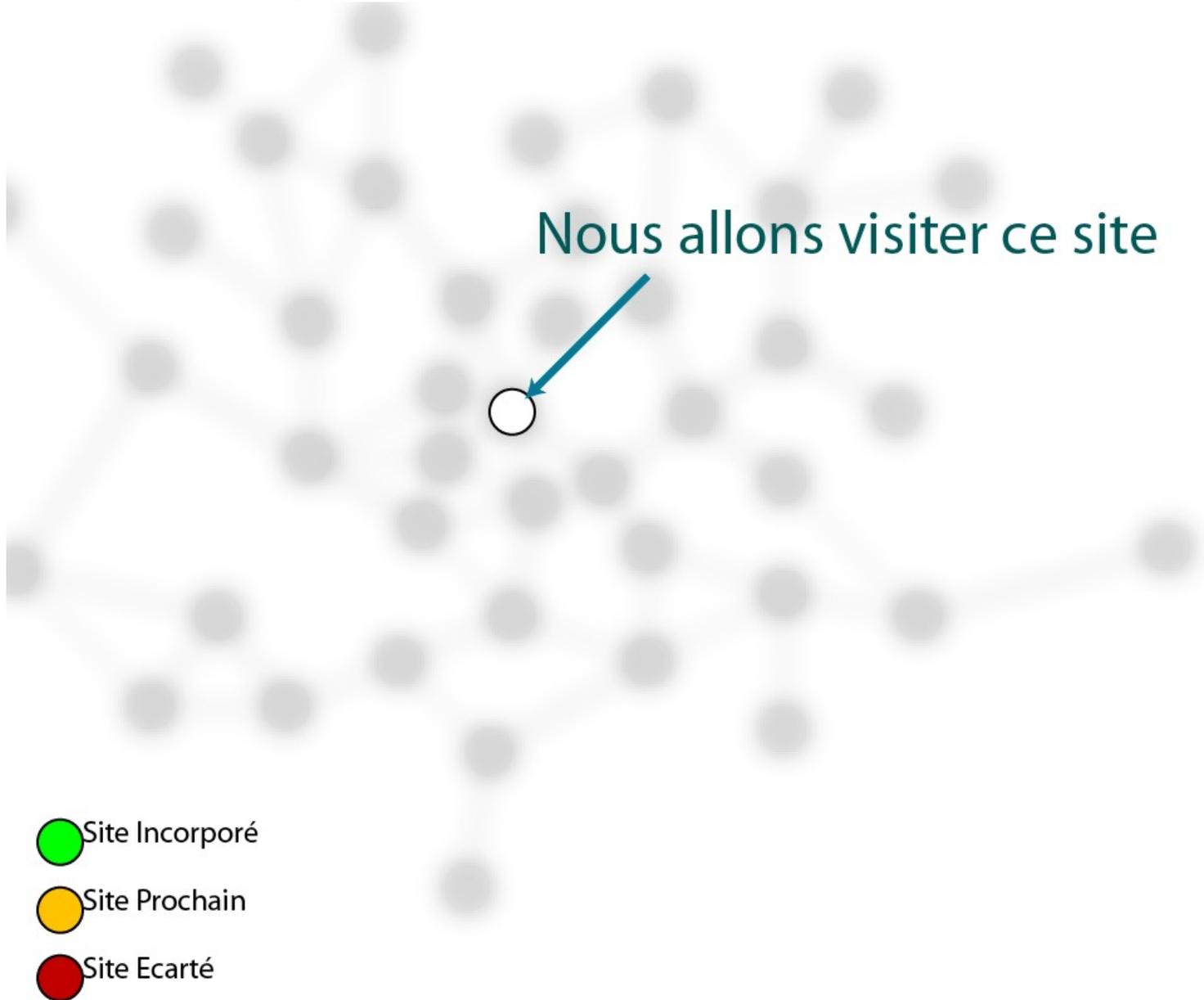


Eléments de méthode Exploration Navicrawler

Franck Ghitalla
Département TSH
Président de WebAtlas
franck.ghitalla@gmail.com

Mathieu Jacomy
Trésorier de WebAtlas
Doctorant CELSA
mathieu.jacomy@gmail.com

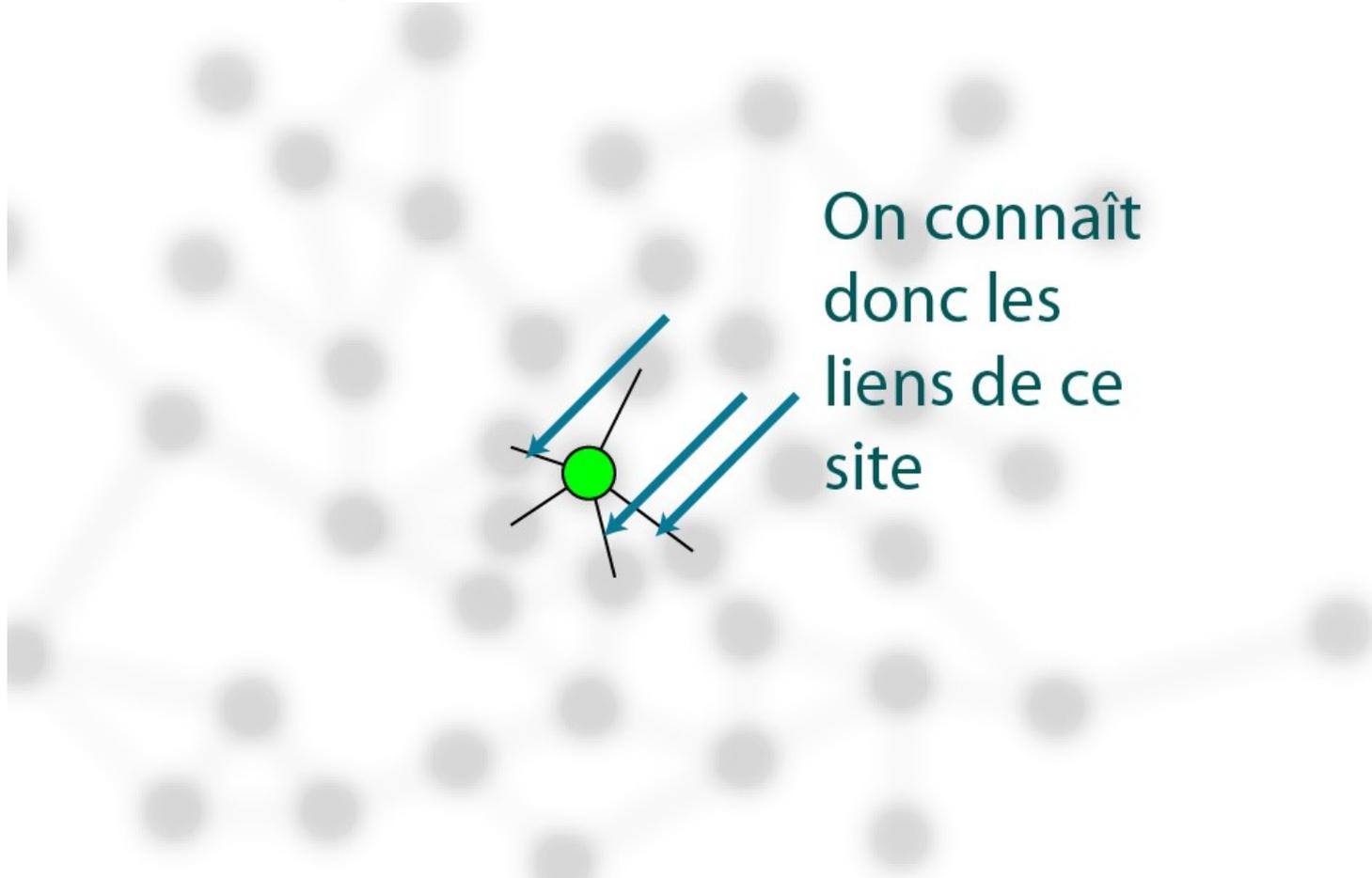
Corpus & Navicrawler



Corpus & Navicrawler



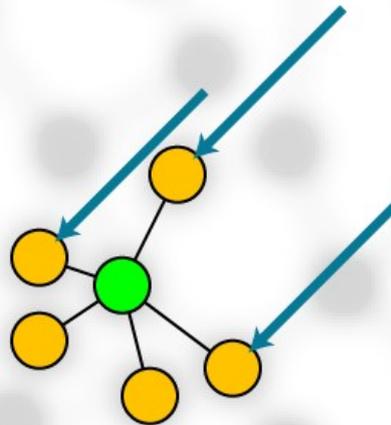
Corpus & Navicrawler



On connaît
donc les
liens de ce
site

-  Site Incorporé
-  Site Prochain
-  Site Ecarté

Corpus & Navicrawler

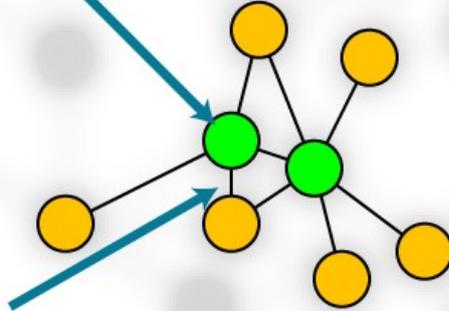


Les sites au bout
des liens sont
détectés :
ce sont
les sites
prochains.

-  Site Incorporé
-  Site Prochain
-  Site Ecarté

Corpus & Navicrawler

On continue : on visite ce site.



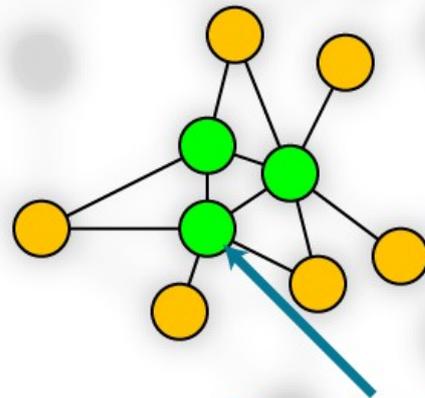
Ce lien est découvert
(on n'avait pas les liens entre
sites prochains)

 Site Incorporé

 Site Prochain

 Site Ecarté

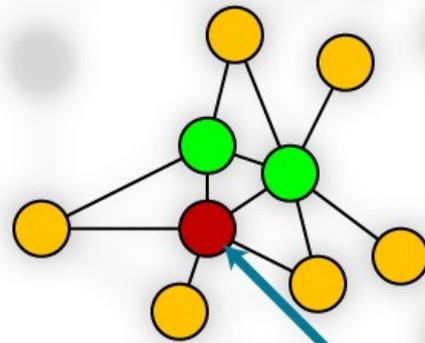
Corpus & Navicrawler



On viste encore ce site
Mais il ne nous
convient pas !

-  Site Incorporé
-  Site Prochain
-  Site Ecarté

Corpus & Navicrawler



Ce site est donc écarté.

-  Site Incorporé
-  Site Prochain
-  Site Ecarté

Corpus & Navicrawler

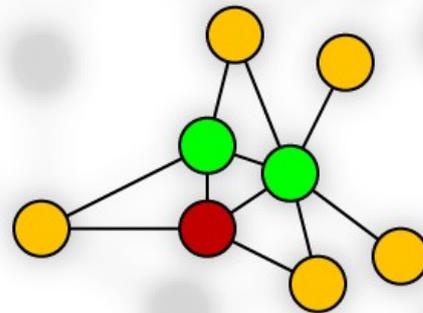


Ce site prochain va être effacé (car on efface les voisins des sites écartés)

Ce site prochain est gardé quand même car il est lié à un incorporé

-  Site Incorporé
-  Site Prochain
-  Site Ecarté

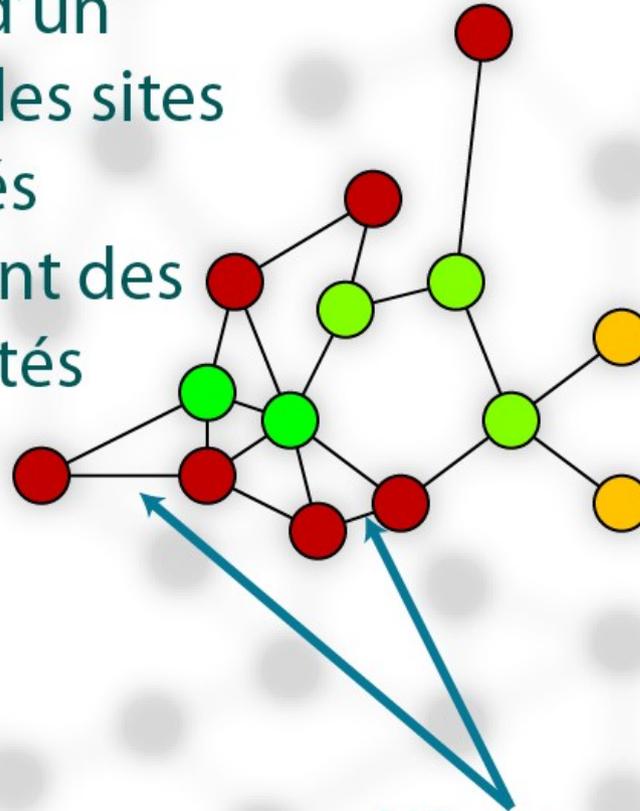
Corpus & Navicrawler



-  Site Incorporé
-  Site Prochain
-  Site Ecarté

Corpus & Navicrawler

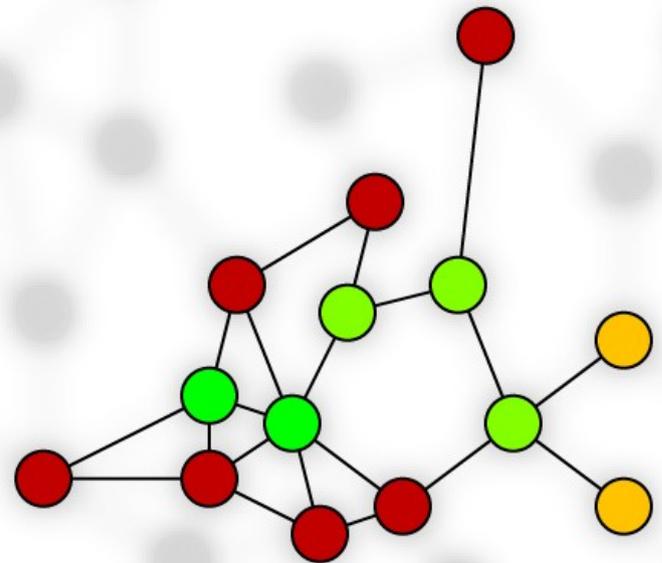
Au bout d'un moment les sites incorporés s'entourent des sites écartés



NB : on connaît les liens entre sites écartés

-  Site Incorporé
-  Site Prochain
-  Site Ecarté

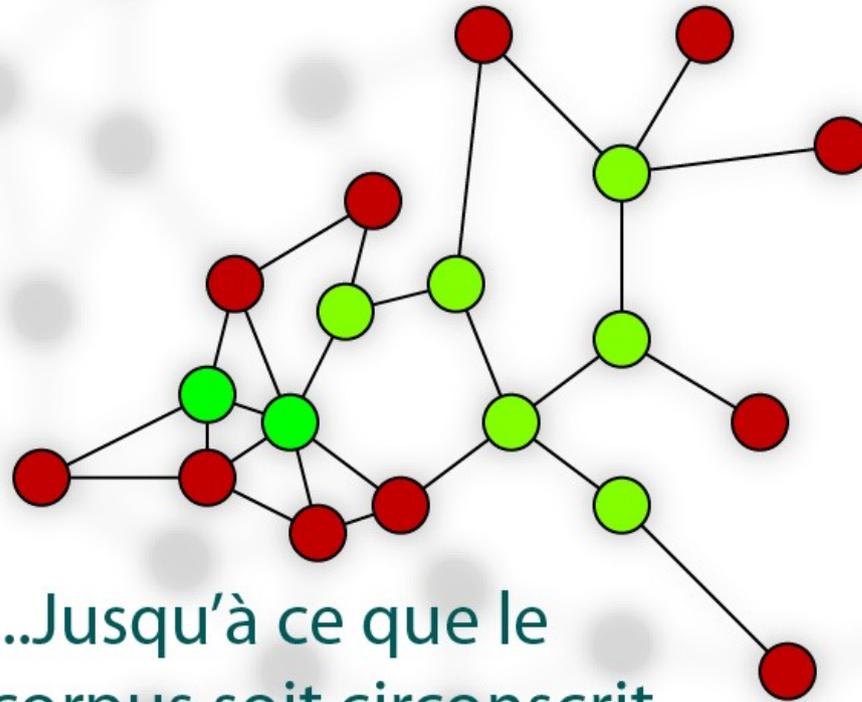
Corpus & Navicrawler



On continue
à traiter
les sites
prochains...

-  Site Incorporé
-  Site Prochain
-  Site Ecarté

Corpus & Navicrawler

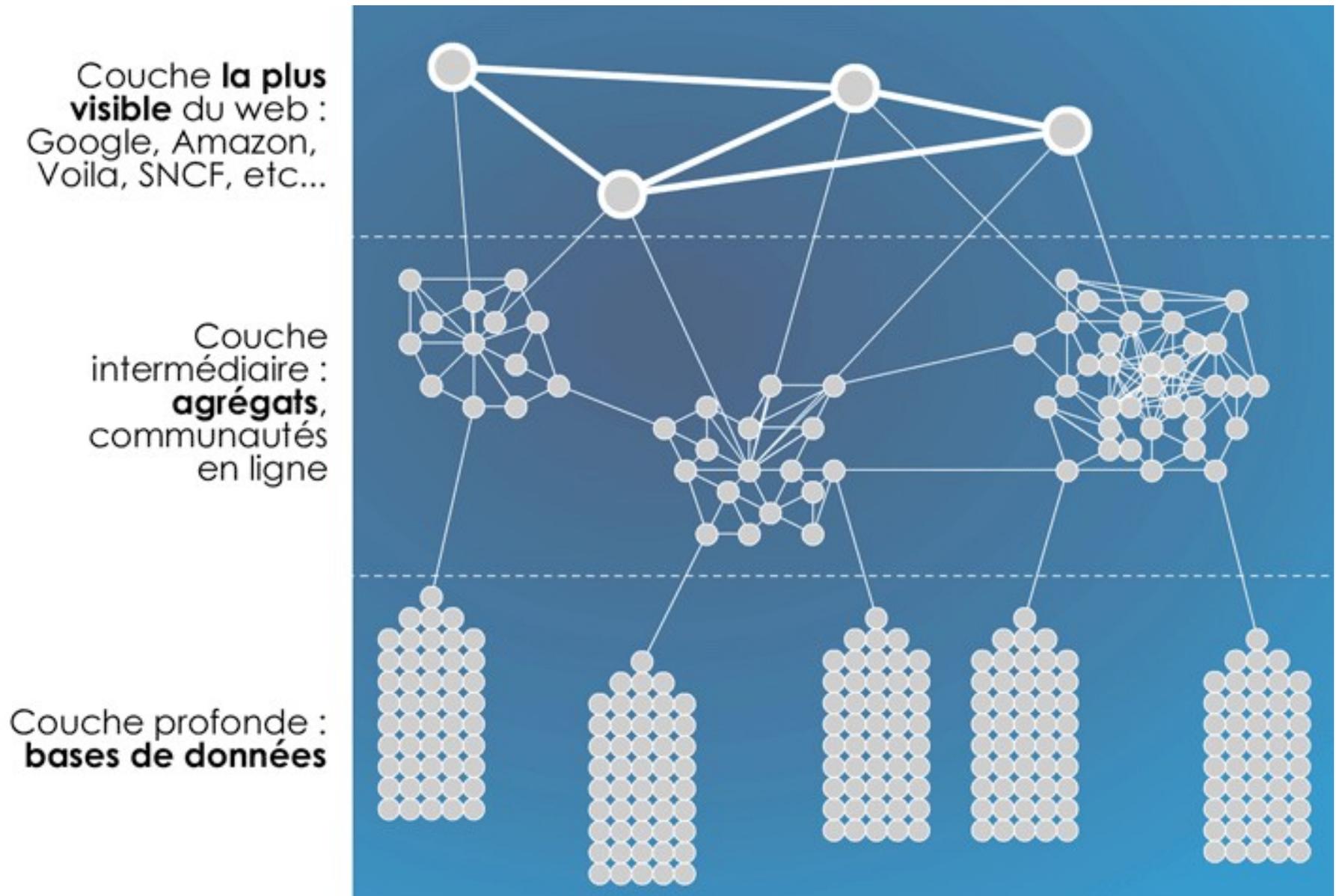


...Jusqu'à ce que le
corpus soit circonscrit.

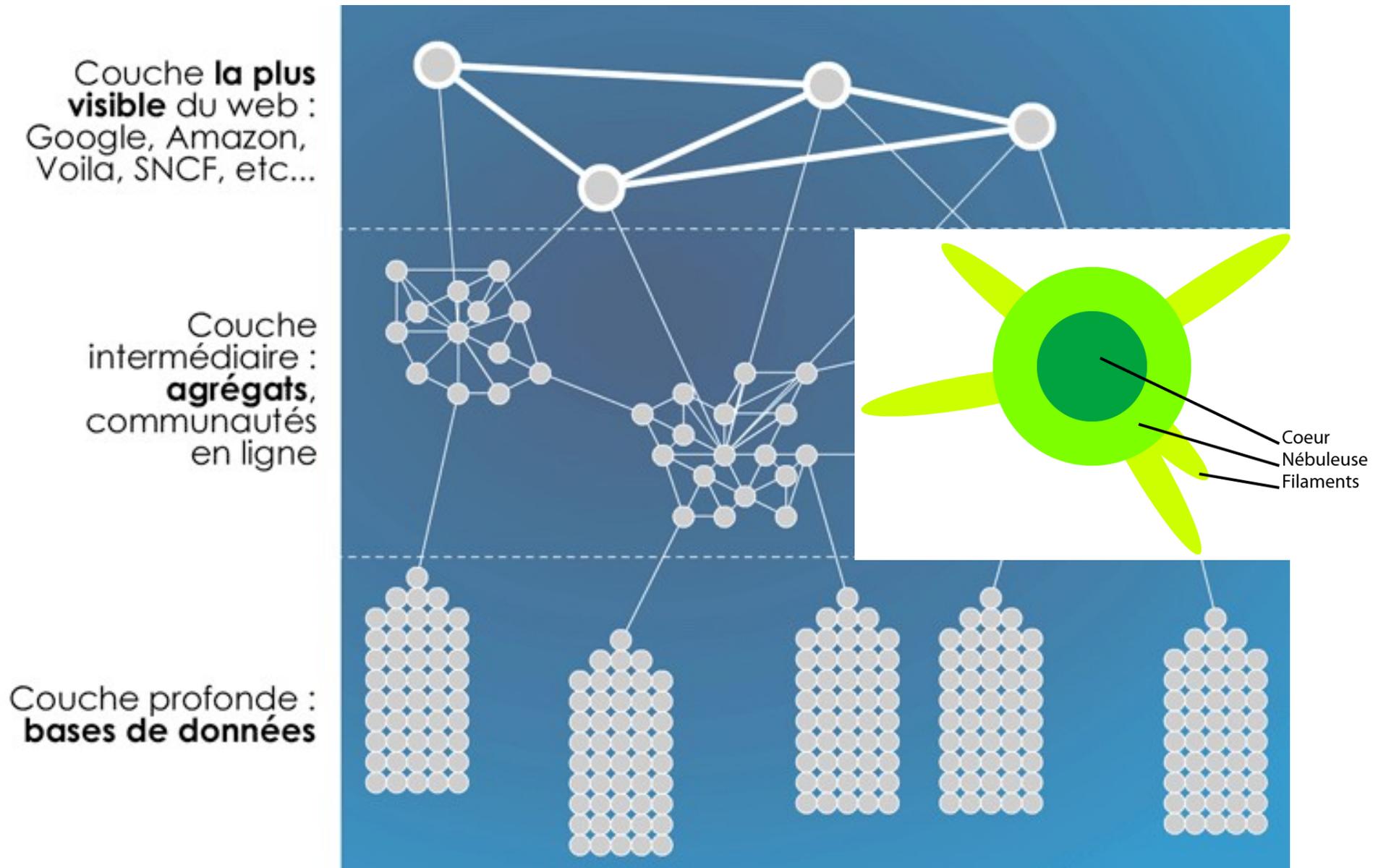
On a conservé les sites écartés
en tant que contexte.

-  Site Incorporé
-  Site Prochain
-  Site Ecarté

Les couches du web

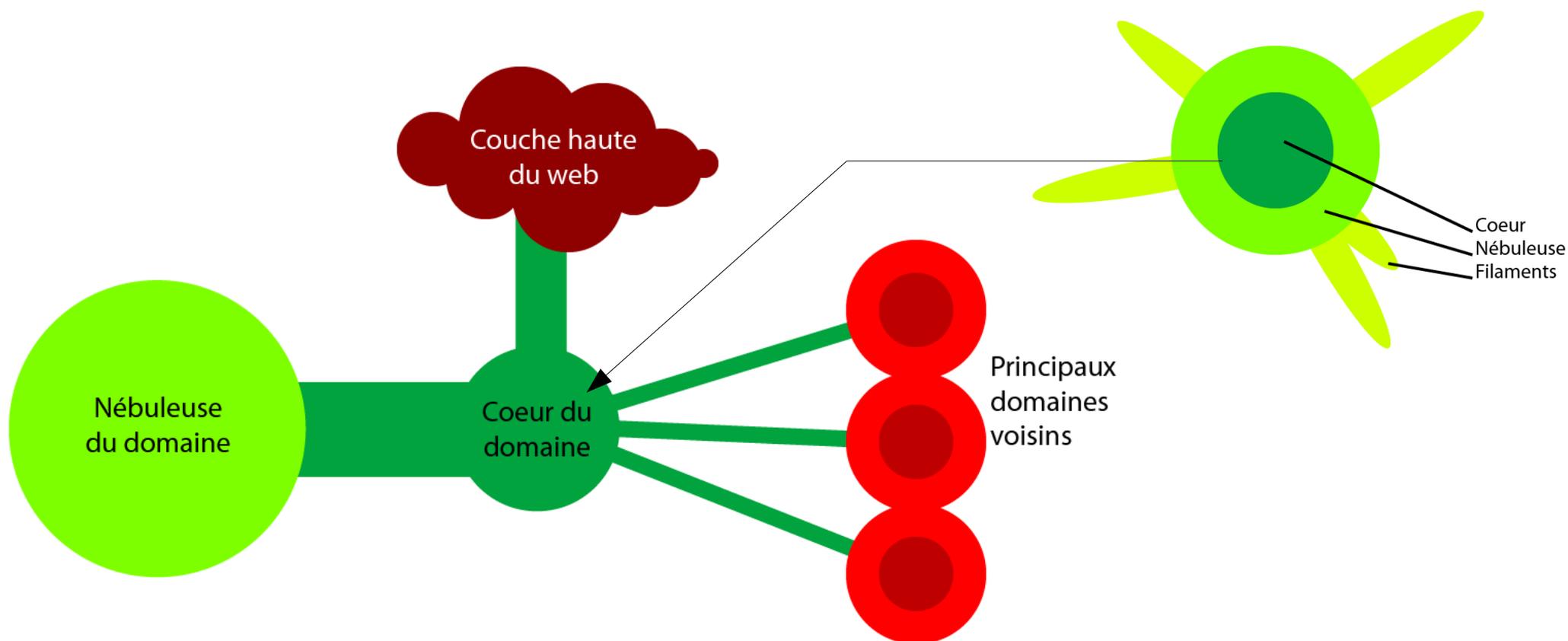


Morphologie d'un domaine



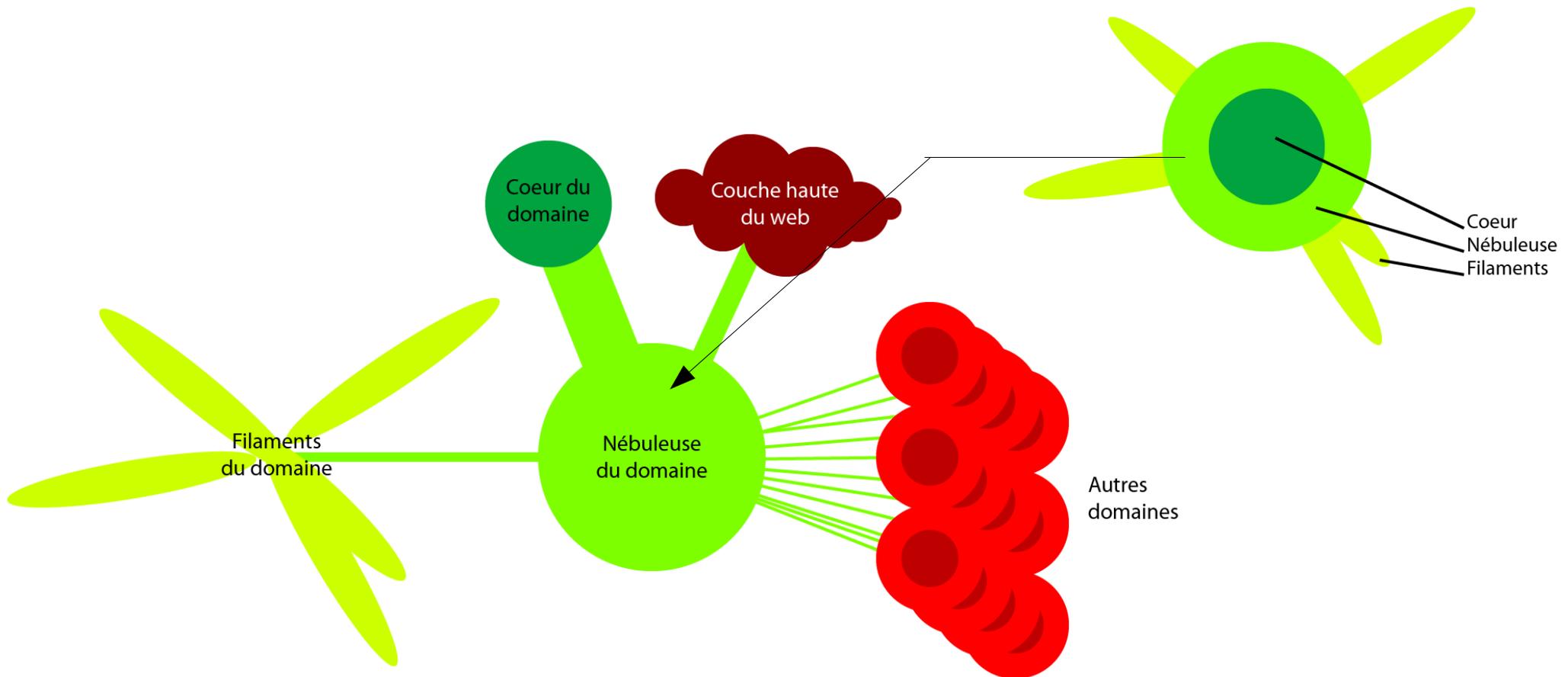
Connectivité du coeur

Avec quoi le coeur est-il connecté ?



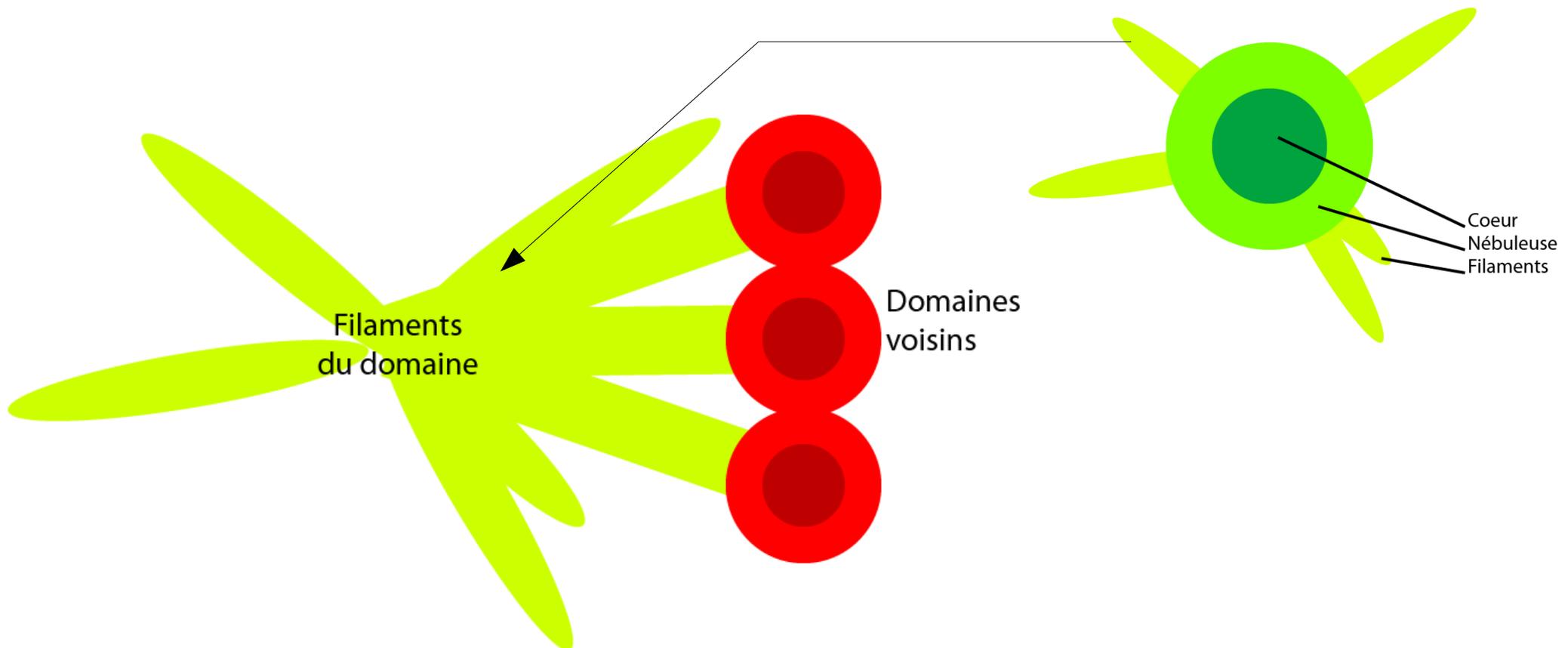
Connectivité de la nébuleuse

Avec quoi la nébuleuse est-elle connectée ?



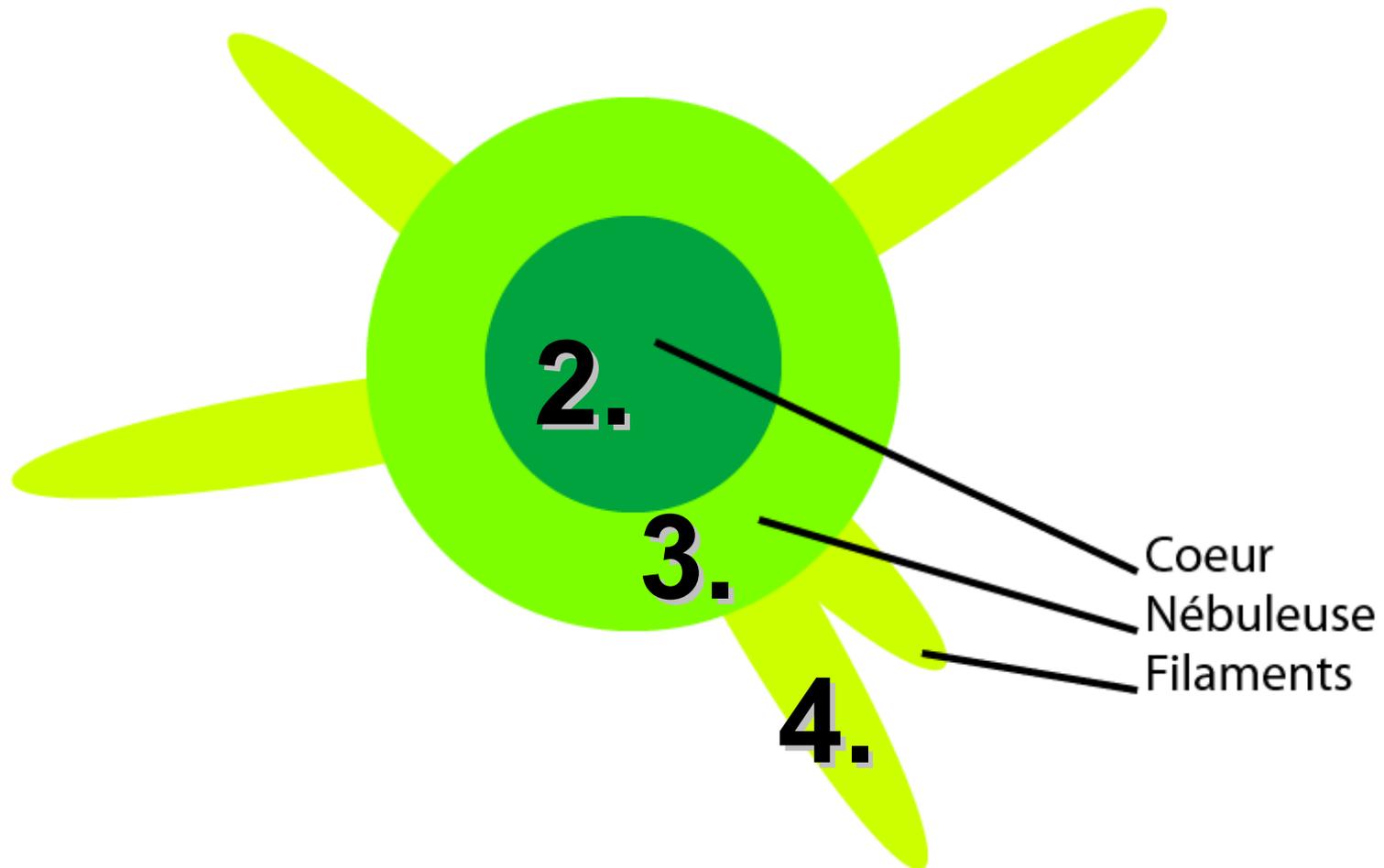
Connectivité des filaments

Avec quoi les filaments
sont-ils connectés ?

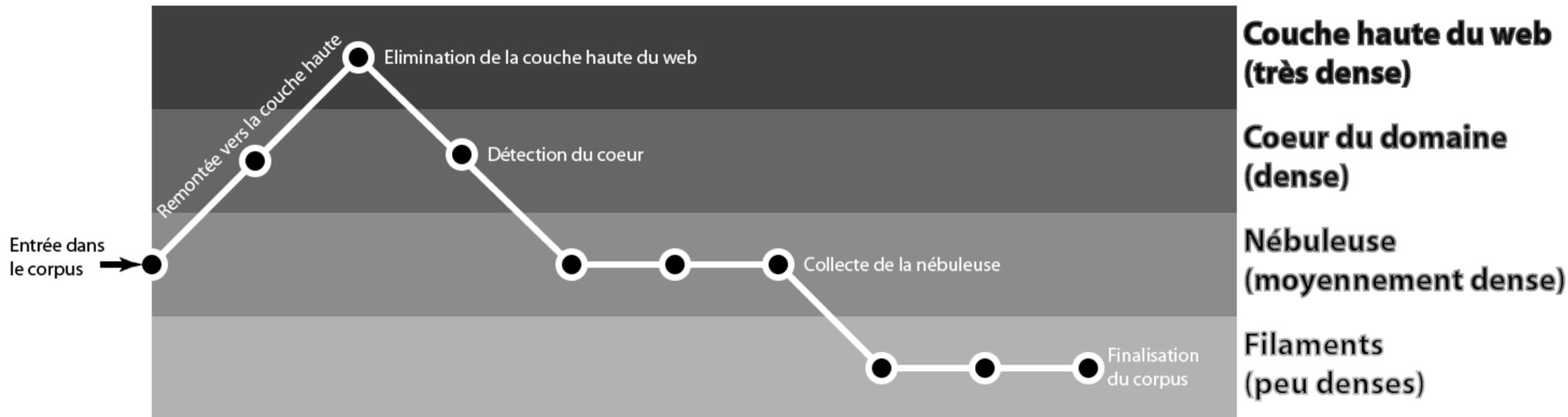


Pour ne pas se perdre...

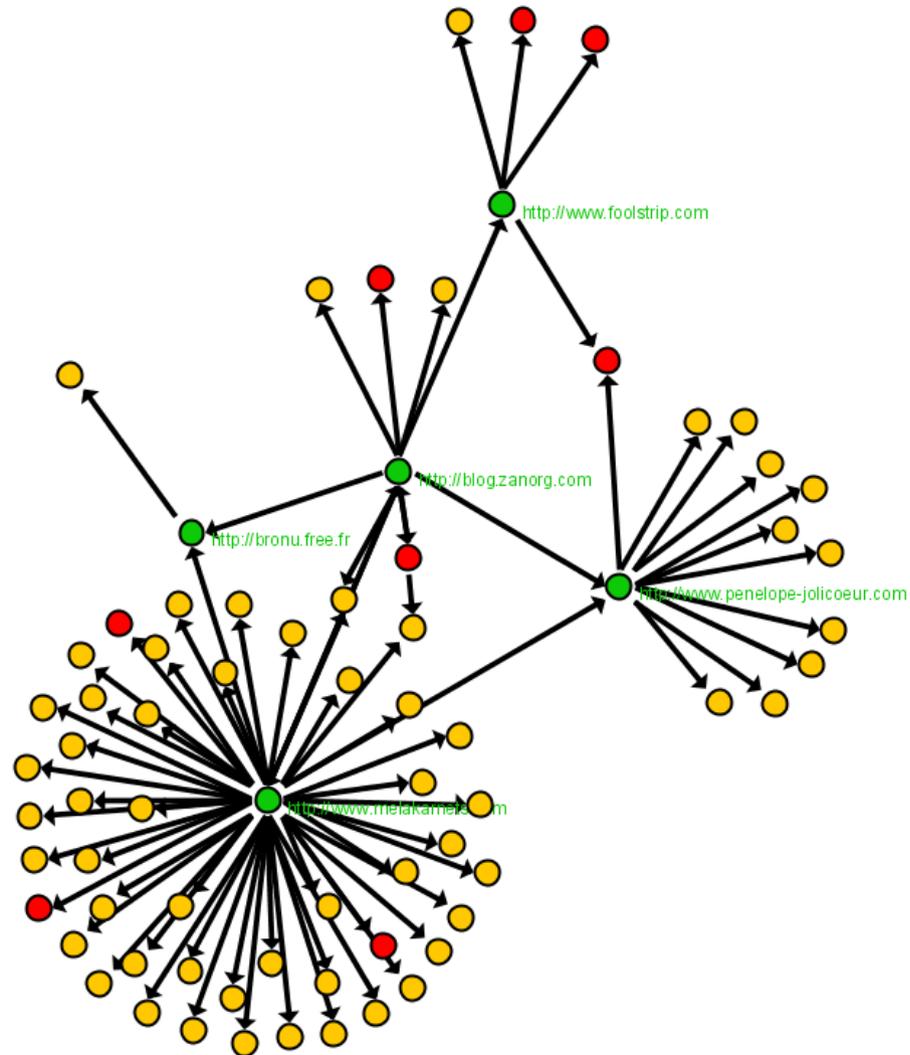
1. Eliminer la couche haute



Méthode du haut vers le bas



Exemple de graphe en cours



Courbe d'avancement de l'exploration

