



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



Escola Tècnica  
Superior d'Enginyeria  
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica  
Universitat Politècnica de València

# Estudio de Link Network Analysis y su influencia en el posicionamiento de búscadores

Trabajo Fin de Grado

**Grado en Ingeniería Informática**

**Autor:** Cristian Alexandre Capraru Pons

**Tutor:** Jose Antonio Ontalba Ruipérez

Cotutor: Jorge Serrano Cobos

2014-2015





# Resumen

---

Este Trabajo de Fin de Grado de la titulación de Grado en Ingeniería Informática consiste en el análisis de las redes de links, con la finalidad de que un enlace en un nicho de mercado, pueda mejorar su visibilidad en los buscadores. El proyecto también estudia una serie de herramientas analíticas para luego, exportar los datos en la herramienta de código libre Gephi. Gracias al uso de grafos, se podrán observar, de manera visual, algunas conclusiones sobre dicho estudio, concretamente, en el universo vinícola.

**Palabras clave:** linkbuilding, SEO, análisis de redes sociales, posicionamiento en buscadores

# Abstract

---

This is the Final Project of the Engineering degree in Computer Science and it is about link network analysis, looking for a way in which a link in a prospective market could improve search engine visibility. The Project also studies some analytical tools from which the data was exported into an open source tool called Gephi. Through the use of graphs, some conclusions about the research, specifically in the world of wine, can be visualised.

**Keywords :** linkbuilding, SEO, social network analysis, Search engine optimization

# Tabla de contenidos

---

## Tabla de contenido

<b>1. Introducción</b>	<b>7</b>
<b>1.1 Marco de estudio y justificación</b>	<b>8</b>
a. Objetivo del análisis	9
b. Estructura del Trabajo de Fin de Grado	11
<b>2. Estado de la cuestión</b>	<b>12</b>
<b>2.1 Motores de búsqueda</b>	<b>12</b>
a. Definición de buscador	12
b. Clases de buscadores	15
c. Algoritmos usados en los motores de búsqueda	17
d. PageRank™	20
<b>2.2 Posicionamiento seo</b>	<b>23</b>
a. Importancia de un buen posicionamiento	23
b. Como mejorar el posicionamiento	25
c. Link Baiting y Link Building	28
<b>3. Material y Métodos</b>	<b>31</b>
<b>3.1 Material</b>	<b>31</b>
a. Ahrefs	31
b. Semrush	37
c. Majestic SEO	44
<b>3.2 Análisis de influencias</b>	<b>50</b>
a. Centralidad	50
b. Uso de la interfaz con la herramienta Gephi	51
<b>3.3 Patrones de enlazado entre una muestra de webs del universo     vinícola</b>	<b>55</b>
<b>4. Análisis de resultados</b>	<b>62</b>
<b>5. Recomendaciones a nivel empresarial</b>	<b>68</b>
<b>6. Conclusiones</b>	<b>70</b>
<b>7. Bibliografía</b>	<b>73</b>
<b>8. Glosario</b>	<b>75</b>



# 1. Introducción

Hoy en día, la sociedad en la que nos encontramos vive con un acceso a una gran parte de la información mundial. Gracias a la creación de la infraestructura “Internet”, portales de todo el mundo permiten a la humanidad estar informada en todo momento ya sea desde un ordenador o desde un móvil con conexión. Actualmente, ya desde niños, la gente pasa muchas horas al día conectados. Estas herramientas se convierten en indispensables y habituales en el día a día de la sociedad, sea cual sea la clase social. La información está a un solo *click*. No olvidemos que estamos ante un entorno del conocimiento donde se juntan intereses variados, tanto privados como públicos.

Internet ejerce una influencia muy grande en la gente, de manera que uno siempre puede estar actualizado, cuando sea y donde sea. Desde las redes sociales a las noticias, pasando por meros blogs y informativos independientes. Compartir contenido esta a la orden del día, y es una nueva forma de proveer información en el mundo digital. El uso de blogs por ejemplo cada vez tiene un uso más extendido, ya que son plataformas gratuitas y que pueden usarse para comercializar productos o servicios, compartir experiencias e interactuar con los usuarios visitantes.

El objetivo de este Trabajo de Fin de Grado es analizar esas influencias que ejercen unos dominios con otros y entender como funcionan los algoritmos de los motores de búsqueda. Para ello se realizará un exhaustivo análisis de los enlaces de la red relacionados con el mundo del vino, concretamente, un subconjunto de dominios de bodegas españolas con un posicionamiento de marca alto y se observará cuál es su influencia en el posicionamiento en los buscadores.

## 1.1 Marco de estudio y justificación

El origen del proyecto de “Estudio de Link Network Analysis y su influencia en el Posicionamiento en Buscadores” surge a partir de un interés en la búsqueda del funcionamiento de los buscadores. Toda empresa hoy en día quiere que su producto esté posicionado para ciertas búsquedas. En este sentido, queda claro que existe una competencia máxima para ciertas **keywords**<sup>1</sup> y lo que queremos averiguar pues, es como se debe trabajar para optimizar una keyword.

La principal motivación a la hora de desarrollar este proyecto, ha sido la oportunidad que me han ofrecido mi tutor y mi cotutor, a los que les estoy muy agradecidos, de comprender el funcionamiento de los buscadores online. Creo que la experiencia adquirida a la hora de realizar este proyecto, implica un aprendizaje en todas las áreas a nivel de web.

Siempre he considerado que tiene mucha importancia el hecho de escoger un trabajo que me motive y que me permita desarrollar mis aptitudes de forma integral como individuo. En este sentido, en este proyecto tengo la posibilidad de:

- Aprender a trabajar y coordinar un proyecto web que una empresa en un futuro me confíe.
- Mejorar mis destrezas técnicas referentes al mundo del posicionamiento en buscadores.
- Poner en práctica mis conocimientos a la hora de participar en un proyecto web durante todas sus fases de desarrollo.

---

<sup>1</sup> Keyword o palabra clave: un término (o términos) relevantes para nuestro proyecto. El objetivo de cualquier estrategia de posicionamiento SEO consiste en posicionar las keywords objetivo en las primeras páginas de los resultados de búsqueda.

## a. Objetivo del análisis

***“En el mundo de la web, donde todo el mundo quiere dejar backlinks en todas partes, este proyecto pretende ser una herramienta para mostrar “Dónde” y “Cómo” deben ser dejados”***

Los objetivos del posicionamiento vienen dados por la finalidad de nuestra web. ¿Qué es lo que se pretende? ¿Potenciar una idea? ¿Vender un servicio o producto? Está claro que lo que uno pretende es llegar ser el primero en una temática o ser el primero para unas keywords determinadas y para ello queremos dirigir al usuario desde el momento que descubre nuestra web hasta que realiza la conversión esperada y abandone nuestra web.

Entre los muchos factores que influyen en el posicionamiento de una web, tenemos:

- Edad del dominio
- Las keywords en el nombre de dominio
- La extensión del dominio para el país
- Outbound Link Quality<sup>2</sup> o calidad de los enlaces salientes<sup>3</sup>
- Inbound Link Quality o calidad de los enlaces entrantes.

...

Se pretende, pues, centrar el análisis en este último punto, dejando a parte el análisis ONpage (mejora de la web de manera interna). Como el algoritmo de Google es cambiante y no es público, se tratará de hacer una inferencia y vamos a intentar pensar como nos podemos acoplar al ritmo de baile de Google.

Así que, lo que se pretende con este trabajo, lo cual es nuestro objetivo, es saber con qué páginas vamos a tener que contactar para que nos enlacen de la forma más correcta y menos penalizable posible por los algoritmos de Google; pero para esto debemos entender como funcionan los buscadores, qué herramientas usar, y cómo dar los pasos de manera segura.

El propósito de este Trabajo de Fin de Grado también es descubrir como gestionar un sitio web para una empresa, en este caso, relacionada con el área de los vinos, ya sean bodegas, tiendas o blogs y en referencia a su posicionamiento en los buscadores de Internet. Se ha elegido este trabajo porque me considero una persona muy conectada al mundo de las nuevas

---

<sup>2</sup> OUTBOUND LINKS AND SEO [https://www.hochmanconsultants.com/articles/link-miser.shtml]. Hochman J. [Consulta: 3-7-2015]

<sup>3</sup> Para Google, no solo importan los enlaces entrantes, sino también los salientes. Si al usuario se le envía a una web basura, puede que la web origen también sea considerada como tal.

tecnologías. Además, he estado trabajado a nivel personal en un blog de temática tecnológica y esta experiencia me permitirá aportar una cierta habilidad y entendimiento en lo que a posicionamiento SEO se refiere. A su vez, quiero ampliar más conceptos sobre la influencia en Internet, que es un tema que considero apasionante.

## b. Estructura del Trabajo de Fin de Grado

Estos han sido los pasos que se han ido siguiendo para llevar a cabo el proyecto:

1. Recopilación de información sobre el “Link Network Analysis” a partir del “Social Network Analysis”
2. Estudio del funcionamiento de las herramientas a usar en el Trabajo: Ahrefs, Semrush, Majestic.
3. Elección de la herramienta más potente entre estas tres.
4. Estudio del funcionamiento de la herramienta Gephi para la interpretación de grafos de manera visual.
5. Obtención datos ampliados del caso de estudio a investigar por el tutor Jose Antonio Ontalba y el cotutor Jorge Serrano. El caso de estudio trata sobre el comportamiento de la palabra “vino” en los buscadores a partir de una muestra de 2.609 enlaces.
6. Introducción de la información exportada a Gephi para una interpretación visual.
7. Análisis de los resultados para obtención de conclusiones

Una vez analizados los resultados, se pretende que los datos obtenidos en el análisis del mundo del “vino”, provenientes de webs de bodegas, blogs y tiendas se puedan extrapolar a otros ámbitos, teniendo así un patrón de trabajo a la hora de examinar nuevos sectores y nichos de mercado.

## 2. Estado de la cuestión

### 2.1 Motores de búsqueda

#### a. Definición de buscador

Tener cierta organización sobre la información que se dispone siempre ha sido un hecho. Internet es un cúmulo de información cambiante y dinámico donde se puede acceder a los datos que uno desea en un momento dado. Gracias a los buscadores, podemos tener, de forma fácil y rápida, la información que se encuentra en la red, convirtiéndose en accesible mundialmente. El número de sitios web ha crecido de manera exponencial y los buscadores se convierten en las puertas de entrada a la información. Desde el inicio del internet, muchos han sido los buscadores que han pasado por nuestros ordenadores:

- Yahoo - 1994
- Web Crawler - 1994
- Lycos - 1994
- Excite - 1995
- Altavista - 1995
- Inktomi - 1996
- Ask Jeeves - 1997
- Google - 1998
- Open Directory Project / DMOZ – 1998
- MSN Search - 1998

Los resultados de algunos de estos buscadores eran introducidos manualmente. Se trataba de directorios de webs donde el factor humano tenía mucha importancia. Moderadores voluntarios de todo el mundo se dedicaban a rellenar cientos de categorías dentro de estos directorios. Pero con el crecimiento de la web, tal trabajo se hizo imposible. Pero en 1998, apareció Google, un buscador creado por dos estudiantes de la US (Universidad de Stanford): Sergey Brin y Larry Page, que consiguieron automatizar esta tarea logrando resultados a una velocidad tremenda.

El motor de búsqueda de Google, a través de un robot (GoogleBot, el robot de indexación de Google), iba guardando en un índice, millones de páginas web, asignándole a estas un valor más o menos relevante. Así, a la hora de que un usuario buscara información sobre ciertas keywords, el motor de búsqueda arrojaba un listado de resultados que habían sido filtrados por el algoritmo y que tenían que ver con el valor asignado anteriormente.

Google no fue el primer motor de búsqueda, pero sí que fue el más eficiente. Actualmente, Googlebot, ha rastreado más de 12.000 millones de documentos y los ha incluido en su base de datos. En estos momentos, la compañía afirma que su índice es tres veces mayor que el de cualquier buscador. Para ello, Google tiene a su disposición la mayor granja de servidores a nivel mundial (más de 1 millón), distribuidos en diferentes centros de datos.

Tal infraestructura necesita ser montada de forma eficiente, con bajos costes de hardware y con un rendimiento perfecto. Estos servidores están basados en el sistema operativo Linux. El software se encarga de indexar contenido junto a algoritmos de popularidad y relevancia para poder clasificar los resultados de búsqueda y que estos sean efectivos para tales búsquedas. Los algoritmos y técnicas, son actualizados cada poco tiempo, ya que a veces fallan y pueden salir a la luz resultados de baja calidad cara al usuario final.

Resumiendo, los 4 cimientos de los motores de búsqueda son:

- El Robot, rastreador, araña o *crawler*: Se trata de un emisario enviado por el motor, esta definido por cada buscador y su misión es enviar peticiones a los servidores web. Con ello, rastrea la WWW y compila información sobre los sitios de la web que visita. La extensión se produce gracias a los enlaces de hipertexto. Estos robots no suelen interpretar imágenes ni animaciones flash. Una función importante del robot es actualizar la información ya existente para modificar la base de datos o descartar portales borrados o retirados.
- El índice: se forma a partir de grandes bases de datos, donde se almacena todo aquello que ha recopilado el robot. Este índice esta formado por millones de documentos y archivos. Cuanto más grande sea el índice, más extenso y variado será el buscador.
- La interfaz de búsqueda: el buscador en sí. Se debe poder acceder de alguna manera, la más usada es a través de una **URL**. La interfaz debe tener alguna especie de campo de búsqueda donde el usuario haga la consulta y se le devuelva una respuesta.
- Un algoritmo que organice la información: se debe disponer de una herramienta de ordenación según la autoridad, ordenando de mayor a menor relevancia dependiendo de la consulta del usuario. Este suele ser secreto y es donde se almacena el *alma mater* de la credibilidad del buscador.



El uso de buscadores aumenta cada día. A través de estadísticas<sup>4</sup> se puede comprobar que los usuarios que se sirven de los buscadores, no suelen sobrepasar la segunda página de resultados. De manera que se puede afirmar que, con los millones de búsquedas que se hacen diariamente, estar entre los primeros veinte resultados se puede considerar todo un privilegio y incluso a veces, un líder de opinión.

Entrando en temas de competencias empresariales, si nuestra web se exhibe entre las 20 primeras posiciones, estaremos dentro de una zona con infinidad de potenciales clientes y por tanto de rentabilidad económica. En caso contrario, será nuestra competencia la que podrá hacerlo, disipándose nuestras posibilidades de conseguir clientes rentables a través de la aparición en las SERPs <sup>5</sup> de los diferentes motores de búsqueda.

---

<sup>4</sup> Relación entre la posición en Google y la cantidad de visitas [<http://www.alquiler-sitio-web.info/estudia-ctr-segun-la-posicion-en-google/>] Pierrick J. [Consulta: 1-6-2015]

<sup>5</sup> SERPs: hace referencia a los resultados que muestra una página de buscador como pueden ser: Google, Bing, y Yahoo! entre los más destacados.

## b. Clases de buscadores

Si se desea entrar en un análisis de clasificación de la información y de los datos que están en Internet, se pueden diferenciar diversos tipos de buscadores. Aunque cada buscador tiene su manera de funcionar y sus algoritmos personalizados, tres son los grandes grupos de buscadores que debemos destacar:

1. Los motores de búsqueda o arañas: son los buscadores más conocidos de todos. Se basan en palabras clave, que son las palabras usadas en la búsqueda de información. Los motores de búsqueda guardan páginas web que clasifican en función de sus algoritmos. El encargado de esto es un software llamado “spider” que guarda en su base de datos lo que recopila en la red. A continuación, cuando el usuario introduce unas palabras clave en el recuadro de búsqueda, por pantalla se mostrará, de manera ordenada (según el algoritmo), un listado de los resultados que más se acerquen a esa palabra clave o expresión.

En caso de que una nueva url se genere, deberá esperar a que la araña del buscador (pequeños programas informáticos con la tarea de recoger información de la WWW) pase por su web para ser indexada. Los buscadores poseen miles de millones de urls en sus bases de datos<sup>6</sup>.

Ejemplos de ello son:

Google: El gran jefe de los buscadores. Con un algoritmo muy preciso y trabajado por unos ingenieros que lo diseñaron para que diese más valor a los enlaces que se ceñían mejor a las keywords introducidas por el usuario y a la popularidad que adquirirían dichos enlaces.

Yahoo!: Un buen buscador, que a su vez es también un directorio de enlaces. Hoy en día tiene su propio algoritmo, aunque la capacidad de generar tan buenos resultados como los de Google no es la misma.

Bing: el más novato de los tres y el tercero más usado a nivel mundial. Los resultados de este motor de búsqueda no son en absoluto comparables con la calidad de resultados que ofrece Google, sin embargo, detrás de él se sitúa una empresa grande como es Microsoft.

---

<sup>6</sup> El uso de los buscadores en Internet: Buscadores en Internet. Conceptos, clasificación y características [http://bvs.sld.cu/revistas/aci/vol11\_3\_03/aci04303.htm] Torres-Pombert A. [Consulta: 1-6-2015]



2. Directorios: Los directorios organizan y reúnen portales en categorías y subcategorías, como en una biblioteca. Los portales son clasificados según su temática y en función de su contenido, pueden aparecer en una o diversas subcategorías.

Cuando una web necesita que se le indexe en un directorio, envía una solicitud de indexación, que es estudiada por el equipo de redacción. Los resultados en estos directorios, son bastante menores que los de los motores de búsqueda. Pero suelen ser de mayor calidad ya que pasan por las manos de una persona real y no por una máquina.

Hay dos maneras de realizar una búsqueda en un directorio:

1. La primera es buscar en forma de categorías y subcategorías, es decir de forma jerarquizada.
2. La segunda es en un campo de búsqueda donde se introducen las palabras clave sin pasar por categorías y esperando un listado de resultados que se ajusten a esa búsqueda.

Ejemplos de esto son:

Origen	Buscadores	Directorios
<b>Español</b>	Altavista(de Yahoo!) Hispavista	
<b>Internacional</b>	Google Yahoo Bing	Yahoo Dmoz (Open Directory) Looksmart
<b>Búsqueda de la información</b>	Indexación de webs Mayor número de webs indexadas Se persigue la exhaustividad Clasificación por orden de relevancia Actualización de la info más rápida	Indexación de webs Menor número de webs indexadas Criterios de calidad Búsqueda por categorías Actualización más tardía
<b>Indexación</b>	Indexación automatizada Indexación de la mayoría de webs	Indexación humana Criterios de calidad bajo supervisión para la indexación

Fig. 1: Resumen de la diferencia entre Buscadores y Directorios. Fuente: Propia

En conclusión: Los buscadores y los directorios no trabajan pues de la misma manera.

3. Metabuscadores: Por último, se encuentran los metabuscadores, es decir, motores de búsqueda que rastrean la información en otros buscadores para devolver los resultados que encuentran en estos. Un ejemplo de ello es el metabuscador español Trivago, que devuelve al usuario los mejores precios que ofrecen distintas webs de viaje.

The screenshot shows the Trivago website interface for a search in Madrid. At the top, it displays '60 mil' and '2,4 millones' of users, and '782.554 hoteles de 269 webs'. The search bar contains 'Madrid' and the dates 'Entrada Domingo, 26.07.15' and 'Salida Lunes, 27.07.15'. The search results are displayed in a grid format, showing three hotel cards:

- NH Barajas**: 3 stars, Hotel, Madrid, Madrid (Provincia), España. Price: 48 €. Web oficial del hotel: 48 €. Hoteles.com: 48 €. otel.com: 500 €. Ver ofertas de 19 webs.
- NH Madrid Puerta de Alcalá**: 3 stars, Hotel, Madrid, Madrid (Provincia), España. Price: 72 €. Web oficial del hotel: 72 €. Venere.com: 72 €. Amoma.com: 125 €. Ver ofertas de 21 webs.
- NH Madrid Ribera del Manzana...**: 3 stars, Hotel, Madrid, Madrid (Provincia), España. Price: 72 €. Web oficial del hotel: 72 €. Hoteles.com: 72 €. 7ideas: 73 €. Ver ofertas de 20 webs.

On the right side, there is a 'Mi lista' section with a 'Comparar hoteles' button and a 'Últimas búsquedas' section with 'No hay historial'.

Fig. 2: Ejemplo de web tipo Metabuscador: Trivago.

Fuente: <http://www.trivago.es/>

### c. Algoritmos usados en los motores de búsqueda

El algoritmo de ordenación de búsqueda es el encargado principal de devolver aquel contenido que considera “relevante”. Mucho se ha hablado del algoritmo de ordenación de Google. Pero tan cambiante es este algoritmo, que hoy puede dejar una página “spam” en las primeras posiciones, y mañana encerrarla en el sand-box<sup>7</sup>. Las actualizaciones de las bases de datos y retoques dentro del algoritmo de Google pueden dar lugar a cambios en las SERP de Google. Por ejemplo, con la llegada del “zoológico” de Google, se les da nombres a cada uno de los grandes cambios que se hace en algoritmo de búsqueda. Los principales y novedosos algoritmos de los últimos tiempos han sido:

- Google Panda 2011: se trata de la primera gran actualización que intentó pararle los pies al SPAM provocado de manera inadecuada. Tal algoritmo se concentró en excluir de las primeras posiciones del buscador a toda web que ofrecía contenido de mala calidad (textos espineados<sup>8</sup>, keyword stuffing<sup>9</sup>) o contenido duplicado. Esto supuso ciertos cambios en el algoritmo de Google, ya que era de vital importancia que las primeras posiciones fuesen solo para contenidos de calidad. Es decir:
  - Contenido único y original de la web.
  - Alto CTR<sup>10</sup>.
  - Contenido Viralizable
  - Porcentaje de rebote bajo

---

<sup>7</sup> Sand-box: Es un efecto observado, que postula un sistema de filtrado especial de Google en su algoritmo, que hace que las páginas web de nueva creación no aparezcan, o aparezcan en posiciones muy inferiores, en los resultados de búsqueda

<sup>8</sup> Espinear: Tomar un texto y sustituirlo por sinónimos, otras formas verbales, etc. para que parezca original cara a los buscadores.

<sup>9</sup> Keyword Stuffing: Acción de inflar artificialmente la densidad de palabras clave de una web

<sup>10</sup> CTR: relación porcentual entre clicks e impresiones (número de veces que se muestra tu página en los buscadores). Es un valor que nos indicará la efectividad de nuestras campañas.

- **Google Penguin 2012-2013:** fue la segunda gran actualización de Google y pretendía focalizarse en el Link Building<sup>11</sup> automatizado por herramientas como ScrapeBox<sup>12</sup>, que permiten gracias al uso de proxies, dejar backlinks usando patrones y footprints<sup>13</sup> de búsqueda de cierta precisión. Google creó esta actualización para excluir de las SERPs aquellos portales que habían recibido SEO engañoso. Para Google, un sitio web de calidad es aquel donde la gente comparte su contenido de forma natural, porque encuentra lo que necesita y porque considera que puede ser útil para otras personas de sus círculos.

Las reglas de la ordenación en los motores de búsqueda residen en el algoritmo. De esta manera, se preserva la calidad del servicio eliminando contenido de naturaleza dañina. A esta agilidad de cambio y a su eterno afán de indexación se le conoce como el efecto *everflux*, que se refleja sobretodo en Google.

Es por esto, que el **posicionamiento SEO** intenta buscar la forma de adecuarse a ese ritmo. El SEO desconoce los factores que influyen en los algoritmos de los buscadores. El algoritmo de Google es un secreto más enigmático que el de la fórmula de la propia Coca-Cola. Casi 200 factores (search engine ranking factors) son tenidos en cuenta en este algoritmo (recordemos que a su vez cambiantes), y uno de ellos, importante para tenerlo en cuenta hacer unos años, era el PageRank.

---

<sup>11</sup> Link Building: Construcción de enlaces manual. Se explicará en el punto 2.2.c

<sup>12</sup> ScrapeBox: Programa para automatizar la escritura de comentarios en miles de webs.

<sup>13</sup> Proxy: intermediario en las peticiones de recursos que realiza un cliente (A) a otro servidor (B)

<sup>13</sup> Footprints: Comandos que podemos teclear en Google para realizar búsquedas específicas.



#### d. PageRank™

En ciertas ocasiones, se ha confundido el algoritmo de Google con el PageRank, aunque hoy en día, este concepto ya no se usa y está desactivado. Por motivos comerciales (y con ello de posicionamiento SEO), el PageRank de un sitio era fácilmente manipulable con técnicas de spam: backlinks desde otras páginas, blogs, libros de visitas, foros, etc. que alteraban la puntuación dada por Google a una página.

El PageRank estaba basado en el modelo de Science Citation Index<sup>14</sup>, confeccionado por Eugene Garfield del Instituto para la Información Científica durante la década de los 50. Tal modelo, ayudaba a los investigadores a tener un cálculo según los factores de impacto, basados en referencias bibliográficas de otros trabajos. Finalmente, el objeto de análisis residía en ver qué entes habían acumulado más popularidad.

Para evitar la manipulación de su algoritmo, en 2005, Google implementó un nuevo atributo “rel=“nofollow” “ para hiper-enlaces en un intento de derrotar al spam masivo. De esta manera, se evitaba que páginas con alto índice de visitas y PageRank derivaran autoridad hacia páginas de dudosa calidad.

Para que este concepto se entienda, se propone un ejemplo. En lenguaje HTML, el código de los enlaces tiene la siguiente estructura:

```
<a href="http://blog.example.com">Mi blog</a>
```

Donde <http://blog.example.com> es la web destino y “Mi blog” es el **anchor text** que aparece una vez se ha ejecutado el código en el navegador del usuario y que permite al usuario acceder a <http://blog.example.com>.

[Mi blog](#)

Fig. 3: Ejemplo de enlace Dofollow. Fuente: Propia

---

<sup>14</sup> Trends in research on global climate change: A Science Citation Index Expanded-based analysis [[http://www.researchgate.net/profile/Yuh-Shan\\_Ho/publication/259892846\\_Trends\\_in\\_research\\_on\\_global\\_climate\\_change\\_A\\_Science\\_Citation\\_Index\\_Expanded-based\\_analysis/links/oc96052f5006d10409000000.pdf](http://www.researchgate.net/profile/Yuh-Shan_Ho/publication/259892846_Trends_in_research_on_global_climate_change_A_Science_Citation_Index_Expanded-based_analysis/links/oc96052f5006d10409000000.pdf)]

Yuh-Shan Ho [Consulta: 4-6-2015]

Si una web introduce este código, cuando la araña de Google rastree dicha web, entiende que si una web A, cita a una web B, es porque confía en esta web. Con lo cual se traspasa parte de la autoridad de la web A a la web B. En los últimos años, se ha abusado de este hecho y muchos **webmasters** se dedicaban a dejar enlaces en páginas de alta autoridad de manera masiva sin importar lo más mínimo en que la web anfitrión era de la misma temática que la web destino. Es decir, si un webmaster de una web de zapatos, dejaba un enlace como el de arriba en un comentario en un **post** de belleza con un PageRank elevado, parte de la autoridad iba a parar a la web de calzado.

Por el uso abusivo como este, Google decidió usar el atributo Nofollow (No seguir), para que diversas webs pudiesen implementarlo en sus portales para que los comentarios que tuviesen enlaces, o widgets con enlaces, banners, no transmitiesen autoridad a esas webs, ya que no merecían la carga de autoridad que tenía página anfitrión. A partir de entonces, los enlaces tendrían esta forma:

```
<a href="http://blog.example.com" rel="nofollow">Mi blog</a>
```

Y cara al público quedaría sin ningún cambio:

[Mi blog](#)

Fig. 4: Ejemplo de enlace Nofollow. Fuente: Propia

De esta manera, se evita que la araña del buscador rastree la página destino. Según Matt Cutts<sup>15</sup>, Director del Departamento contra el Spam en Web de Google:

“Originalmente, el atributo nofollow aparece en la metaetiqueta de nivel de página e indica a los motores de búsqueda que no sigan (es decir, que no rastreen) los enlaces de salida (marcados con ese atributo) de la página.”

La verdad en todo este mundo reside en la idea de que la araña de Google, mientras trabaja, crece y recibe *feedback* de forma constante, cual ser vivo, para adaptarse a las nuevas formas de engaño que puedan surgir día a día.

---

<sup>15</sup> Cómo usar rel="nofollow" para enlaces específicos  
[<https://support.google.com/webmasters/answer/96569?hl=es#>] Cutts M.  
[Consulta: 4-6-2015]

Según Matt Cuts, el PageRank dejó de usarse hace tiempo por Google. Pero, ¿Quién nos dice que esto sea verdad? Si el PageRank era una unidad de medida de autoridad de un portal y dejó de usarse, habrá que buscar otras maneras de medir el crédito del que dispone una web.

## 2.2 Posicionamiento SEO

### a. Importancia de un buen posicionamiento

Como ya sabemos, millones de personas buscan información en internet a diario. Un SEO que lleva una web quiere que su sitio sea visible a la hora de que un usuario busque un determinado tipo de información, ya sea por intereses comerciales o simplemente por tener autoridad.

Debemos conocer dos conceptos importantes: el posicionamiento SEO y la indexación: el primer concepto consiste en colocar un sitio web entre los primeros puestos de una keyword determinada. El segundo concepto consiste en conseguir que una página aparezca en las SERPs usando las herramientas necesarias.

Diferenciamos también entre 2 tipos de posicionamiento: el orgánico, filtrado por el algoritmo de autoridad del buscador y con ello gratuito, y el publicitario, que permite colocar anuncios de pago en los resultados de los buscadores web y se diferencian porque se sitúan al principio y a la derecha de la página de resultados. Estos anuncios se colocan en las posiciones que según los estudios de calor, son los más visualizados por el ojo humano.

El SEO o posicionador SEO, debe tener conocimientos en las siguientes disciplinas

1. Altas nociones de marketing en internet, buscadores y su usabilidad: Esto incluye el conocimiento referente a los buscadores, sobre como funcionan y como trabajan a la hora de indexar contenido, cual es la forma más rápida y eficiente de conseguir la máxima valoración por parte del algoritmo de los motores de búsqueda.
2. Conocimientos sobre accesibilidad, diseño y programación web: Entender el lenguaje HTML a la hora de trabajar con páginas web, así como las hojas de estilo (CSS).
3. Facilidad con la redacción a la hora de necesitar un volumen amplio de palabras. Tener habilidad a la hora de generar contenido es otro de los talentos que debe poseer un SEO. Con los últimos algoritmos de Google, se buscan páginas llenas de contenido de calidad, tarea básica para el SEO.



4. Desenvoltura a la hora de trabajar con grandes flujos de datos y estadísticas: El SEO debe dominar todo lo relacionado con las estadísticas web (Analytics) y métricas de publicidad. También deberá revisar las recomendaciones que las herramientas de *webmaster*<sup>16</sup> le sugieran.

Diferenciar SEO y SEM (Search Engine Marketing) es vital, ya que el SEO abarca el sector de las búsquedas orgánicas y el SEM el de la publicidad inserida en las SERPs:

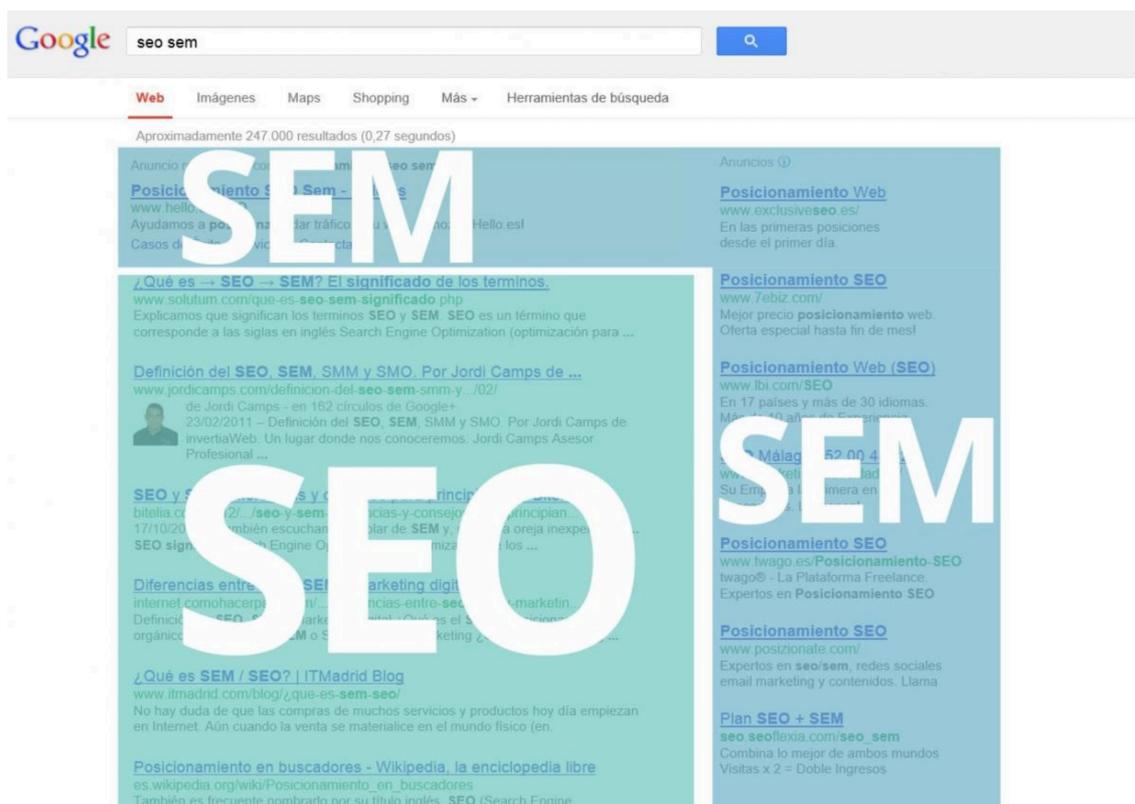


Fig. 5: Distribución del SEO y del SEM en las SERPs.

Fuente: <http://youmarketing.com/wp-content/uploads/2014/02/seo-vs-sem-youmarketing-social-media-sevilla.jpg>

<sup>16</sup> Herramientas de Webmaster: Web que proporciona Google para que los Webmasters (Gestionador de la web) puedan ver sus posiciones en la SERPs y puedan ayudar a Google a entender mejor la estructura de su web a través del envío de mapas (Sitemaps). [Consulta 5-6-2015]]

## b. Como mejorar el posicionamiento

Antes de empezar a posicionar, debemos aprender a diferenciar dos conceptos:

### **Factores ON-page:**

Son aquellos que tienen que ver con el contenido de la página web. Entre otras cosas, tiene que ver con la densidad de las palabras clave que se quiere posicionar. Esto se obtiene:

Densidad = Número de veces que aparece nuestra frase o Keyword / número de palabras totales

Según los mejores manuales SEO<sup>17</sup>, es importante tener una  $\rho$  de alrededor del 3% y de no sobrepasar el 6%, ya que en ese caso, el buscador podría tomar la web como spam (por motivos de Keyword Stuffing).

Una herramienta que permite ver tal densidad es:

<http://www.posicionar-web.info/herramientas-seo/densidad-palabras-clave/>

Por otra parte, en 2011, Matt Cutts advirtió que no hay un porcentaje de densidad de palabras clave mínimo. Tampoco lo hay máximo. Directamente, descartaba por completo la idea de “densidad de palabras clave”, asegurando que “no por mucho repetir tus palabras clave vas a posicionarte mejor en Google”. Matt Cutts considera hoy en día que las palabras clave deben ser usadas con naturalidad<sup>18</sup>. Se recomienda que el uso de estas keywords deban aparecer en:

- URL de nuestras páginas y si es en el dominio, aún mejor.
- Títulos de la página (H1, H2, H3)
- Tags o etiquetas de nuestras publicaciones.
- El primer y segundo párrafo, para establecer desde el principio la temática de la página.
- Los campos “title” y “alt” de las imágenes con las “keywords”
- Puestos en formatos como la negrita ya que les darán más importancia en el texto

---

<sup>17</sup> Orense- Fuentes M. & Rojas-Orduña O. (2010). SEO, Cómo triunfar en los buscadores: Explicación de las técnicas Black Hat SEO más comunes. Madrid: ESIC. [Consulta: 5-6-2015]

<sup>18</sup> What is the ideal keyword density of a page?  
[<https://www.youtube.com/watch?v=Rk4qgQdp2UA>] Cutts M [Consulta 5-6-2015]



## Factores OFF-page

Estos son totalmente lo contrario a los factores anteriores, es decir, son ajenos al contenido de la página. Se basan en la popularidad de la web, en la viralidad que puedan tener. Cuanto más noble sea la naturaleza de los backlinks, mejor situados en las SERPs se estará. Google, en sus últimas actualizaciones, está dando cada vez más importancia a los enlaces de retroceso<sup>19</sup> y a sus anchor text y con ello, penaliza a los portales que lo usan de manera desproporcionada.

Los tres factores OFF-page más importantes en el verano de 2015 son:

- **CTR (Click Through Ratio):** Porcentaje de clicks. Se basa en el número de clicks que recibe una web, en proporción de sus impresiones en las SERPs. Es decir: si una web aparece en las SERPs 1000 veces, pero la gente solo ha hecho click 50 veces, entonces tenemos  $(50/1000)*100 = 5\%$ . A mayor número de clicks, mayor será nuestra reputación en Internet. Es por esto que tanto el título de la página como la “meta description” deben ser atractivos. De esta manera, aumentaremos nuestro posicionamiento en los buscadores.
- **Índice de popularidad:** Depende de los vínculos externos que apunten hacia nuestra web. Otra forma de aumentar el número de visitas es con la contratación de anuncios, aunque eso suponga un coste económico.
- **El botón +1 de Google Plus,** debe estar incorporado en la web, ya que es una funcionalidad de Google que da la posibilidad al visitante o usuario de que vote el portal si lo considera oportuno. Google Plus, al pertenecer al propio Google, es la red social perfecta para ayudar a mejorar el posicionamiento y mantener una web en los primeros resultados en una búsqueda (no quita que Facebook, Twitter no sean importantes).

---

<sup>19</sup> Backlinks

Los SEO se dedican a programar sus webs de manera que queden lo más “amigables” cara a Google. Es decir, adaptan su web pensando qué puede ser lo mejor para que el buscador piense que se trata de una web que acierta en todos los puntos de su algoritmo.

Para hacer un símil, en el argot del vino, se dice que “hay algunos productores que "parkerizan" su vino, es decir, lo adaptan a lo que creen el gusto de este crítico (Robert Parker) con la finalidad de obtener una alta puntuación en sus libros”



### c. Link Baiting y Link Building

En ciertas ocasiones, conseguir enlaces por parte de otros portales suele ser complicado. Generar enlaces ayuda al robot de Google a ver el desarrollo de una web y la evolución de esta a medida que otros portales apuntan hacia ella. Cuando un portal no recibe enlaces es por motivos como:

- Ser poco conocido
- No aparecer en las SERPs
- Tener poca autoridad
- Tener contenido de baja calidad

Es bueno pues, empezar a dejar pequeñas semillas en diversos lugares de la web y cuanto más autoridad tengan los portales, mejor. Como se ha dicho, esto a veces no es fácil, pero si se sabe diseñar un portal y el contenido es de calidad de manera que el este se mueva por las redes sociales, estaremos entrando en la técnica del *link bait*. El *link baiting* consiste pues, en la creencia de que el si el argumento y el contenido son de calidad, por si solo será compartido en las redes sociales y además será referenciado por otras webs del sector.

Por otra parte el *Link building* o “construcción de enlaces” es una técnica que se basa en la generación de enlaces de forma manual. Se pretende con esto aumentar el número de enlaces que vienen desde diferentes portales, pero a diferencia del link baiting, estos se hacen de manera artificial. Este proceso se divide en:

- Alta en buscadores y directorios: consiste en darse de alta en los directorios de blogs y webs. Se recomienda hacerlo de manera natural y nunca de golpe. Por ejemplo en DMOZ, un directorio donde los moderadores son gente voluntaria que ayudan a mantener la información de la web organizada y categorizada.
- Inclusión de enlaces en Private Blog Network (PBN): Agencias e incluso particulares disponen de su propia Red Privada de Blogs donde uno puede contactar para colocar links en zonas de relevancia.
- Intercambio de enlaces: Otra opción puede ser el intercambio de enlaces entre dos o más portales. El hecho de que estos portales sean de la misma temática ayuda a que Google vea cierta afinidad entre estos. Lo que no es en absoluto lógico es que una web de venta de coches enlace a una web de venta de vinos y esta enlace a la primera.

- Usando determinadas herramientas para encontrar dominios expirados y hacerles una redirección 301<sup>20</sup> hacia nuestra web. Google entiende que si un dominio A es movido a otro B y se desea que se traspase toda la autoridad, la redirección 301 traslada hasta un 75% de la autoridad de A hasta B.
- Otra opción es la compra de enlaces ¿Pero, es la compra de enlaces penalizable?

En las políticas de Google leemos sobre la compra-venta de enlaces<sup>21</sup>:

“No todos los enlaces de pago vulneran nuestras directrices. La compra y venta de enlaces es una actividad económica normal en la Web cuando se realiza para fines publicitarios y no para la manipulación de los resultados de búsqueda. Los enlaces que se adquieren para publicidad se deben identificar como tales. Esto puede realizarse de diversas maneras:

1. Añadiendo el atributo rel="nofollow" a la etiqueta <a href>
2. Redirigiendo los enlaces a una página intermedia que se encuentra bloqueada para los motores de búsqueda con un archivo robots.txt”

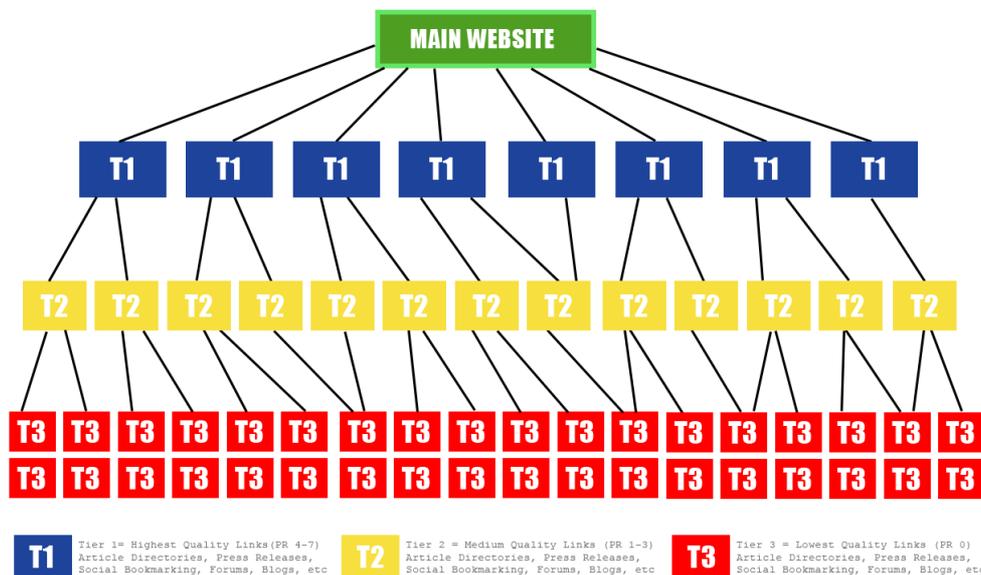


Fig 6: Técnica de Link Building piramidal.

Fuente: <http://posicionamientoweb.cat/seo/esquema-tier-linkbuilding-niveles.png>

<sup>20</sup> Redirección 301: Las redirecciones 301 son comandos que permiten enviar a usuarios y buscadores de una URL a otra automáticamente

<sup>21</sup> Orense- Fuentes M. & Rojas-Orduña O. (2010). SEO, Cómo triunfar en los buscadores: Explicación de las técnicas Black Hat SEO más comunes. Madrid: ESIC.



# 3. Material y Métodos

## 3.1 Material

Como este trabajo pretende centrarse en análisis de los backlinks (SEO OFF-page), tenemos a nuestro alcance una serie de herramientas que ayudarán al SEO a examinar todos los hilos que se tejen hacia una web. Cualquier SEO deberá disponer de una o diversas herramientas de las que se habla a continuación, ya que sin ellas, realizar una base de link building en una web será tarea difícil.

### a. Ahrefs

Esta potente herramienta nos permite examinar los backlinks de las webs que introduzcamos. Con ello, se puede estar al día de los backlinks entrantes a nuestra web o lo que es mejor, a nuestra competencia. Ahrefs puede llegar a indexar miles de webs nuevas en cuestión de minutos, y cara al usuario, se actualiza cada 30 minutos. Poseer esta herramienta no es sinónimo de triunfo en el ámbito del SEO, ya que hay que saber como usarla.

En la sección Site Explorer, una interfaz muy amigable para el usuario básico, podremos introducir la web a analizar:

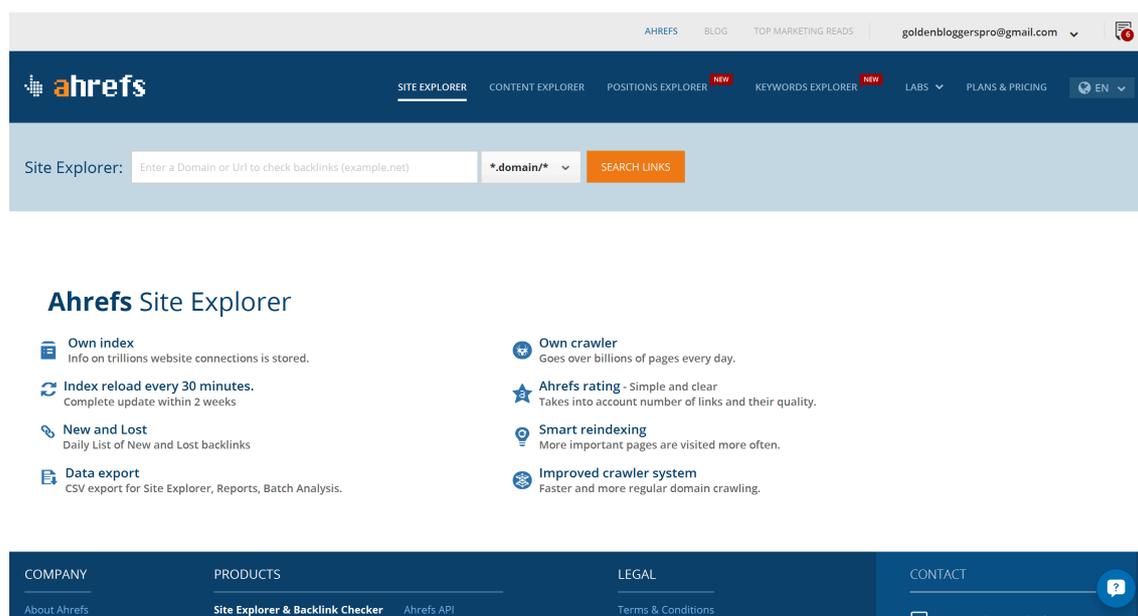


Fig. 7: Pantalla principal de Ahrefs. Fuente: <https://ahrefs.com/>



Se propone como ejemplo que se desea que se devuelva, a modo de resumen rápido, información de la tienda online: [www.bodeboca.com](http://www.bodeboca.com)

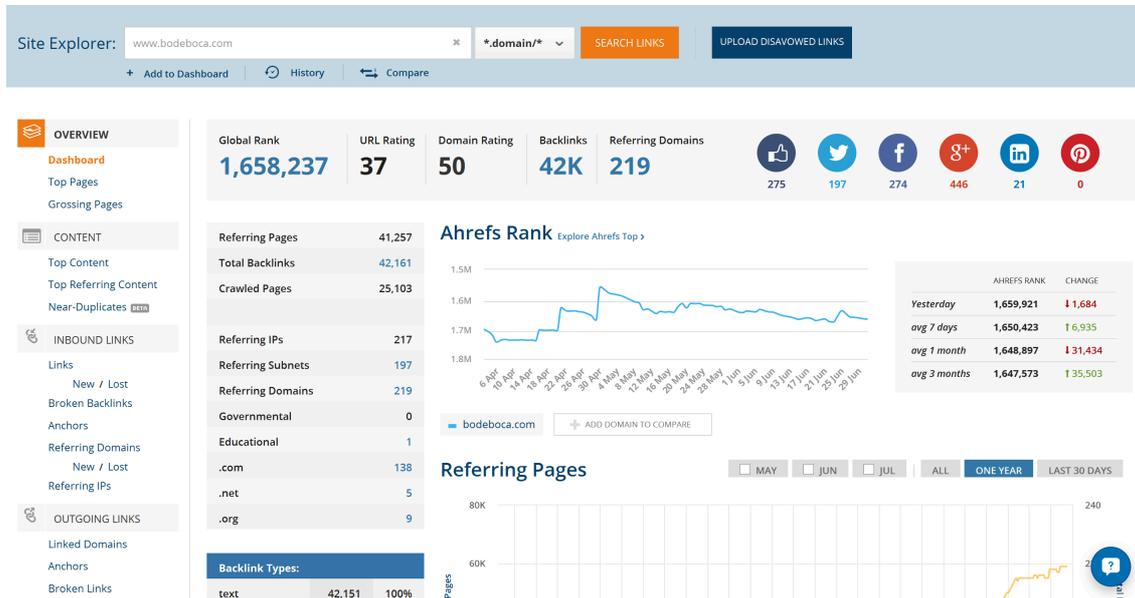


Fig. 8: Pantalla del Site Explorer de Ahrefs. Fuente: <https://ahrefs.com/>

Con esto podemos interpretar muchas cosas:

	Valor	Lectura
<b>Ranking Global</b>	1,658,237	Posición para Ahref en su ranking global de páginas web
<b>Puntuación de URL</b>	37	Puntuación sobre 100 de la página <a href="http://www.bodeboca.com">www.bodeboca.com</a>
<b>Puntuación de Dominio</b>	50	Puntuación sobre 100 del dominio <a href="http://bodeboca.com">bodeboca.com</a>
<b>Interacciones sociales</b>	275 Likes, 197 Tweets, 274 Shares, 446 +1, etc	Nos indica cuantas interacciones sociales ha tenido esa web gracias a los botones sociales.
<b>Número de backlinks</b>	42K (42.000)	Indica cuantas webs apuntan hacia este portal. En este caso son muchas las webs que han decidido poner un backlink nombrando <a href="http://www.bodeboca.com">www.bodeboca.com</a>

Fig. 9: Tabla resumen de los valores dados por Ahrefs. Fuente: Propia

Se puede observar en la gráfica que la web tiene un crecimiento regular y natural en lo que a ranking de la web se refiere. Sospechoso sería ver picos fuertes en muy poco tiempo, cosa que indicaría que se ha estado haciendo linkbuilding masivo. Esta información, permite saber qué lugares pueden ser penalizados por el algoritmo de Google.

En la columna izquierda, se nos permite acceder a la sección de “Links”, la más interesante de esta herramienta y donde se pueden examinar los backlinks de una web.

ALL 42,169		DOFOLLOW 41,716	REDIRECT 3	GOVERNMENTAL 0	EDUCATIONAL 1	NOTSITEWIDE 42,169	NOFOLLOW 443	SITEWIDE 0
1 - 50 of 42,169 backlinks found								
Referring Page Title	Referring Page URL	DR	UR	Int	Ext	Link URL	Link Anchor	Last Seen
1. Mola.com	<a href="http://www.mola.com/es/">www.mola.com/es/</a>	8+ 11	49	34	7	<a href="http://www.bodeboca.com/">www.bodeboca.com/</a>	Bodeboca	17 Jun '14 13 d
2. En Alquiler.com alquila vacaciones, eventos, equipos, remolques...	<a href="http://www.alquiler.com/">www.alquiler.com/</a>	8+ 16	42	34	75	<a href="http://www.bodeboca.com/">www.bodeboca.com/</a>	Bodeboca	8 Sep '14 37 h
3. Mola.com	<a href="http://mola.com/es/">mola.com/es/</a>	8+ 0	49	26	7	<a href="http://www.bodeboca.com/">www.bodeboca.com/</a>	Bodeboca	2 Jul '14 22 d
4. Mola.com	<a href="http://www.mola.com/en/">www.mola.com/en/</a>	8+ 9	49	25	7	<a href="http://www.bodeboca.com/">www.bodeboca.com/</a>	Bodeboca	22 Jun '14 12 d
5. CATANATURA HERBOLARIO - Google+	<a href="https://plus.google.com/116597691114100317753">plus.google.com/116597691114100317753</a>	8+ 0	94	18	44	<a href="http://www.bodeboca.com/user/registre...">www.bodeboca.com/user/registre...</a>	Bodeboca   Regístrate gratis y sin compromiso, y disfruta de las ventaja...	27 Apr '14 188 d
6. Aquilino Peña: "Las grandes compañías de España no se toman en serio el e-c...	<a href="http://www.territoriocreativo.es/etc/2012/02/aquilino-companias-espana-e-commerce-tcdesayunos.html">www.territoriocreativo.es/etc/2012/02/aquilino-companias-espana-e-commerce-tcdesayunos.html</a>	8+ 23	58	16	113	<a href="http://www.bodeboca.com/">www.bodeboca.com/</a>	mal". El e-commerce tiene que desintemediar. Como ejemplo, Bodeboca , una de sus participadas, en la que	10 Aug '13

Fig. 10: Pantalla de backlinks de Ahrefs. Fuente: <https://ahrefs.com/>

Esta sección viene en un formato de tabla, en la que cada columna nos indica una atributo diferente:

- Título de la página donde origen ([mola.com](http://mola.com)) y URL de esta: ([www.mola.com/es/](http://www.mola.com/es/))
- Interacciones sociales de la web origen (11 +1, 247 tweets, etc)
- Puntuación del dominio (49)
- Puntuación de la URL (34)
- Links entrantes y salientes de la web destino (7 y 59 respectivamente)
- URL destino del backlink ([www.bodeboca.com/](http://www.bodeboca.com/)) y el anchor usado (“Bodeboca”)



También podemos tomar datos curiosos que no podríamos ver de otras maneras como el total de enlaces y como son en su totalidad:

ALL 42,169	DOFOLLOW 41,716	REDIRECT 3	GOVERNMENTAL 0	EDUCATIONAL 1	NOTSITEWIDE 42,169	NOFOLLOW 443	SITEWIDE 0
---------------	--------------------	---------------	-------------------	------------------	-----------------------	-----------------	---------------

Fig. 11: Pantalla de backlinks de Ahrefs. Fuente: <https://ahrefs.com/>

El hecho de que la web tenga 41,716 enlaces sin el atributo rel="nofollow" da a entender que todos esos enlaces transmiten autoridad. Y solo 443 tienen el atributo rel="nofollow", probablemente enlaces en comentarios o en anuncios de publicidad (periódicos, blogs, etc). El ratio de enlaces dofollow y nofollow puede dar bastante información de la calidad de una web, en este caso:

- 41.716 -> 99%
- 443 -> 01%

Todo esto, tratándose de una web que se posiciona en el 6 lugar para la palabra clave "vino".

Los enlaces "Redirect" son aquellos enlaces que simplemente son redirecciones. Los de tipo "Educational", son aquellos que provienen de entidades educativas, como universidades o .edu. Google da mucho valor a estos backlinks<sup>22</sup>.

En la sección Referring Domains podremos ver datos muy importantes, ya que indican cuantos dominios han apuntado a una web. El primer enlace que se hace desde una web es el más importante de todos. Aquí prevalece la variedad sobre la cantidad. Es decir, es mejor 50 dominios con 1 enlace cada uno, que 50 enlaces del mismo dominio. Aplicando los filtros New/Lost podremos ver enlaces nuevos y perdidos.

---

<sup>22</sup> Are .Edu and .Gov Links Really Worth it? [<http://tripleseo.com/are-edu-gov-links-worth-it/>] Dyson C. [Consultado 7-6-2015]

1 - 50 of 219 results | Export | Settings

ALL 219		DOFOLLOW 177	REDIRECT 1	GOVERNMENTAL 0	EDUCATIONAL 1	NOTSITEWIDE 219	NOFOLLOW 43	SITEWIDE 0
ALL TLDs: 219								
#	Referring Domain	DR	Ahrefs Rank	Backlinks		First Seen	Backlinks	
				Dofollow	Total			
1.	<a href="#">alquiler.com</a>	42	-	31,589 / 31,589		1 Apr '15	BACKLINKS	
2.	<a href="#">mola.com</a>	49	2,353,040	5,894 / 5,894		30 Mar '15	BACKLINKS	
3.	<a href="#">bodeboca.co.uk</a>	37	-	2,904 / 2,905		22 Mar '15	BACKLINKS	
4.	<a href="#">diariodelvino.com</a>	48	2,863,721	340 / 340		11 Jun '15	BACKLINKS	
5.	<a href="#">san-alejandro.com</a>	48	2,957,961	259 / 259		26 Mar '15	BACKLINKS	
6.	<a href="#">silktide.com</a>	63	23,555	98 / 396		18 May '15	BACKLINKS	
7.	<a href="#">eccion.es</a>	45	7,432,887	98 / 98		28 Mar '15	BACKLINKS	
8.	<a href="#">gastronomiayunapizca.com</a>	47	4,697,055	51 / 51		22 Mar '15	BACKLINKS	
9.	<a href="#">acatarlotodo.com</a>	37	-	34 / 34		19 Mar '15	BACKLINKS	
10.	<a href="#">forodecrueros.com</a>	43	-	32 / 32		25 Mar '15	BACKLINKS	
11.	<a href="#">alternatiba.com</a>	45	7,072,869	19 / 19		19 Mar '15	BACKLINKS	

Fig. 12: Pantalla de dominios de referencia de Ahrefs. Fuente: <https://ahrefs.com/>

La versión gratuita, permite la consulta de 3 URLs por día pero devuelve información muy limitada. Se recomienda a cualquier persona que esté inmersa en una campaña de SEO contratar, por lo menos un mes, una cuenta premium, y así exportar los datos en **CSV**, tanto de su web como los de la competencia.

Los precios del usuario premium van desde los 79 dólares (71 euros al cambio) hasta los 1,299 dólares (1171 euros) mensuales dependiendo de la suscripción.

		LITE	STANDARD	PLUS	PREMIUM
		The best if you work with small websites and do not use advanced features.	The best for small to medium websites. All features are included.	The best for medium to big websites. All features are included.	Get up to 1 billion backlinks per month. All features are included.
			Most Popular		
Monthly	Annual 20% discount	\$79 /mo	\$179 /mo	\$499 /mo	\$1,299 /mo
User - <i>Need more users?</i>		1	1	1	1
Total Data Rows Limit		5M	30M	125M	1,000M
Crawl Reports		15	60	180	600
Social Metrics		✓	✓	✓	✓
Positions Explorer		10 rows	30K rows	50K rows	100K rows
Content Explorer		20 rows	✓	✓	✓
Alerts		5	10	40	100
Site Explorer		✓	✓	✓	✓
Backlink Analysis		✓	✓	✓	✓
Broken Backlinks & Links		5 rows	✓	✓	✓
Content Analysis		5 rows	✓	✓	✓
		SUBSCRIBE	SUBSCRIBE	SUBSCRIBE	SUBSCRIBE

Fig. 13: Pantalla de precios del usuario premium de Ahrefs.

Fuente: <https://ahrefs.com/>

**Conclusión:** Ahrefs es una herramienta muy útil a la hora de hacer una campaña de linkbuilding en una web. Es muy difícil no sacarle provecho ya que puede mostrar enlaces realmente valiosos y que se pueden replicar. Un precio justo por un servicio justo.

## b. Semrush

Semrush es otra de las herramientas más importantes del mercado. Se trata de un instrumento perfecto para monitorizar las palabras clave y analizar el comportamiento de la competencia. Aunque en diversos foros se comenta que es una herramienta dedicada al SEM (Search Engine Marketing o publicidad contextual), se puede extraer un gran volumen de datos, como keywords o Long Tails<sup>23</sup> con las que está trabajando la competencia, tanto a nivel orgánico como a nivel de publicitario. Esto nos permitirá descubrir cuales son las mejores keywords y obtener nuevas ideas para aumentar nuestro tráfico

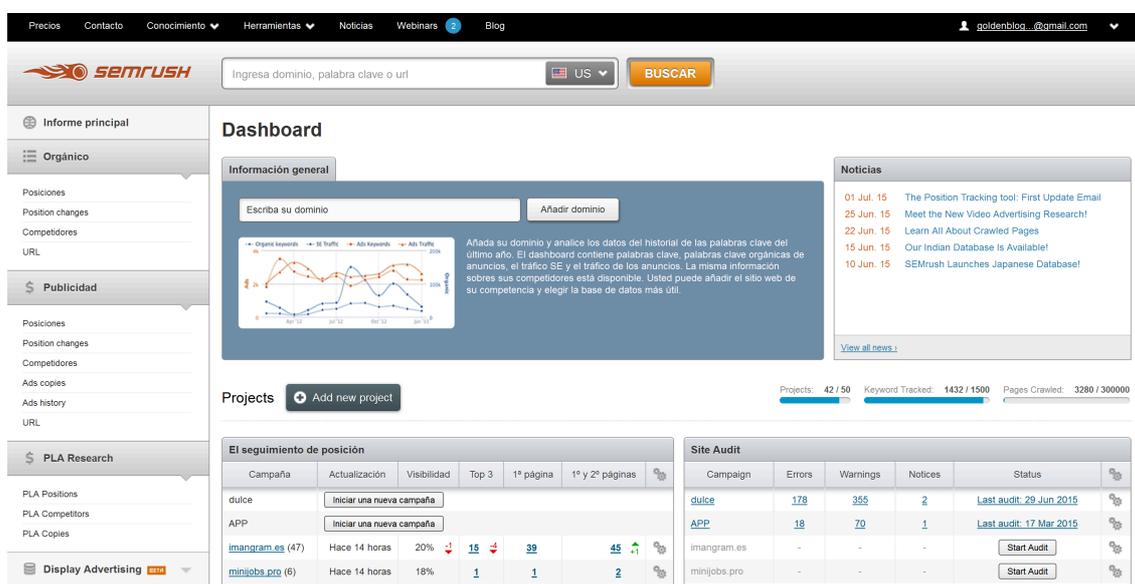


Fig. 14: Pantalla principal de Semrush. Fuente: <http://es.semrush.com/>

La web ofrece una interfaz muy fácil de asimilar y de entender. Arriba, nos da opción para introducir una web y que nos devuelva los datos recopilados en su base de datos. En este caso, hemos analizado la web <http://www.wines.com/>.

<sup>23</sup> Long tail: Es un término que hace referencia a la utilización de 3 o más frases o palabras como texto ancla para la construcción de enlaces de retroceso.

## Estudio de Link Network Analysis y su influencia en el posicionamiento de buscadores

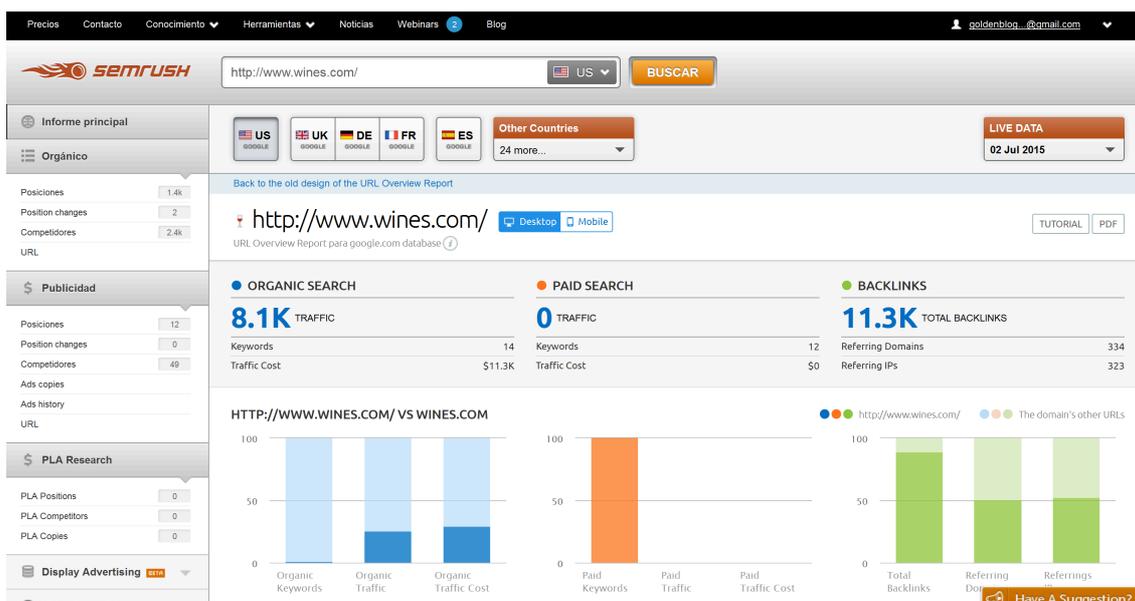


Fig. 15: Pantalla de análisis. Fuente: <http://es.semrush.com/>

Al igual que Ahrefs, se presenta un pequeño resumen con los datos más importantes. En este caso, al lado del recuadro de introducción de la URL, permite seleccionar el país con el que queremos trabajar. Para búsquedas a nivel global y EEUU, se usa la pestaña US ([google.com](http://google.com)). Para el caso de España, la ES ([google.es](http://google.es)). El problema de Semrush es que ver a simple vista el porcentaje de enlaces dofollow y nofollow es un poco más complicado.

A primera vista, podemos obtener algunos datos:

	Valor	Lectura
<b>Organic Search</b>	8.1K	Cantidad de visitas que recibe a través de las búsquedas orgánicas de Google
<b>Keywords</b>	14	Número de palabras clave que tiene este portal posicionando en las primeras 20 posiciones
<b>Traffic Cost</b>	11.3K	Costo de la transferencia de usuarios valorado en 11.300 dólares.
<b>Paid Search</b>	0	Pago por SEM
<b>Backlinks</b>	11.3K	Número de links que apuntan a la web
<b>Referring Domains</b>	334	Número de dominios que apuntan a la web

Fig. 16: Tabla resumen de los valores dados por Semrush. Fuente: Propia

Si se desea ver estas keywords, se puede entrar en una ventana que permite la visualización en tablas. Ofrece datos muy concretos de la web respecto a las keywords. Esta información es muy valiosa para la toma de decisiones: saber en qué keywords invertir, cuales descartar por poca afluencia de visitas, precios del CPC, etc.

En la columna izquierda se puede analizar la web tanto en modo orgánico como en modo publicitario.

ORGANIC SEARCH POSITIONS 1 - 14 (14) 

 Exportar

Palabra clave	Pos	Volumen	CPC	Com.	Tráfico %	Costos %	Resultados	Tendencia	Última actualización	SERP source
<a href="#">wine</a>	4	90,500	1.55	0.22	20.33	25.45	772,000,000		12 hr ago	
<a href="#">wine.com</a>	10	12,100	0.30	0.18	0.23	0.05	147,000,000		11 hr ago	
<a href="#">wines</a>	2	8,100	1.04	0.44	3.37	2.83	239,000,000		11 hr ago	
<a href="#">internet wines</a>	11	480	0.96	0.25	0.01	0.01	3,820,000		17 Jun 2015	
<a href="#">wines.com</a>	1	480	0.59	0.36	0.72	0.34	18,900,000		17 Jun 2015	
<a href="#">www.wine.com</a>	2	390	0.56	0.46	0.03	0.01	767,000,000		18 Jun 2015	
<a href="#">wine.com reviews</a>	9	170	2.92	0.08	0.01	0.03	327,000,000		21 Jun 2015	
<a href="#">www.wine</a>	3	110	1.32	0.45	0.03	0.03	150,000,000		25 Jun 2015	
<a href="#">wines on line</a>	19	50	3.27	0.77	0.00	0.00	93,600,000		18 Jun 2015	
<a href="#">internet wine</a>	4	50	2.11	0.76	0.00	0.00	293,000,000		19 Jun 2015	
<a href="#">internet wines.com</a>	7	50	1.13	0.93	0.00	0.00	3,680,000		19 Jun 2015	
<a href="#">wine.com review</a>	9	40	13.16	0.08	0.00	0.04	392,000,000		23 Jun 2015	
<a href="#">www.wines.com</a>	1	20	1.06	0.73	0.03	0.02	108,000,000		8 Jun 2015	
<a href="#">wine wines</a>	16	20	0.00	0.19	0.00	0.00	104,000,000		10 Jun 2015	

Fig. 17: Pantalla de palabras clave de wines.com. Fuente: <http://es.semrush.com/>

Lo bueno de Semrush es que ofrece una API que permite que integremos muchos elementos de la herramienta en la web si somos los webmasters autorizados. La base de datos cuenta con más de **120.000.000** palabras clave analizadas y más de 98.000.000 dominios analizados.

Permite exportar a informes en formato CSV y compara dominios para ver las características principales entre la web a posicionar y las de la competencia. Esta herramienta sirve para ver cuáles son las keywords que está posicionando la competencia, pero no la página que queremos posicionar.

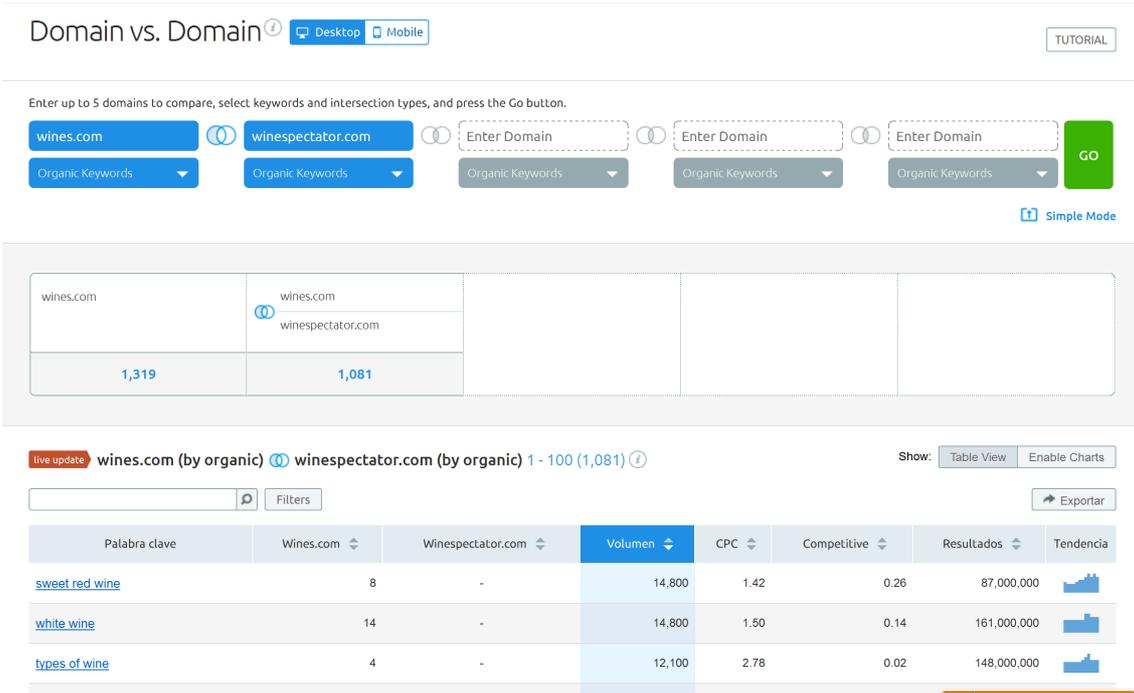


Fig. 18: Pantalla de comparación entre wines.com y winespectator.com.

Fuente: <http://es.semrush.com/>

Gracias a Semrush, se pueden examinar las Keyword por las que está compitiendo wines.com pero que no compite winespectator.com. Se observa que para las palabras “sweet red wine”, “white wine” y “types of wine”, wines.com está en las posiciones 8, 14 y 4 respectivamente en las SERPs de Google. Este estudio minucioso ayudaría a winespectator.com a realizar un mejor trabajo y en invertir en estas keywords, ya que aportan casi 40.000 visitantes mensuales. De forma visual, serían todas las palabras que están de color azul puro.

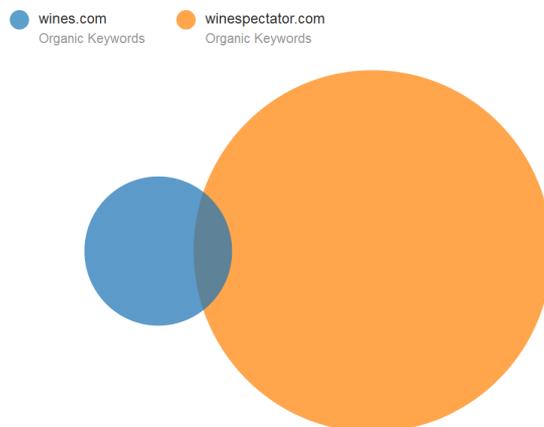


Fig. 19: Pantalla de comparación entre wines.com y winespectator.com.

A diferencia de la URL, si introducimos el dominio podremos observar que wines.com compite por muchas más palabras clave. Casi 1400 keywords con un tráfico de 31.100 visitas orgánicas mensuales.

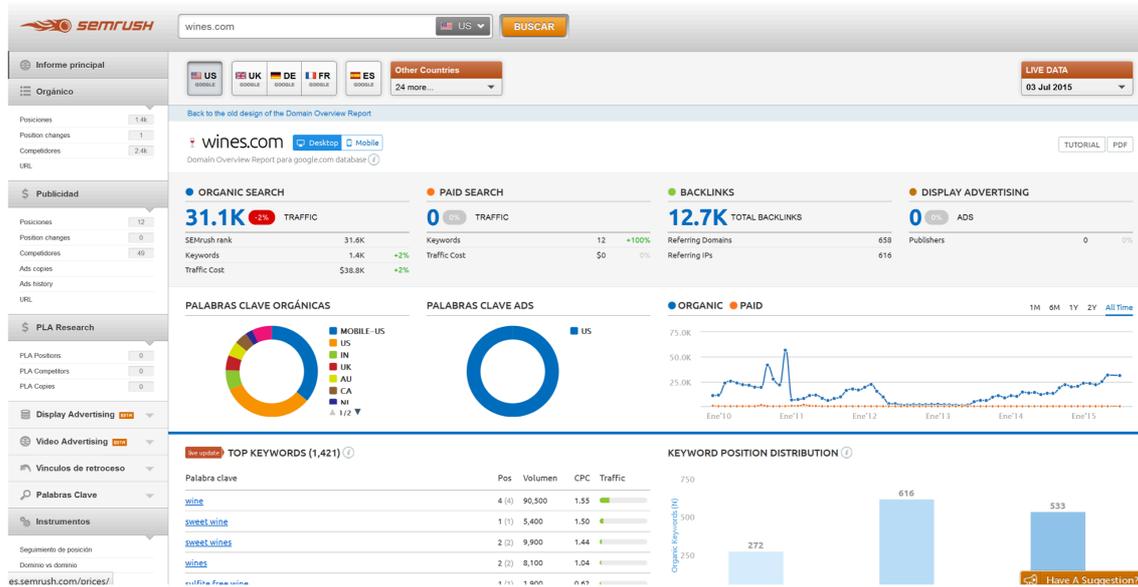


Fig. 20: Pantalla de análisis. Fuente: <http://es.semrush.com/>

Con Semrush, la monitorización de las keywords es algo fácil y rápido. Se puede tener a los competidores controlados y investigar de mejor manera el tráfico orgánico.

El gasto económico no se aleja mucho de Ahrefs. Las cuentas premium van desde los 70 hasta los 549 dólares (63 y 495 euros al cambio). Puede parecer caro, pero se trata de un instrumento para encontrar keywords que normalmente no se habrían encontrado con Ahrefs

Plan	Price (mensual)	Best Choice
PRO	\$69.95	No
GURU	\$149.95	Yes
BUSINESS	\$549.95	No

Plan	PRO	GURU	BUSINESS
<b>Análisis</b>			
Resultados por informe	10 000	30 000	50 000 (para cada uno de los 4 usuarios)
Informes al día	3 000	5 000	10 000 (para cada uno de los 4 usuarios)
<b>Proyectos</b>			
Proyectos	5	50	ilimitados
Rastreo de palabras clave	500	1 500	6 000 (1 500 per user)
Rastreo de páginas	100 000	300 000	1 200 000 (300 000 per user)

Fig. 21: Pantalla de precios del usuario premium de Semrush.

Fuente: <http://es.semrush.com/>

## c. Majestic SEO

Majestic ha sido la herramienta más completa de las tres que se han visto. Se trata de una herramienta que nos ayudará en la fase de Linkbuilding. No se trata de una herramienta cualquiera, ya que Majestic nos ofrece la posibilidad de ver la calidad de los enlaces entrantes a una web, y eso ayudará a la hora de ajustarnos al algoritmo de Google Panda.

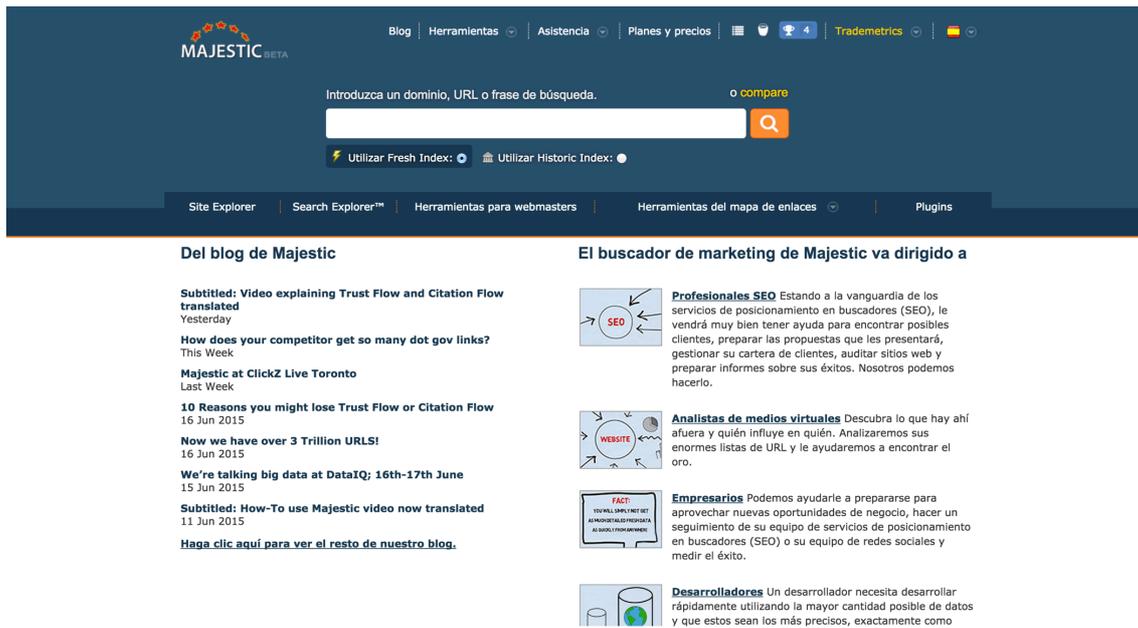


Fig. 22: Pantalla principal de Majestic. Fuente: <https://es.majestic.com/>

Para observar el funcionamiento de la herramienta, se ha procedido a analizar la web de venta de vinos <http://www.majestic.co.uk/>, (no confundir con la web de la herramienta [es.majestic.com](http://es.majestic.com))

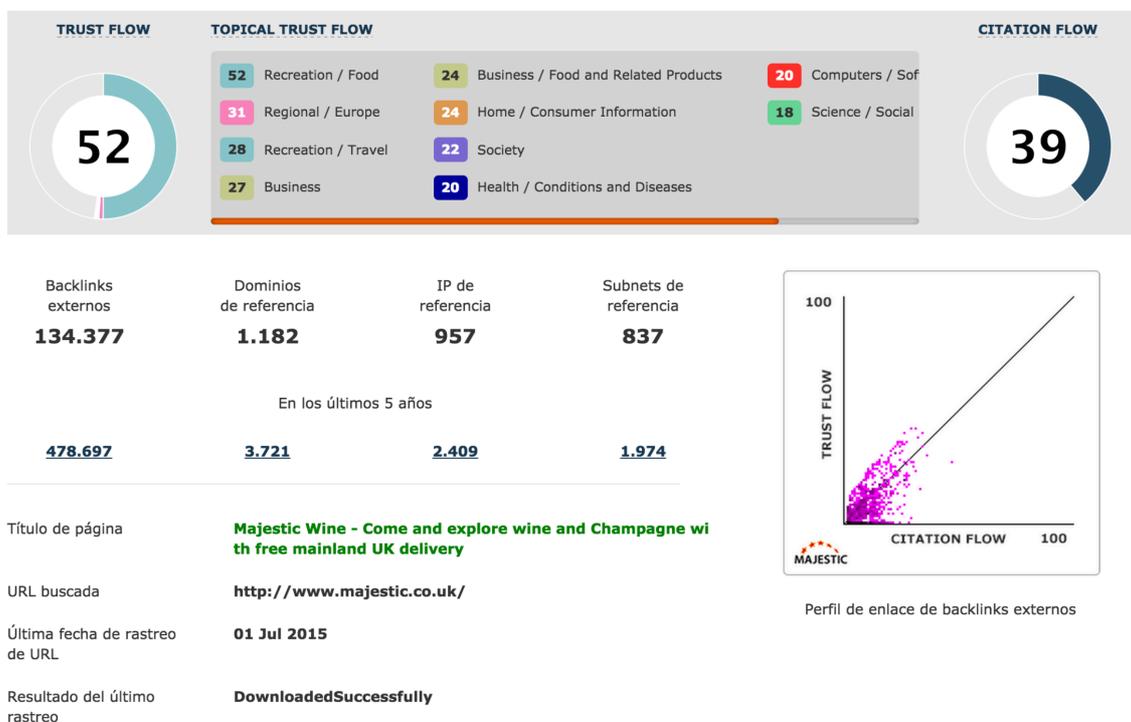


Fig. 23: Pantalla de análisis de [www.majestic.co.uk](http://www.majestic.co.uk). Fuente: <https://es.majestic.com/>

La lectura que se puede hacer de esto es la siguiente:

	Valor	Lectura
<b>Trust Flow</b>	52	Indica la puntuación que recibe según la confianza
<b>Citation Flow</b>	39	Indica la puntuación que recibe según la popularidad
<b>Topical Trust Flow</b>	52/Recreation/Food 31/Regional/Europe	Indica la puntuación que recibe según las diferentes temáticas de la web
<b>Backlinks Externos</b>	134.377	Número enlaces hacia esta web

Fig. 24: Tabla resumen de los valores dados por Majestic. Fuente: Propia

Además, se acompaña una gráfica donde se puede ver el perfil de los enlaces entrantes según su Trust Flow y su Citation Flow. Como se explicará luego, cuantos más puntos estén en el sector superior, más confianza ofrece la web.

Las métricas que ofrece Majestic son:

- **Citation Flow (CF):** Se trata de una métrica que se basa en la cantidad de enlaces entrantes. Algunos enlaces tienen más peso que otros. Se mide de 0 a 100 y está basado en una escala logarítmica. Si esto fuese así: si un portal web tiene un CF de 20 y otro portal tiene un CF de 40, el segundo no tiene el doble del primero, sino 100 veces más. Es decir, no hay una proporción lineal sino exponencial. Es por eso que es mucho más fácil pasar de una puntuación de 10 a 20 que de 80 a 90.
- **Trust Flow (TF):** A diferencia de la anterior métrica se enfoca en la calidad más que en la cantidad de los enlaces. También va de 0 a 100 puntos y se basa en una escala logarítmica. Para poder asignar una puntuación, Majestic observa la pirámide formada debajo del link a analizar. Desde las raíces, hasta la punta, Majestic descompone todos los links que llegan y examina la calidad de todos ellos, asignándole al link objetivo una puntuación adecuada.

Este sistema de puntuación se parece bastante al llamado TrustRank de Yahoo!, elaborado en 2004. La potencia del Trust Flow reside en que, a diferencia del PageRank, que medía simplemente la citación y por este motivo cayó en desgracia, el Trust (Confianza) Flow mide la calidad. En ningún caso puede tener la misma calidad una web con un texto redactado por un profesional del sector, que uno que ha “espineado” un texto y lo ha pegado.

- **Topical Trust Flow (TTF):** la última novedad de Majestic incluye también una sección dedicada la puntuación de las webs según la confiabilidad que merecen y asignándolas a algunas de las 1.000 categorías diferentes que existen. Como es sabido, el proceso de indexación de Google otorga mayor o menor peso a los portales que difunden contenido de calidad. De tal manera, hacer Linkbuilding con webs que sean de pura confianza para Google fomentará que nuestra web crezca de forma cualitativa y cuantitativamente en las SERPs de Google. De nada sirve pues, tener una web con grandes cantidades de texto especializado, si periódicos, blogs de nuestra misma temática u otros entes influyentes no nos enlazan.

Si examinamos la web <http://www.vinoseleccion.com/> podemos ver como el algoritmo de Majestic le ha asignado la siguiente notoriedad para los “topics”:

1. Shopping 38/100
2. Regional / Europe 27/100
3. Recreation / Food 26/100
4. Shopping / Food 26/100
5. Business / Food and Related Products 21/100

Con lo que se comprueba que la potencia de esta herramienta es altísima.  
Para sintetizar:

- **Citation Flow:** Cantidad de Enlaces
- **Trust Flow:** Calidad de Enlaces
- **Topical Trust Flow:** Calidad dentro de una determinada temática.

Tener un Trust Flow alto es bastante más difícil que tener un Citation Flow.  
Para poder hacernos una idea rápida de cómo de bien lo está haciendo una web podemos hacer una ratio dividiendo el TF entre el CF. Por ejemplo:

La url: <http://www.guiarepsol.com/es/gastronomia/vinos-y-bodegas/> está en la **décima** posición para la búsqueda de la palabra “vino” y tiene 35 puntos de TF y 43 puntos de CF:

$$35/43 = 0,8139$$

La url <http://www.todovino.com/> está en la **cuarta** posición:

$$43TF/32CF = 1,3437$$

La url <http://www.vinoseleccion.com/> se sitúa en **segunda** posición:

$$45/33 = 1,3636$$

La url <https://es.wikipedia.org/wiki/Vino>, **primera** para esta misma búsqueda:

$$31/22 = 1,4090$$

Se ha observado tras una serie de comprobaciones con diferentes métricas, que esta ratio puede ayudar mucho a la hora de medir la autoridad de un sitio web. No podemos afirmar al 100% que cuanto más alta sea esta ratio, más se acercará una url a la primera posición para aquella keyword por la que se está disputando. Pero sí que se puede afirmar que si esta ratio es mayor que 1 (más Trust Flow que Citation Flow) se está haciendo un buen trabajo de posicionamiento.

A diferencia de Semrush, Majestic ofrece la posibilidad de hacer análisis masivos de hasta 300 URL o hasta 10.000 URL si están en un documento de excel o txt.

Con Bulk Backlink Checker puede analizar los números de backlinks de hasta 400 URL gracias a la función de pegar URL, y de hasta 100 000 URL a partir de la carga de un archivo (en función de su categoría de suscripción).

Los resultados muestran un resumen del número total de enlaces que encontramos, junto con el número total de dominios de referencia, educativos y gubernamentales.

Esta herramienta está diseñada para comprobar **muy rápidamente** los recuentos de backlinks y dominios de referencia para los primeros 10 000 millones de URL más enlazadas de nuestra base de datos, todos los subdominios y dominios raíz presentes en nuestra base de datos.

Fig. 25: Pantalla de análisis masivo de Majestic. Fuente: <https://es.majestic.com/>

Majestic va más allá del posicionamiento en buscadores. No da en ningún momento seguimientos en las SERPs de Google, ni ofrece grandes bases de datos relacionadas con las keywords. Eso sí, tiene opciones de comparación con los competidores. Majestic trata de dar un resultado de posicionamiento en la red, es decir, trata de atribuir valores según su influencia. Cabe destacar que el posicionamiento en buscadores, es una probable consecuencia del posicionamiento de influencias.

El buen SEO es el que no necesita de Google y no se dedica solo a mirar las posiciones de las SERPs. Una buena red montada alrededor de una web es lo que dictará la posición posteriormente en las SERPs, cambie o no el algoritmo de Google en el tiempo. En esta sección se ha visto que el Trust Flow es un valor muy interesante y se usará más adelante.

## 3.2 Análisis de influencias

### a. Centralidad

Ya en los años 50<sup>24</sup> se hablaba de algo llamado “social network analysis” como técnicas para poder detectar relaciones y corrientes en la sociedad. Hacer análisis de este tipo en la sociedad podía llegar a ser una ayuda potente a la hora tomar decisiones. Según la definición que se da en el libro<sup>25</sup> “Social Network Analysis” de John Scott, se trata de un recurso de valor incalculable para los investigadores a través de las relaciones sociales, ayudando al estudio de las teorías sociales.

Para empezar, hay que tener en cuenta que cuando se estudia un grupo de personas, se analizan en su globalidad atributos, valores, cualidades, propiedades, etc. Es decir, características que permitan, posteriormente, agrupar y encontrar comunidades. Esto se consigue a través de técnicas estadísticas que estudian el comportamiento y relación de la sociedad. La centralidad de grado (degree centrality), busca los entes más influyentes dentro de un conjunto o muestra de datos. Por ejemplo, un ser con muchas amistades, es un ser popular y con ello, puede ejercer influencia en sus amigos. La centralidad busca, dentro de una muestra, estos nodos que pueden ejercer esta influencia en los nodos que estén conectados a él.

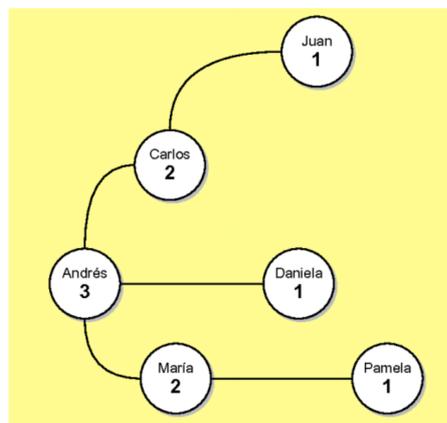


Fig. 26: Grado de popularidad.

Fuente: <http://homepage.divms.uiowa.edu/~mmonsaltv/bio/SNA.pdf>

<sup>24</sup> La centralidad en las redes sociales

[<http://revistas.ucm.es/index.php/POSO/article/viewFile/POSO0000130131A/24622>]

Freeman Linton [Consultado 10-6-2015]

<sup>25</sup> Social Network Analysis

[[https://books.google.es/books?hl=es&lr=&id=MJoIGBfYDGEC&oi=fnd&pg=PP2&dq=%22social+network+analysis%22&ots=zwAy\\_ZYkd8&sig=Nef9cS-engS2KGZaPBpulZXa7Io&redir\\_esc=y#v=onepage&q=%22social%20network%20analysis%22&f=false](https://books.google.es/books?hl=es&lr=&id=MJoIGBfYDGEC&oi=fnd&pg=PP2&dq=%22social+network+analysis%22&ots=zwAy_ZYkd8&sig=Nef9cS-engS2KGZaPBpulZXa7Io&redir_esc=y#v=onepage&q=%22social%20network%20analysis%22&f=false)] Scott J. [Consultado 12-6-2015]

## b. Uso de la interfaz con la herramienta Gephi

El origen del análisis de grafos tiene orígenes militares para examinar grupos terroristas. Luego, este tipo de análisis se exportó a la búsqueda de enfermedades y análisis de comportamientos sociales. Pero este trabajo tiene como objetivo centrarse en el análisis de links.

El software que se usará en este proyecto es Gephi. Gephi es un programa de código abierto para la exploración y visualización de todo tipo grafos. Al usar esta plataforma, se puede trabajar con infinidad de nodos conectados mediante aristas ya sean dirigidas o no. Su uso más difundido es el que abarca las redes sociales, pero al tratarse de una herramienta tan potente, la usaremos para examinar la influencia que ejercen grupos de webs con otras.

**Plugin para Facebook - Netvizz**

**Plugin para Twitter - NodeXL**

En nuestro caso, las aristas serán dirigidas. Por ejemplo si una web A apunta a una web B, del nodo A veremos salir una arista en dirección del nodo B.

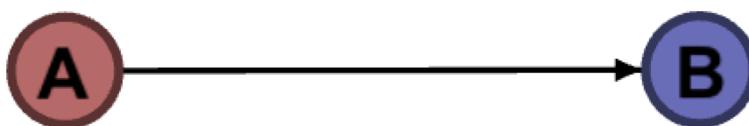


Fig. 27: Nodos interconectados. Fuente: Propia

Gephi permite la creación de grandes cantidades de nodos y aristas que ayudan a la interconexión de estos. La aplicación se divide en tres fracciones:

La vista general, donde aparecen tres columnas: la primera es particionamiento junto con otra pestaña de clasificación, la segunda es el Grafo y la tercera es la de estadísticas, filtros y contexto. En la pestaña estadística, accederemos a opciones como la “Modularidad” o la de “Centralidad del Vector Propio”. La Modularidad trata de encontrar subcomunidades en los grafos. Por otra parte, la Centralidad del Vector Propio, asigna valores según la influencia ejercida por el nodo en relación a los demás.

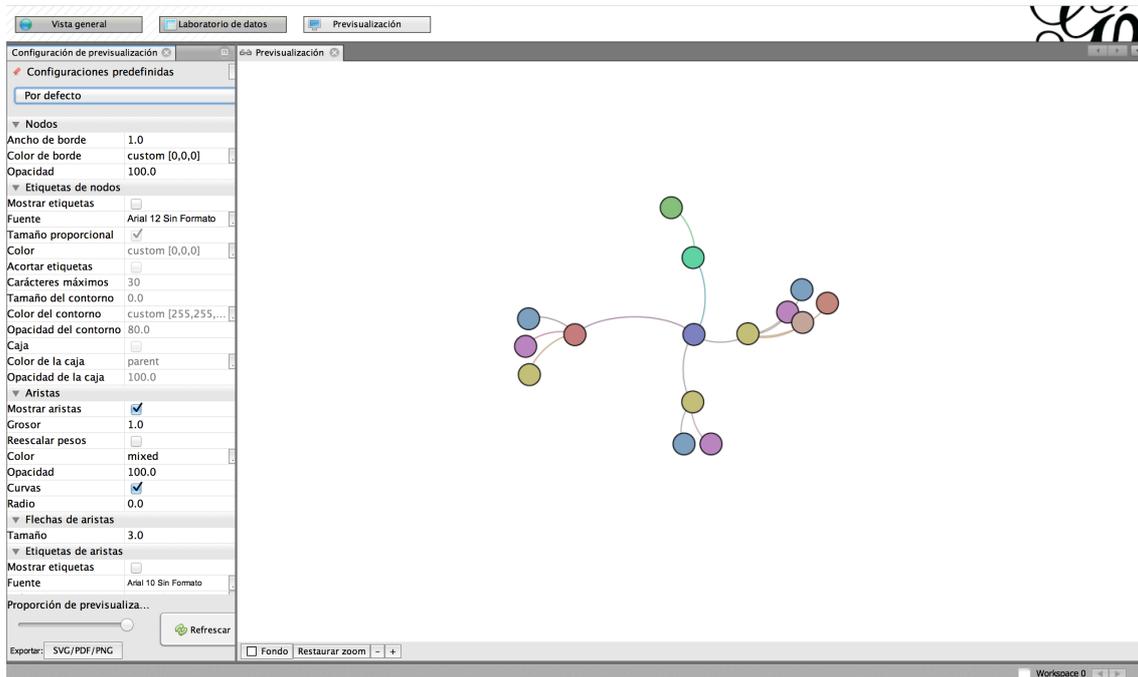


Fig. 28: Vista General de Gephi. Fuente: propia

Este mismo grafo de ejemplo, si lo modificamos con los ajustes adecuados nos puede llegar a dar mucha más información de manera visual. Si realizamos la clasificación de manera que Gephi realice un coloreado a partir de la Modularidad ejecutada anteriormente, podremos observar como se crean pequeñas comunidades. Cada color nos indica las diferentes comunidades que se han formado. A continuación, se ha realizado el mismo proceso, pero esta vez ajustando el tamaño de los nodos, según el atributo asignado por la Centralidad del Vector Propio.

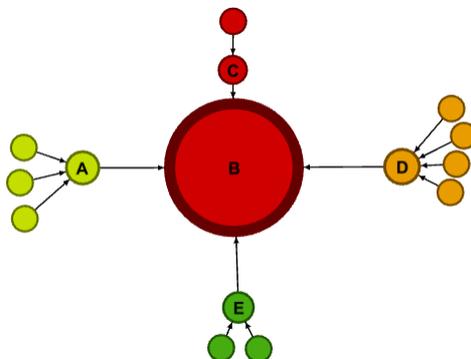


Fig. 29: Muestra aleatoria de nodos. Fuente: Propia

Por otra parte tenemos el laboratorio de datos, el cual permite administrar el grafo, añadiendo, eliminando o modificando nodos y aristas a través de sus tablas. Tales tablas se pueden importar y exportar a partir de formatos CSV: Source, Target, id,...

Origen	Destino	Tipo	Id	Label	Weight
8634	8636	Dirigida	6223		1
8637	8636	Dirigida	6224		1
8638	8636	Dirigida	6225		1
8639	8636	Dirigida	6226		1
8640	8634	Dirigida	6227		1
8641	8634	Dirigida	6228		1
8642	8634	Dirigida	6229		1
8643	8639	Dirigida	6237		1
8644	8639	Dirigida	6236		1
8645	8638	Dirigida	6233		1
8646	8638	Dirigida	6231		1
8647	8638	Dirigida	6232		1
8648	8638	Dirigida	6234		1
8649	8637	Dirigida	6230		1

Fig. 30: Laboratorio de datos. Fuente: Propia

Nodos	Id	Label	Modularity Class	Eigenvector Centrality
A	8634	A	2	0,067
B	8636	B	0	1
C	8637	C	0	0,022
D	8638	D	1	0,089
E	8639	E	3	0,044
8640	8640		2	0
8641	8641		2	0
8642	8642		2	0
8643	8643		3	0
8644	8644		3	0
8645	8645		1	0
8646	8646		1	0
8647	8647		1	0
8648	8648		1	0
8649	8649		0	0

Fig. 31: Laboratorio de datos. Fuente: Propia



Por último, Gephi ofrece la opción de poder hacer una previsualización antes de exportar el grafo a SVG/PDF. Es un punto positivo para el programa, ya que los archivos SVG son gráficos vectoriales. Esto significa que la calidad del grafo varía muy poco al ampliarlo, así que pueden ser visualizadas en terminales de alta calidad o incluso impresas sin perder resolución.

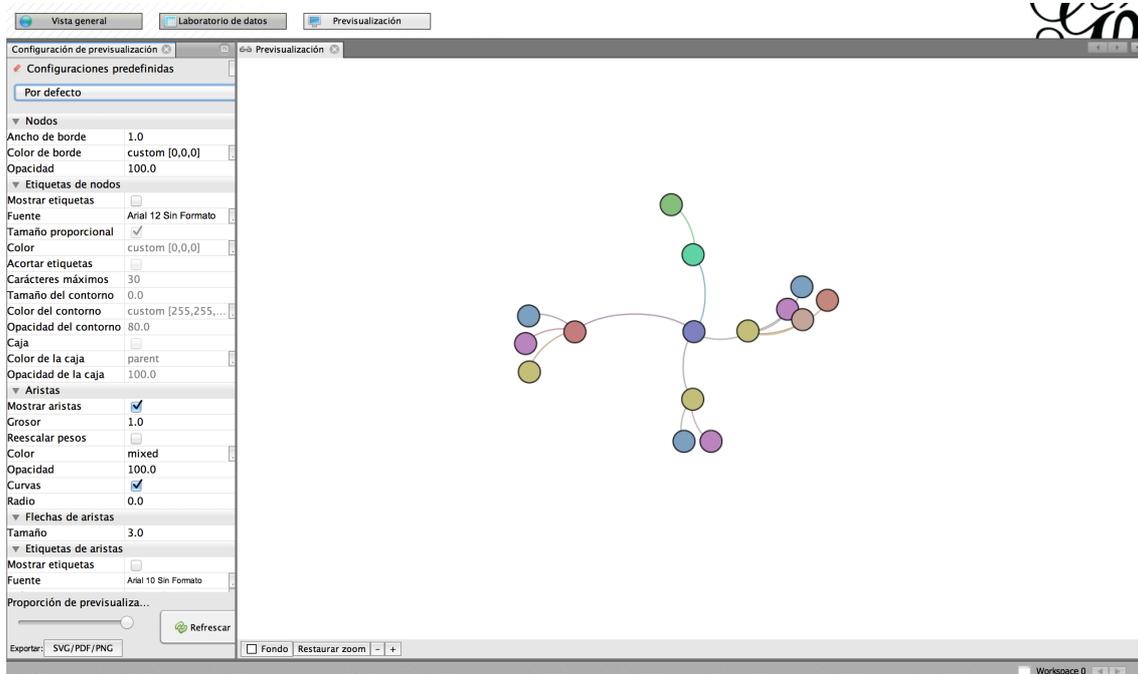


Fig. 32: Previsualización final del grafo. Fuente: Propia

En el siguiente punto, se introducirán los datos de una muestra del mundo del vino para analizarlo a fondo.

### 3.3 Patrones de enlazado entre una muestra de webs del universo vinícola

En esta segunda parte del trabajo, más práctica que la anterior, se pretende hacer un análisis de un grafo a partir de una base de datos proporcionada por los tutores. Este listado, en formato .xlsx, es un archivo con tres libros divididos en enlaces de bodegas, enlaces de bloggers y enlaces de tiendas online. La idea es introducir estos datos en la herramienta Majestic y poder exportar los datos completos de todos los enlaces. Para la adaptación de Majestic se ha adaptado el formato de los listados de la siguiente manera:

	A	B	C
1	<b>Nombre Bodega</b>	<b>URL WEB</b>	
2	Abadía Retuerta	<a href="http://www.abadia-retuerta.com/">http://www.abadia-retuerta.com/</a>	
3	Bodegas Monje	<a href="http://www.bodegasmonje.com/">http://www.bodegasmonje.com/</a>	
4	Edra	<a href="http://www.bodega-edra.com/">http://www.bodega-edra.com/</a>	
5	Pago del vicario	<a href="http://www.pagodelvicario.com/web/guest/bodega">http://www.pagodelvicario.com/web/guest/bodega</a>	
6	Freixenet	<a href="http://www.freixenet.es/">http://www.freixenet.es/</a>	
7	Cepa 21	<a href="http://www.cepa21.com/">http://www.cepa21.com/</a>	
8	Mas perinet	<a href="http://www.masperinet.com/">http://www.masperinet.com/</a>	
9	Vall Llach	<a href="http://www.vallllach.com/">http://www.vallllach.com/</a>	
10	Art Cava	<a href="http://www.artcava.com/">http://www.artcava.com/</a>	
11	Marqués de Riscal	<a href="http://www.marquesderiscal.com/">http://www.marquesderiscal.com/</a>	
12	Ysios	<a href="http://www.ysios.com/">http://www.ysios.com/</a>	
13	Viña Tondonia	<a href="http://www.tondonia.com/">http://www.tondonia.com/</a>	
14	Vicente Gandía	<a href="http://www.vicentegandia.es/">http://www.vicentegandia.es/</a>	
15	Viña Real	<a href="http://www.cvne.com/web/bodegas.php?bod=2">http://www.cvne.com/web/bodegas.php?bod=2</a>	
16	Bodegas Rioja Santiago	<a href="http://www.bodegasriojasantiago.com/">http://www.bodegasriojasantiago.com/</a>	
17	Viñas del Vero	<a href="http://www.vinasdelvero.es/">http://www.vinasdelvero.es/</a>	
18	Bodegas Habla	<a href="http://www.vinoshabla.com/">vinoshabla.com/</a>	
19	Viña Mayor	<a href="http://www.vina-mayor.es/">http://www.vina-mayor.es/</a>	
20	Bodegas Protos	<a href="http://www.bodegasprotos.com/es/">http://www.bodegasprotos.com/es/</a>	
21	Viña Sastre	<a href="http://www.vinasastre.com/sp/index.htm">http://www.vinasastre.com/sp/index.htm</a>	
22	Grupo Pesquera	<a href="http://www.grupopesquera.com/index.htm">http://www.grupopesquera.com/index.htm</a>	
23	Viña Pedrosa	<a href="http://www.perezpascuas.com/">http://www.perezpascuas.com/</a>	
24	Bodegas Muga	<a href="http://www.bodegasmuga.com/index.html">http://www.bodegasmuga.com/index.html</a>	
25	Bodegas Prado Rey	<a href="http://www.pradorey.es/home.php">pradorey.es/home.php</a>	
26	Bodegas Mauro	<a href="http://www.bodegasmauro.com/">http://www.bodegasmauro.com/</a>	
27	Bodegas Lan	<a href="http://www.bodegaslan.com/">http://www.bodegaslan.com/</a>	
28	Bodega Arzuaga-Navarro	<a href="http://www.hotelarzuaga.com/index.php/es/bodega">http://www.hotelarzuaga.com/index.php/es/bodega</a>	
29	Pago de Carraovejas	<a href="http://www.pagodecarraovejas.com/">http://www.pagodecarraovejas.com/</a>	
30	Viña Alcorta	<a href="http://www.alcortavino.com/">http://www.alcortavino.com/</a>	
31	Bodega Matarromera	<a href="http://www.matarromera.es">http://www.matarromera.es</a>	

Fig. 33: Muestra de datos del mundo del vino. Fuente: Propia

Se ha transformado el archivo proporcionado por los tutores a un formato de tipo CSV

PAIS	Nombre	URL
EspaĐa	Punto Vino	<a href="http://www.puntovino.es/">http://www.puntovino.es/</a>
EspaĐa	Uvinum	<a href="http://blog.uvinum.es/mejor-blog-vinos-espana-1175007">http://blog.uvinum.es/mejor-blog-vinos-espana-1175007</a>
EspaĐa	Blog de Vinos	<a href="http://www.blogdevinos.net/">http://www.blogdevinos.net/</a>
EspaĐa	Vinos-blog	<a href="http://www.vinos-blog.com/">http://www.vinos-blog.com/</a>
EspaĐa	Saber de vinos	<a href="http://www.saberdevino.com/">http://www.saberdevino.com/</a>
EspaĐa	ELVI.net	<a href="http://www.elvi.net/">http://www.elvi.net/</a>
EspaĐa	De vinis.org	<a href="http://www.devinis.org/">http://www.devinis.org/</a>
EspaĐa	Terroaristas	<a href="http://www.terroaristas.com/">http://www.terroaristas.com/</a>
EspaĐa	Cupatges	<a href="http://www.cupatges.cat/">http://www.cupatges.cat/</a>
EspaĐa	Envinats	<a href="http://blogs.elpunt.cat/envinats/">http://blogs.elpunt.cat/envinats/</a>
EspaĐa	Akata vino	<a href="http://www.akatavino.com/">http://www.akatavino.com/</a>
EspaĐa	Weirdo	<a href="http://www.weirdo.es/">http://www.weirdo.es/</a>
EspaĐa	De uvas a vinos	<a href="http://deuvasavinos.blogspot.com.es/">http://deuvasavinos.blogspot.com.es/</a>
EspaĐa	Abre un vinito	<a href="http://abreunvinito.blogspot.com.es/">http://abreunvinito.blogspot.com.es/</a>
EspaĐa	Experto en vinos	<a href="http://expertoenvinos.com/blog/">http://expertoenvinos.com/blog/</a>
EspaĐa	Cristina Alcalá	<a href="http://www.cristinaalcala.com/">http://www.cristinaalcala.com/</a>
EspaĐa	Donde el viento nos lleve	<a href="http://dondeelvientonoslleve.com/">http://dondeelvientonoslleve.com/</a>
EspaĐa	La casa de Antociano	<a href="http://www.antociano.net/blog/">http://www.antociano.net/blog/</a>
EspaĐa	Bodega CauzÑn	<a href="http://bodegacauzon.blogspot.com.es/">http://bodegacauzon.blogspot.com.es/</a>
EspaĐa	Adictos a la lujuria	<a href="http://www.adictosalalujuria.com/">http://www.adictosalalujuria.com/</a>
EspaĐa	Va de bacus	<a href="http://www.vadebacus.com/">http://www.vadebacus.com/</a>
EspaĐa	Baba O'Wines	<a href="http://iglegorburu.wordpress.com/">http://iglegorburu.wordpress.com/</a>
EspaĐa	Descorche	<a href="http://blogs.20minutos.es/descorchevinos/">http://blogs.20minutos.es/descorchevinos/</a>
EspaĐa	Gourmet del prado	<a href="http://gourmetdelprado.blogspot.com.es/">http://gourmetdelprado.blogspot.com.es/</a>
EspaĐa	Sibaritastur	<a href="http://sibaritastur.blogspot.com.es/">http://sibaritastur.blogspot.com.es/</a>
EspaĐa	Polakia	<a href="http://www.polakia.com/">http://www.polakia.com/</a>
EspaĐa	Roco & Wines	<a href="http://www.rocowines.net/">http://www.rocowines.net/</a>
EspaĐa	Sobre vino	<a href="http://www.sobrevino.com/">http://www.sobrevino.com/</a>
EspaĐa	Gota a gota	<a href="http://gotaagotablog.com/">http://gotaagotablog.com/</a>
EspaĐa	Blod de vinivicultura de Nacho Bueno	<a href="http://nachobueno.wordpress.com/">http://nachobueno.wordpress.com/</a>
EspaĐa	Parrasounds	<a href="http://www.parrasounds.com/">http://www.parrasounds.com/</a>
EspaĐa	Aprende a catar vino	<a href="http://aprendeacatarvino.wordpress.com">http://aprendeacatarvino.wordpress.com</a>
EspaĐa	Las webs del vino	<a href="http://websvino.blogspot.com.es/">http://websvino.blogspot.com.es/</a>
EspaĐa	Clásicos de Vanguardia	<a href="http://blogs.larioja.com/vinosclasicos/">http://blogs.larioja.com/vinosclasicos/</a>
EspaĐa	El mundo de Rafa Ibarra	<a href="http://www.elmundoderafaibarra.com/">http://www.elmundoderafaibarra.com/</a>
EspaĐa	La trastienda def	<a href="http://traslavitualla.blogspot.com.es/">http://traslavitualla.blogspot.com.es/</a>
EspaĐa	Vinos Kupel	<a href="http://vinoskupel.blogspot.com.es/">http://vinoskupel.blogspot.com.es/</a>
EspaĐa	Oenoblog	<a href="http://www.oenoblog.info">http://www.oenoblog.info</a>
EspaĐa	Vins i roses	<a href="http://www.vinsiroses.com/">http://www.vinsiroses.com/</a>
EspaĐa	Tu propio estilo de vino	<a href="http://www.tupropioestilodevino.com/">http://www.tupropioestilodevino.com/</a>

Fig. 34: Muestra de datos del mundo del vino en CSV. Fuente: Propia

De esta manera, Majestic acepta hasta 10.000 links en su sección de **bulk backlinks**<sup>26</sup>, como se ha comentado antes. En este caso se han introducido los 2.609 enlaces para su análisis. Y este ha sido el resultado después de dos minutos de trabajo por parte del servidor:

<sup>26</sup> Bulk Backlinks: comprobación masiva de enlaces. Herramienta proporcionada por Majestic.

Id	Status	External Bac	External Bac	External Bac	Referring Do	Referring Do	Referring Do	IP Addresses	Class C	Subn	Trust Flow	Citation Flow	Topical Trust
<a href="http://www.abadia-retuerta.com/">http://www.abadia-retuerta.com/</a>	Found	11109	3	0	369	1	0	324	294	46		32	Business
<a href="http://www.bodegasmonje.com/">http://www.bodegasmonje.com/</a>	Found	4011	0	0	333	0	0	259	238	22		30	Computers/I
<a href="http://www.bodega-cdra.com/">http://www.bodega-cdra.com/</a>	Found	443	0	0	76	0	0	64	57	24		25	Regional/Eur
<a href="http://www.sagodelvicario.com/web/guest/bodega/">http://www.sagodelvicario.com/web/guest/bodega/</a>	Found	455	0	0	22	0	0	20	18	17		18	Recreation/T
<a href="http://www.freixenet.es/">http://www.freixenet.es/</a>	Found	85965	14	0	1063	4	0	919	761	54		51	Business/Foc
<a href="http://www.cepa21.com/">http://www.cepa21.com/</a>	Found	1642	0	0	249	0	0	182	160	29		32	Business
<a href="http://www.masperinet.com/">http://www.masperinet.com/</a>	Found	420	0	0	120	0	0	101	96	21		25	Recreation/F
<a href="http://www.valllach.com/">http://www.valllach.com/</a>	Found	2218	0	0	241	0	0	206	193	45		30	Business/Foc
<a href="http://www.artcava.com/">http://www.artcava.com/</a>	Found	1249	0	0	90	0	0	80	77	38		27	Business/Foc
<a href="http://www.marquesderiscal.com/">http://www.marquesderiscal.com/</a>	Found	10380	1	0	1226	1	0	1012	870	35		36	Sports/Cyclir
<a href="http://www.ysisos.com/">http://www.ysisos.com/</a>	Found	995	0	0	215	0	0	186	171	21		27	Reference/IV
<a href="http://www.tondonia.com/">http://www.tondonia.com/</a>	Found	379	3	0	64	1	0	61	57	27		24	Recreation/F
<a href="http://www.vicentegandia.es/">http://www.vicentegandia.es/</a>	Found	1055	0	0	149	0	0	129	124	39		27	Business
<a href="http://www.cvne.com/web/bodegas.php?bod=2">http://www.cvne.com/web/bodegas.php?bod=2</a>	Found	142	0	0	37	0	0	36	34	25		22	Society/Gov
<a href="http://www.bodegasriojasantiago.com/">http://www.bodegasriojasantiago.com/</a>	Found	123	0	0	58	0	0	57	53	34		20	Business
<a href="http://www.vinasdelvero.es/">http://www.vinasdelvero.es/</a>	Found	2574	27	0	381	3	0	336	303	47		32	Regional/Eur
<a href="http://vinoshabla.com/">http://vinoshabla.com/</a>	Found	39	0	0	12	0	0	12	12	7		6	Business
<a href="http://www.vina-mayor.es/">http://www.vina-mayor.es/</a>	Found	33194	0	0	73	0	0	53	50	18		27	Business
<a href="http://www.bodegasprotos.com/es/">http://www.bodegasprotos.com/es/</a>	Found	1907	5	0	256	1	0	228	211	41		32	Regional/Eur
<a href="http://www.vinassastre.com/sg/index.htm">http://www.vinassastre.com/sg/index.htm</a>	Found	19	0	0	5	0	0	5	5	4		13	Recreation/I
<a href="http://www.gruposquera.com/index.htm">http://www.gruposquera.com/index.htm</a>	MayExist												
<a href="http://www.perezpascuas.com/">http://www.perezpascuas.com/</a>	Found	34622	0	0	232	0	0	166	158	43		31	Regional/Eur
<a href="http://www.bodegasmuga.com/index.html">http://www.bodegasmuga.com/index.html</a>	Found	42	0	0	10	0	0	10	10	6		16	Home/Cooki
<a href="http://pradorey.es/home.php">http://pradorey.es/home.php</a>	Found	303	0	0	9	0	0	9	9	14		16	Arts/Televisi
<a href="http://www.bodegasmauro.com/">http://www.bodegasmauro.com/</a>	Found	35090	0	0	223	0	0	190	177	21		27	Regional/Eur
<a href="http://www.bodegaslan.com/">http://www.bodegaslan.com/</a>	Found	53784	0	0	351	0	0	293	271	28		31	Shopping/Fo
<a href="http://www.hotelarzuaga.com/index.php/es/bodega">http://www.hotelarzuaga.com/index.php/es/bodega</a>	Found	38	0	0	15	0	0	11	11	8		16	Science/Tech
<a href="http://www.pagodecarravejas.com/">http://www.pagodecarravejas.com/</a>	Found	5262	0	0	185	0	0	153	139	28		29	Business/Foc
<a href="http://www.alcortavino.com/">http://www.alcortavino.com/</a>	Found	1223	0	0	149	0	0	117	106	20		26	Regional/Eur
<a href="http://www.matarromera.es/">http://www.matarromera.es/</a>	Found	9936	0	0	284	0	0	251	231	22		29	Business
<a href="http://www.bodegasramonbilbao.es/">http://www.bodegasramonbilbao.es/</a>	Found	13012	0	0	435	0	0	360	313	49		33	Business
<a href="http://www.viE_aherminia.es">http://www.viE_aherminia.es</a>	MayExist												
<a href="http://www.vinavilano.com/">http://www.vinavilano.com/</a>	Found	851	0	0	92	0	0	84	80	23		25	Recreation/F
<a href="http://www.bodegasjuangil.com/">http://www.bodegasjuangil.com/</a>	Found	2124	0	0	183	0	0	164	154	18		27	Recreation/F
<a href="http://www.bodegasillon.com/">http://www.bodegasillon.com/</a>	Found	3706	0	0	192	0	0	163	158	20		28	Recreation/F
<a href="http://www.beronia.es/home.php?idioma=1">http://www.beronia.es/home.php?idioma=1</a>	Found	15	0	0	4	0	0	4	4	8		14	News/News
<a href="http://www.dinastiavivanco.com/inicio.asp">http://www.dinastiavivanco.com/inicio.asp</a>	Found	4917	0	0	218	0	0	201	180	23		29	Regional/Eur
<a href="http://www.pierola.com/">http://www.pierola.com/</a>	Found	977	0	0	95	0	0	88	84	19		24	Regional/Eur
<a href="http://www.grupoartevino.com/">http://www.grupoartevino.com/</a>	Found	25886	0	0	236	0	0	161	146	24		35	Business/Foc
<a href="http://www.bodegashcalvente.com/home.html">http://www.bodegashcalvente.com/home.html</a>	Found	3862	0	0	9	0	0	8	8	16		27	Business
<a href="http://www.eguren.com/inicio?acc=home&amp;idc=1">http://www.eguren.com/inicio?acc=home&amp;idc=1</a>	Found	37	0	0	12	0	0	12	12	1		16	Recreation/F
<a href="http://www.4kilos.com/">http://www.4kilos.com/</a>	Found	4615	1	0	234	1	0	204	187	21		29	Recreation/T
<a href="http://www.familiamartinezbuja.com/antigua/?idc=27">http://www.familiamartinezbuja.com/antigua/?idc=27</a>	Found	416	0	0	41	0	0	32	31	21		22	Society/Peop

Fig. 35: Muestra de datos del mundo del vino en CSV y con datos. Fuente: Propia

Como se muestra en la imagen, cada uno de los enlaces viene acompañado de mucha más información de la que tenía antes de ser insertado en Majestic. Una vez se posee este archivo, se importará desde Gephi como si de nodos de un grafo se tratasen.

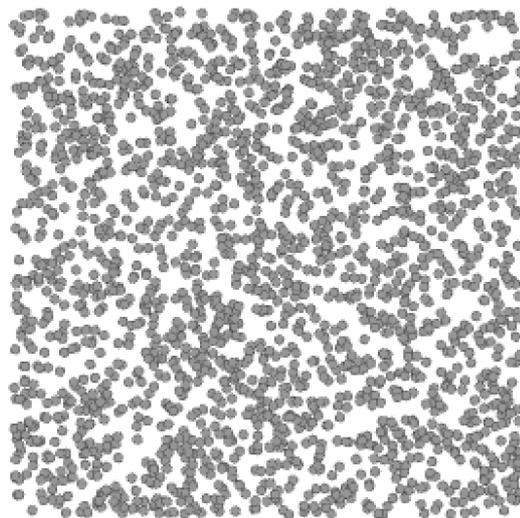


Fig. 36: Grafo con los datos introducidos. Fuente: Propia



Se comprueba que todos los nodos hayan sido introducidos correctamente. La distribución, los colores y los tamaños serán configurados una vez hayamos introducido las aristas.

Para la importación de las aristas, se ha procedido a descubrir los backlinks de más de 200 enlaces. Gracias a la herramienta Majestic, se ha podido importar entre 300 y 1000 backlinks de cada enlace analizado. Y el resultado han sido más de 200 documentos de en excel con el siguiente formato:

A	B	C	D	E	F	G	H	I	J	K
SourceURL	AnchorText	SourceTrustFlow	SourceCitationFlow	Domain	DomainTrustFlow	DomainCitationFlow	FirstIndexedDate	LinkType	LinkSubType	TargetURL
http://www.chadds-winetasting.com		41	28	chaddsford.com	43	28	#####	TextLink	TextLink_Normal	http://www.winetasting.com/
http://www.winesa-winetasting.com		29	28	winesandvines.co	48	37	#####	TextLink	TextLink_Normal	http://www.winetasting.com/
http://www.noircity-winetasting.com		33	22	noircity.com	35	25	#####	TextLink	TextLink_Normal	http://www.winetasting.com/
http://www.1800lowersinc.com/brands.asp		28	25	1800lowersinc.c	23	26	#####	ImageLink	ImageLink_Normal	http://www.winetasting.com/
http://www.internetwineguid.com		30	20	internetwineguid	38	26	#####	TextLink	TextLink_Normal	http://www.winetasting.com/
http://www.geerwade.com/		24	25	geerwade.com	15	13	#####	Redirect	Redirect_Permanent_Chain	http://www.winetasting.com/

Fig. 37: Archivo excel extraido de Majestic. Fuente: Propia

Se ha programado una macro especial para Excel con el fin de juntar los 200 archivos. Este código permite coger los archivos de una carpeta y ponerlos uno detrás de otro en una sola hoja para que, de manera eficiente, Gephi no tenga que importar 200 archivos, sino uno solo.

```

1 Sub simpleXlsMerger()
2 Dim bookList As Workbook
3 Dim mergeObj As Object, dirObj As Object, filesObj As Object, everyObj As Object
4 Application.ScreenUpdating = False
5 Set mergeObj = CreateObject("Scripting.FileSystemObject")
6
7 'change folder path of excel files here
8 Set dirObj = mergeObj.GetFolder("C:\Users\cricappo\Desktop\unir\prueba")
9 Set filesObj = dirObj.Files
10 For Each everyObj In filesObj
11 Set bookList = Workbooks.Open(everyObj)
12
13 'change "A2" with cell reference of start point for every files here
14 'for example "B3:IV" to merge all files start from columns B and rows 3
15 'If you're files using more than IV column, change it to the latest column
16 'Also change "A" column on "A65536" to the same column as start point
17 Range("A1:IV" & Range("A65536").End(xlUp).Row).Copy
18 ThisWorkbook.Worksheets(1).Activate
19
20 'Do not change the following column. It's not the same column as above
21 Range("A65536").End(xlUp).Offset(1, 0).PasteSpecial
22 Application.CutCopyMode = False
23 bookList.Close
24 Next
25 End Sub

```

Fig. 38: Macro en excel para juntar los archivos. Fuente: Propia

Gephi permite la importación de documentos de tipo CSV que no tengan columnas repetidas y que tengan claramente, las columnas que nombren a los nodos que hacen referencia al inicio y fin de las aristas. Estas columnas deben llamarse Source para la columna de los nodos origen y Target, para la columna de los nodos objetivo:

A	B	C	D	E
Source	AnchorText	SourceTrustFlow	SourceCitationFlow	Target
http://congresodietamediterranea.com/patrocinadores/		36	70	http://www.freixenet.es/
http://www.freixenet.com/index.asp	spain	59	42	http://www.freixenet.es/
http://cenoia.com/	toggle navigation	47	33	http://www.freixenet.es/
http://www.tv3.cat/marato/empreses	freixenet	43	33	http://www.freixenet.es/

Fig. 39: Archivo excel modificado con columnas Origen y Destino. Fuente: Propia

Con el archivo en formato CSV, se han importado desde Gephi las aristas.

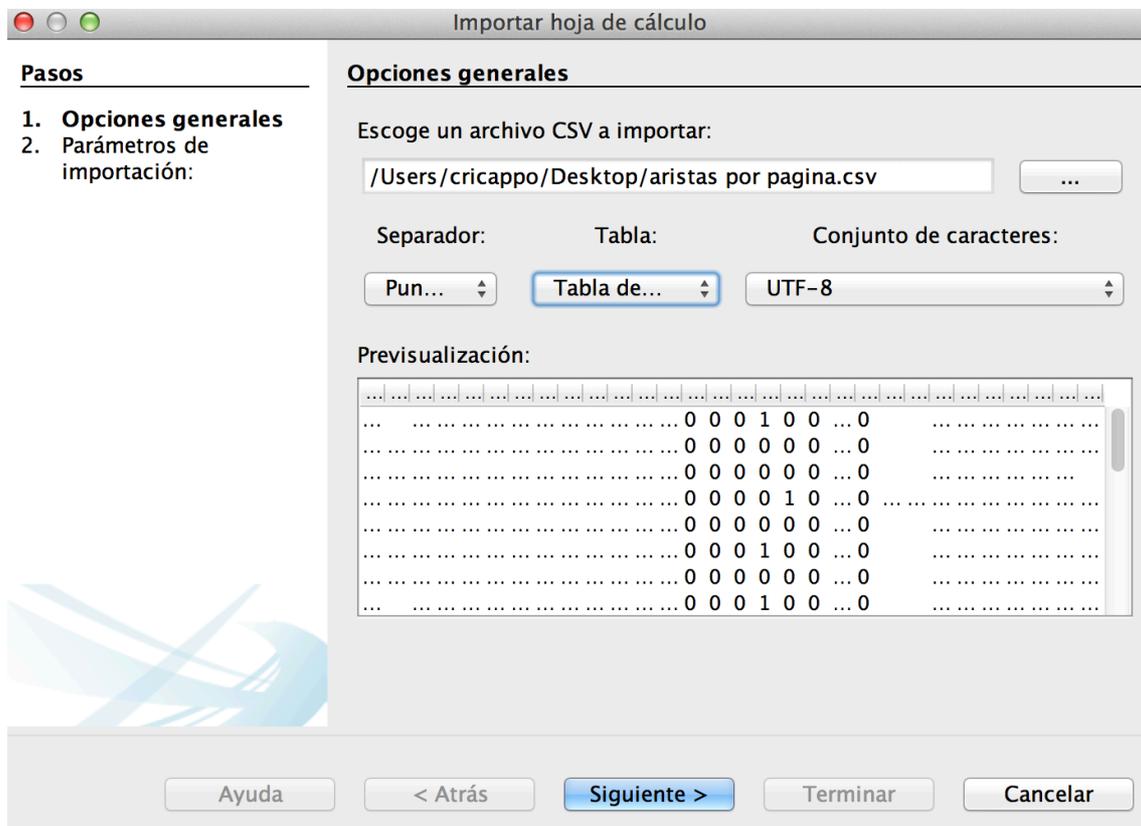


Fig. 40: Proceso de importación de las aristas. Fuente: Propia

Una vez se ha importado el archivo de aristas, ya se tiene el grafo preparado para ser procesado. Como se ha comentado anteriormente, el grafo no está aun ordenado. No se le ha aplicado ninguna distribución ni color ni tamaño.

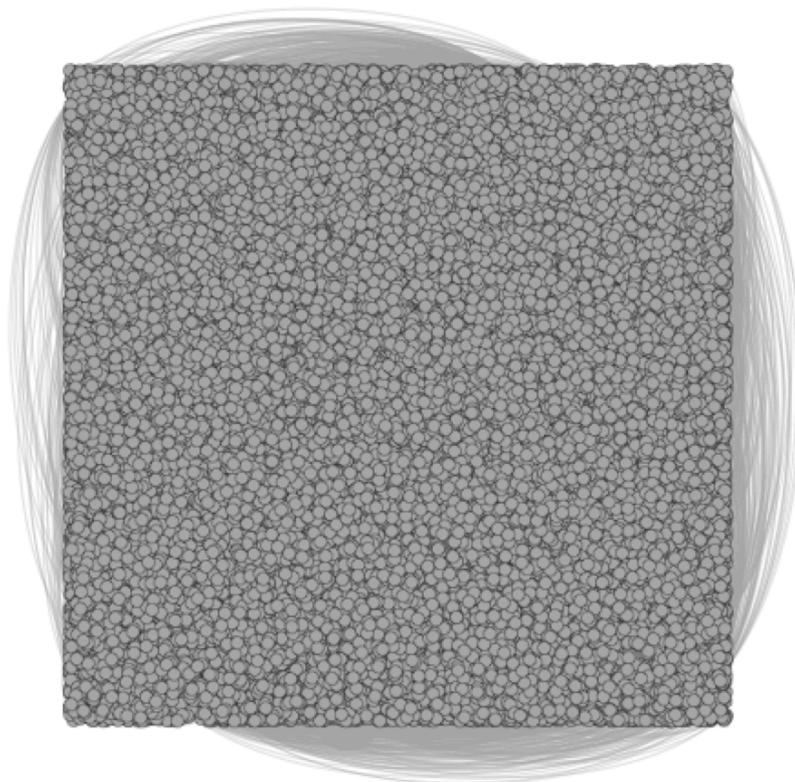


Fig. 41: Grafo con los datos de aristas introducidos. Fuente: Propia

En la sección de estadísticas se han ejecutado las métricas de modularidad, para la búsqueda de comunidades y subcomunidades y la de Centralidad del Vector Propio, para la búsqueda de los nodos más influyentes. Para la distribución se ha elegido la de “Fruchterman Reingold”, el cual permite que unos nodos tiren de otros hasta dejar fuera los menos influyentes y dejar dentro los más comunicados. Este proceso requiere mucho tiempo. En este caso, se ha esperado hasta 10 minutos para la reorganización de los 43.221 nodos y las 46.019 aristas.



## 4. Análisis de resultados

Para empezar a ver datos interesantes, se ha exportado la imagen a PDF, cosa que ha permitido observar ciertos comportamientos por parte de las webs relacionadas con el mundo de los vinos.

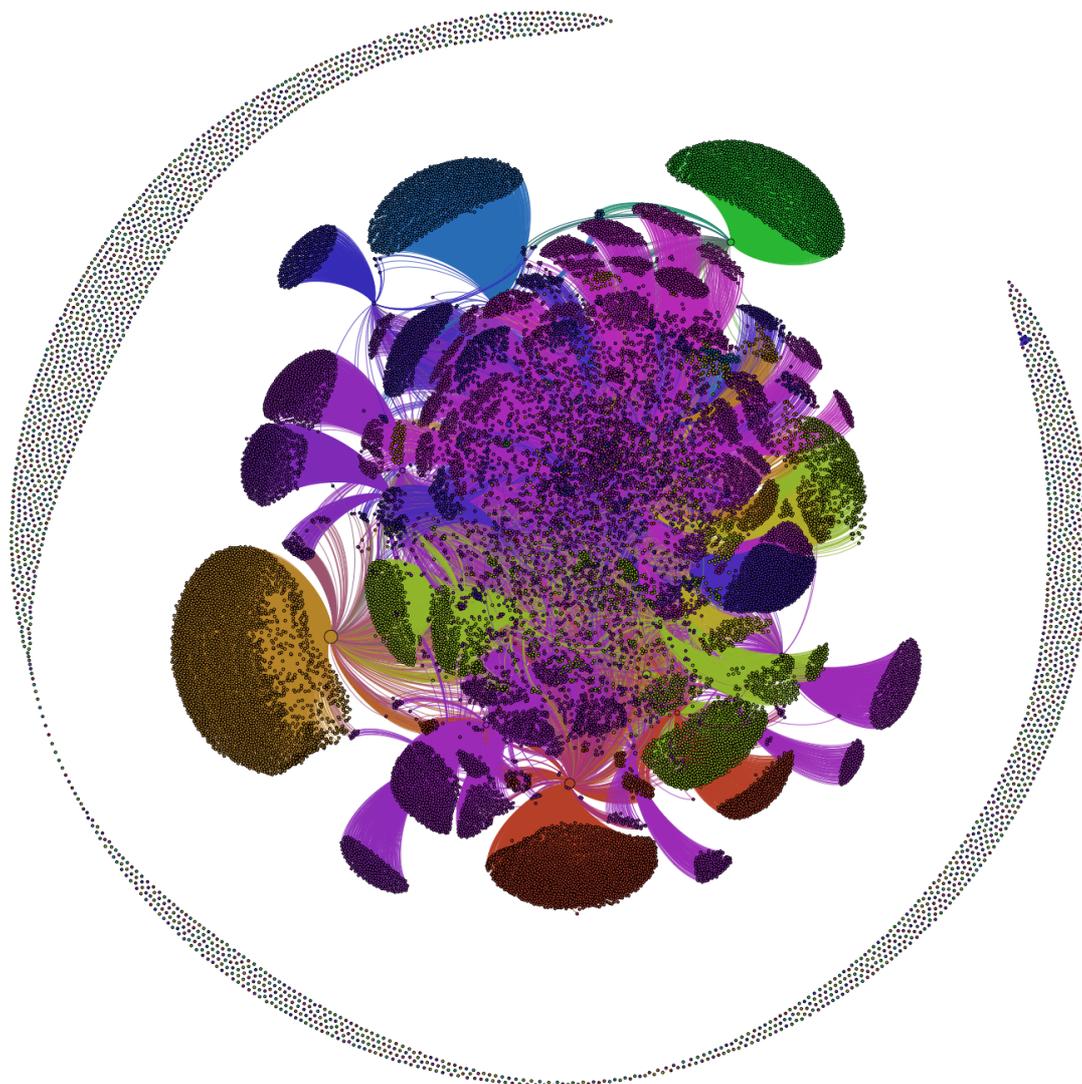


Fig. 42: Grafo con los datos. Color: Modularidad. Tamaño: Centralidad del Vector Propio.

Fuente: Propia

Gracias al coloreado por comunidades (Modularity), a simple vista se pueden ver grupos y subgrupos de webs que se enlazan entre ellas y que tiene nodos “padre” muy influyentes. Podemos ver sus pertenencias en el apartado de particionado según su Topical Trust Flow principal (o, el más alto)

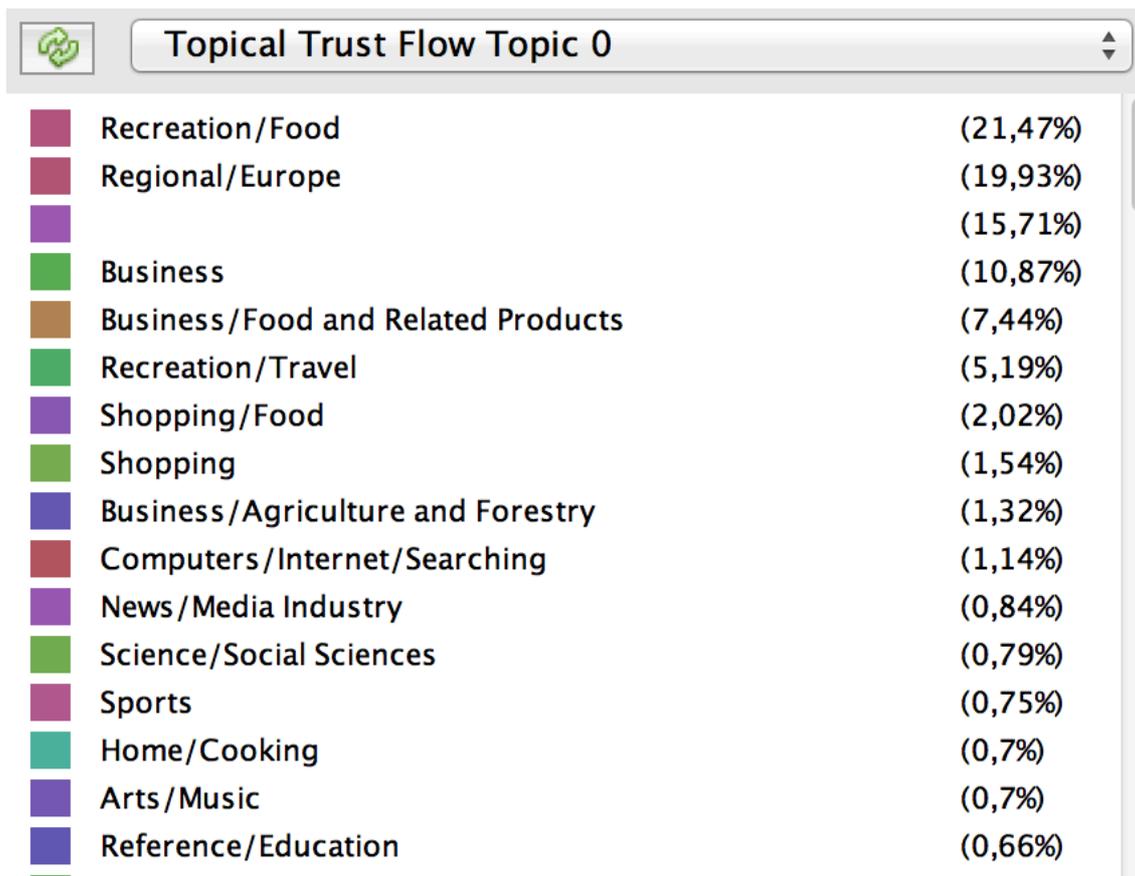


Fig. 43: Particionamiento a partir del atributo “Topical Trust Flow Topic 0”.

Fuente: Propia

Aquí se puede observar cómo el 21,47% de las webs son de temática Recreation/Food reconocida por Majestic. Este tipo de vistas permiten ver rápidamente porcentajes de atributos.

Sorprende el hecho que las páginas relacionadas con el mundo del vino se relacionan muy poco entre ellas. Se forman como islas, todo son competidores. Cada web intenta realizar su hazaña de manera singular. Las pocas veces se enlazan, se observa que las temáticas de los enlaces entrantes son:

- Sports: -> Blogs
- Society / People -> Blogs
- Shopping / Food -> Bodegas y Blogs
- Regional Europe -> Bodegas y Blogs
- Business -> Tiendas
- Business / Agriculture and Forestry -> Bodegas y Tiendas

Es decir, la herramienta Majestic, puede detectar en cada página que analizamos, de que temáticas son los enlaces que le llegan. Y esto es interesante, ya que en una web de vinos, puede llegar a detectar qué páginas son de venta (Shopping o Business) y cuales son de simplemente de opinión (Society o People)

Si se hace un filtrado con Centralidad del Vector Propio, es decir, si se quiere recoger una muestra de los nodos más influyentes, obtendremos un número reducido de nodos. Los más importantes y los que predominan sobre los demás.

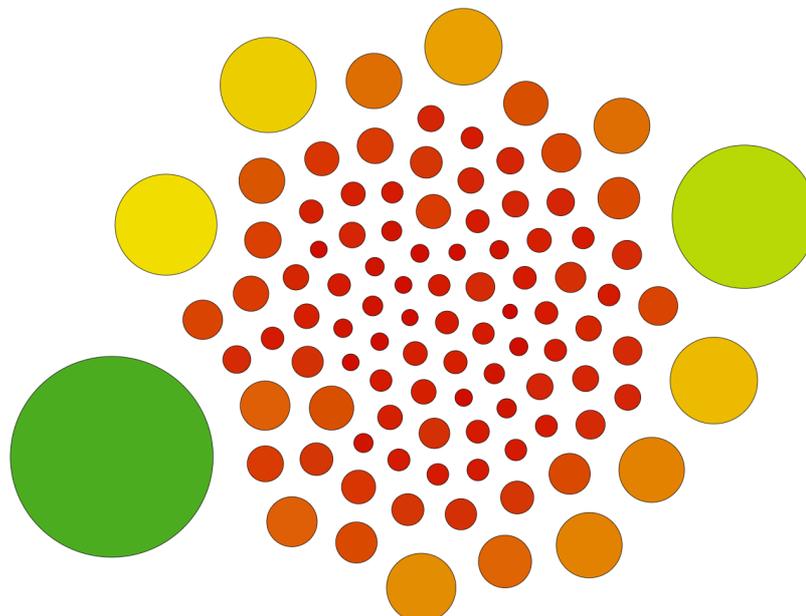


Fig. 44: Filtrado del grafo a partir de la Centralidad del Vector Propio. Fuente: Propia

En el laboratorio de datos se puede observar cual ha sido el resultado. Se consigue pues tener un listado con los 80-90 webs más influyentes en lo que a el mundo del vino se refiere. Se han ordenado según su Trust Flow.

Nodes	Trust Flow	Citation Flow	External Backlinks
<a href="https://www.erobertparker.com/entrance.aspx">https://www.erobertparker.com/entrance.aspx</a>	69	55	41661
<a href="http://www.lvmh.com/">http://www.lvmh.com/</a>	61	47	68128
<a href="https://winelibrary.com/">https://winelibrary.com/</a>	61	47	4936
<a href="http://www.puig.com/">http://www.puig.com/</a>	60	45	140622
<a href="http://www.larioja.com/">http://www.larioja.com/</a>	58	54	1393453
<a href="http://www.pernod-ricard-espana.com/">http://www.pernod-ricard-espana.com/</a>	56	39	7696
<a href="http://www.paginasamarillas.es/">http://www.paginasamarillas.es/</a>	56	50	4359401
<a href="http://www.bbr.com/">http://www.bbr.com/</a>	56	43	104235
<a href="http://www.freixenet.es/">http://www.freixenet.es/</a>	54	51	85965
<a href="http://www.klwines.com/">http://www.klwines.com/</a>	54	44	31255
<a href="http://www.wines.com/">http://www.wines.com/</a>	52	37	100184
<a href="http://www.majestic.co.uk/">http://www.majestic.co.uk/</a>	52	38	132410
<a href="http://www.gonzalezbyass.com/">http://www.gonzalezbyass.com/</a>	51	40	6242
<a href="http://www.wine-pages.com/">http://www.wine-pages.com/</a>	51	40	74115
<a href="http://www.thewinedoctor.com/">http://www.thewinedoctor.com/</a>	51	41	86420
<a href="http://www.laithwaites.co.uk/">http://www.laithwaites.co.uk/</a>	51	38	135611
<a href="http://www.terrasgauda.com/">http://www.terrasgauda.com/</a>	50	36	65954
<a href="http://www.bodegasmuga.com/">http://www.bodegasmuga.com/</a>	50	38	24365
<a href="http://www.bodegasramonbilbao.es/">http://www.bodegasramonbilbao.es/</a>	49	33	13012
<a href="http://www.codorniu.es/">http://www.codorniu.es/</a>	49	33	6242
<a href="http://www.riojalta.com/">http://www.riojalta.com/</a>	49	34	5751
<a href="http://www.nakedwines.com/">http://www.nakedwines.com/</a>	49	43	103812
<a href="http://www.martincodax.com/">http://www.martincodax.com/</a>	48	35	71962
<a href="http://www.vega-sicilia.com/">http://www.vega-sicilia.com/</a>	48	34	16975
<a href="http://www.doriasbaixas.com/">http://www.doriasbaixas.com/</a>	48	39	9185
<a href="http://www.gramona.com/">http://www.gramona.com/</a>	48	35	25147
<a href="http://www.lopezdeheredia.com/">http://www.lopezdeheredia.com/</a>	48	32	7536
<a href="http://www.oddbins.com/">http://www.oddbins.com/</a>	48	24	9291

Fig. 45: Laboratorio de datos una vez aplicado el filtro. Fuente: Propia

De esta manera, de una muestra que llegaba hasta los 43.000 nodos, se recogen ahora los más influyentes. Si nos quedamos con los más importantes según cada país, tenemos:

- EEUU: [erobertparker.com](https://www.erobertparker.com) y [lvmh.com](http://www.lvmh.com)
- Francia: [puig.com](http://www.puig.com)
- Gran Bretaña: [bbr.com](http://www.bbr.com)
- España: [freixenet.es](http://www.freixenet.es)

Los resultados obtenidos tienen mucho sentido: se trata de páginas web que giran entorno al universo del vino. Si analizamos el link más valioso de todos en lo que a Trust Flow se refiere, nos encontramos con la web del crítico de vino Robert Parker. Una web con mucha importancia y influencia. Según “<http://elmundovino.elmundo.es/>”, se trata del “hombre más poderoso, alabado y temido en el panorama vinícola”.

En el caso de España, la web más influyente es [freixenet.es](http://freixenet.es), ya que aunque en la actualidad se dediquen a hacer cava, en sus principios era una productora vinícola. Burbujas de oro, natación, anuncios con personajes de ámbito internacional, etc. Es evidente que Freixenet es un grupo que habrá invertido una gran cantidad de dinero en marketing, con lo cual la influencia que desprende es alta.

Si vamos más allá de lo que a posicionamiento SEO se refiere, se observa que en mundo del vino, no se hablan entre sí. Como se ha comentado antes, no hay comunicación ni enlaces entre estos portales. Al parecer, por tema de competencias, no parece que quieran enlazarse unas con otras.

Al contrario que aquí, en el mundo de la tecnología, una gran parte de las webs se enlazan unas con otras para determinar de donde se ha sacado cierta información, gracias a la sección “Fuente”. Esto permite más movimiento entre webs de la misma temática.

Un posicionador SEO que vaya a hacer un estudio en el mundo del vino, puede ver a través de estos grafos las conexiones que existen entre unos nodos y otros.

La herramienta Majestic ha ayudado mucho gracias a su instrumento de trabajo del Topical Trust Flow. Ya que así, una web puede saber con que “barrios” puede comunicarse sin problemas. Sería extraño, por ejemplo, que una web de vinos tuviese backlinks con procedencia de otros sectores que nada tienen que ver con el vino.

Haciendo comparaciones entre el Fresh Index y el Historic Index, podríamos llegar incluso a tener una noción del grado de conversación que tienen estas webs. Es decir, no es lo mismo que una web enlace solo una vez a otra cada 10 años, que mensualmente.

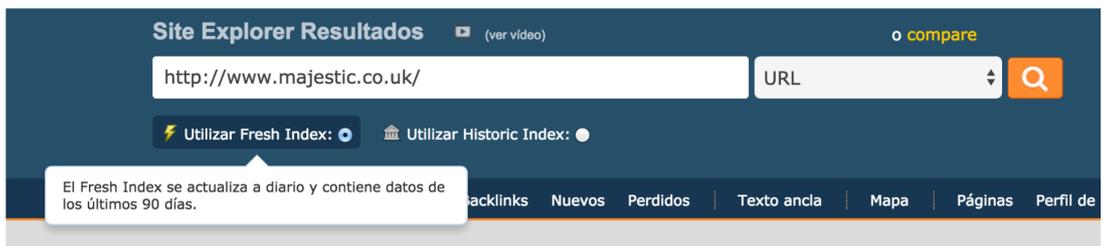
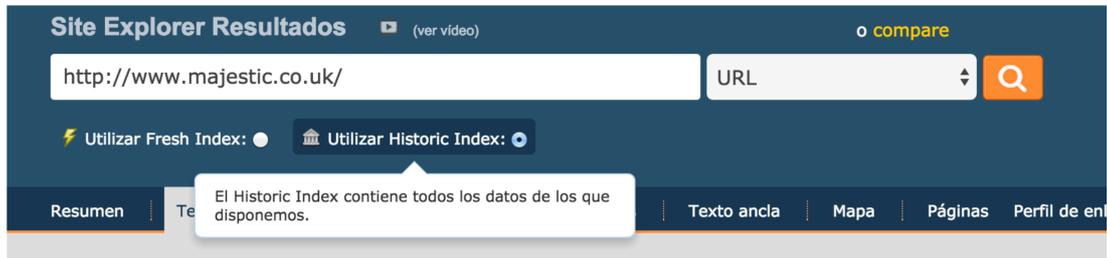


Fig. 46: Site Explorer Fresh Index – Historic Index. Fuente: Propia

## 5. Recomendaciones a nivel empresarial

En una encuesta, se preguntó a los usuarios que opinión tenían sobre los resultados que las búsqueda devolvían:

**¿Está usted de acuerdo con la siguiente afirmación: “La empresa listada en la primera posición de los buscadores es la compañía líder en su sector”?**

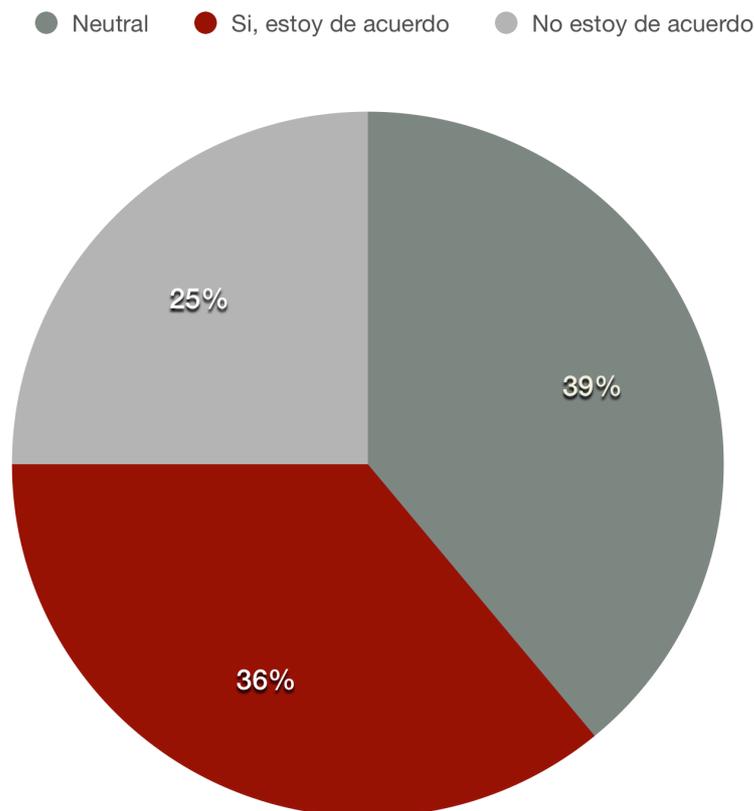


Fig. 47: Encuesta

Fuente: Maciá-Domene, F. & Gosende-Grela J. (2006). *Posicionamiento en Buscadores*. Madrid: Grupo Anaya

Es decir, si una empresa no está entre las primeras treinta posiciones para una búsqueda, las probabilidades de contactar con los clientes potenciales es improbable. Estar entre las primeras posiciones puede representar una oportunidad de negocio para una empresa. En ciertas ocasiones, la clave suele residir en el estudio del nicho de mercado de la empresa y en un adecuado posicionamiento de este.

Sin entrar en internet, uno puede preguntarse cual es el número de competidores reales sobre un nicho de mercado. Con ello, se debe buscar cual es el valor añadido que tiene nuestra empresa que no tenga la del competidor. Probablemente se llegará a un número reducido de empresas que trabajen así, puede que menos de diez. Si esto es así en el mundo real, también puede ser igual en Internet.

Y surge otra pregunta ¿Qué pasa con las bodegas y tiendas que no tienen valor alguno en lo que autoridad en la red se refieren? Miles de bodegas no saben como moverse en la red a diferencia de las 5 o 10 tiendas online más potentes, que sí que habrán cuidado su posicionamiento en buscadores. Sería positivo que las bodegas enlazaran a las tiendas y que unas se encargaran de la gestión de la fabricación y del embotellado y la otra se encargara de venderlo.

Las recomendaciones serían pues, que hubiese una mayor conexión entre webs del universo vinícola. Conseguir que haya una mayor unificación en este mundo permitiría a los nodos menos influyentes aspirar a moverse en zonas de más autoridad, ya que menos influencia en la red, no tiene porqué significar baja calidad del producto. ¿Quién sabe si el mejor vino del mundo no ha sido probado todavía por el mejor crítico del mundo, simplemente porque el encargado de la bodega, no ha sabido promocionar su producto en Internet?

Se aconseja, pues a las tiendas, promover sus webs y usar campañas de Adwords<sup>27</sup> que puedan atraer a grandes grupos de compradores potenciales.

Si se pudiese elaborar para el mercado del posicionamiento en buscadores una matriz DAFO (Debilidades, Amenazas, Fortalezas y Oportunidades) de un producto, el cuadrante de las oportunidades estaría lleno de elementos que ayudarían a los comerciantes a apostar por el marketing en Internet. En este proyecto, se ha demostrado que se puede apostar por la promoción en buscadores y en esforzarse por introducirse en los primeros puestos de las búsquedas orgánicas y que ello tiene como consecuencia, ganancias que compensaría el esfuerzo realizado.

---

<sup>27</sup> AdWords: sistema de publicidad de Google



## 6. Conclusiones

La gran pregunta es ¿Podemos estar todos en la primera página del buscador?

Se trata de una tarea larga, la de ubicar un enlace en las primeras posiciones de un buscador, ya sea Google, Yahoo!, o Bing. Para muchos Webmasters puede llegar a ser incluso una obsesión. Como se ha comentado al principio del trabajo, encontrar información en los buscadores de Internet se ha convertido en una de las actividades más frecuentes por el ser humano. La gente busca información y con ello se influencia por periódicos, blogs, empresas, etc. Y situarse en las páginas de resultados establece la diferencia entre estar presente o no en Internet.

Numerosos son los libros<sup>28</sup> que determinan que los tres primeros resultados de un buscador son leídos por la práctica totalidad de las búsquedas realizadas por personas. La posibilidad de conseguir un clic decae a partir de la segunda página y menos de un 10% llega ni si quiera a entrar a la tercera página de resultados.

El proceso del posicionamiento SEO es largo y necesita una constancia en el día a día. Los resultados que he podido obtener aseguran en un alto porcentaje que si se realiza el estudio como se ha visto, y se respetan los enlaces por temáticas, un portal puede entrar en las primeras posiciones de las SERPs.

La respuesta a la pregunta es ¡SI!, podemos. Al menos en las páginas de resultados de búsqueda que se relacionen con los criterios de búsqueda de sus clientes potenciales.

Compartir información te hace ser reconocido, visible. Aportando datos a los usuarios se consigue que siempre que deseen una opinión recurrirán al criterio de alguien influyente, alguien experto en un sector. Pero para poder ser alguien influyente, hay que saber con que webs tratar y en este proyecto, se da un indicio de por donde se debe recurrir a la hora de tratar con otras webs de una misma temática, es decir, con quien se debe contactar en un futuro para conseguir clientes.

Una tienda, por ejemplo, con los algoritmos actuales, de nada le sirve tener un enlace proveniente de una universidad, por mucha puntuación PageRank que tuviese en los años donde el PR aún se usaba como factor de posicionamiento. Esta universidad no le va a ayudar a vender sus productos.

---

<sup>28</sup> Prat M. (2009). Posicionamiento de su sitio Web: La posición ideal. Cornellá de Llobregat(Barcelona): ENI. [Consulta: 20-6-2015]

En cambio, una web como un blog, o una web de noticias que tenga un volumen elevado y continuo en el tiempo de visitas si que puede aportar un número de usuarios elevado que puedan ser al fin y al cabo, conversiones.

Sería interesante ver como hubiesen actuado los grafos en función del tiempo. Añadiendo la variable tiempo, se podría observar cual ha sido la evolución de las aristas (backlinks) y así tener una visión dinámica del progreso en la muestra.

Aunque lo logrado en este trabajo ha sido entender empíricamente la complejidad de las relaciones entre dominios web en un caso sectorial concreto, el del vino, será interesante ampliar esos conocimientos a otros ámbitos, como la tecnología, la moda, la salud, etc. y ver si los otros sectores funcionan de la misma manera o no.

Personalmente, un primer contacto con el Social Network Analysis ha conseguido que me salte la curiosidad por conocer nuevos conceptos. Todo lo relacionado con este proyecto ha ayudado a aprender gracias a herramientas de estudio muy potentes y que desconocía. Reconocer también el trabajo que hacen los programadores de Software Libre como es Gephi. Se pueden sacar conclusiones importantes y puede que ignoradas para otros que no hayan investigado en este tema.

Esto puede ser un inicio en la investigación que podría resultar muy amplia, puesto que es un sector, el del internet, vivo. Y día a día hay que descubrir nuevas técnicas de análisis para no de quedarse perdido en el pasado.



## 7. Bibliografía

- OUTBOUND LINKS AND SEO  
[<https://www.hochmanconsultants.com/articles/link-miser.shtml>].  
Hochman J.
- Relación entre la posición en Google y la cantidad de visitas  
[<http://www.alquiler-sitio-web.info/estudia-ctr-segun-la-posicion-en-google/>] Pierrick J.
- El uso de los buscadores en Internet: Buscadores en Internet. Conceptos, clasificación y características  
[[http://bvs.sld.cu/revistas/aci/vol11\\_3\\_03/aci04303.htm](http://bvs.sld.cu/revistas/aci/vol11_3_03/aci04303.htm)] Torres-Pombert A.
- Trends in research on global climate change: A Science Citation Index Expanded-based analysis [http://www.researchgate.net/profile/Yuh-Shan\_Ho/publication/259892846\_Trends\_in\_research\_on\_global\_climate\_change\_A\_Science\_Citation\_Index\_Expanded-based\_analysis/links/oc96052f5006d10409000000.pdf] Yuh-Shan Ho
- Cómo usar rel="nofollow" para enlaces específicos  
[<https://support.google.com/webmasters/answer/96569?hl=es#>] Cutts M.
- Orense- Fuentes M. & Rojas-Orduña O. (2010). SEO, Cómo triunfar en los buscadores: Explicación de las técnicas Black Hat SEO más comunes. Madrid: ESIC.
- What is the ideal keyword density of a page?  
[<https://www.youtube.com/watch?v=Rk4qgQdp2UA>] Cutts M
- Are .Edu and .Gov Links Really Worth it? [http://tripleseo.com/are-edu-gov-links-worth-it/] Dyson C.
- La centralidad en las redes sociales  
[<http://revistas.ucm.es/index.php/POSO/article/viewFile/POSO0000130131A/24622>] Freeman Linton
- Prat M. (2009). Posicionamiento de su sitio Web: La posición ideal. Cornellá de Llobregat(Barcelona): ENI.
- Social Network Analysis  
[[https://books.google.es/books?hl=es&lr=&id=MJoIGBfYDGEC&oi=fnd&pg=PP2&dq=%22social+network+analysis%22&ots=zWY\\_ZYkd8&sig=Nef9cS-engS2KGZaPBpulZXa7Io&redir\\_esc=y#v=onepage&q=%22social%20network%20analysis%22&f=false](https://books.google.es/books?hl=es&lr=&id=MJoIGBfYDGEC&oi=fnd&pg=PP2&dq=%22social+network+analysis%22&ots=zWY_ZYkd8&sig=Nef9cS-engS2KGZaPBpulZXa7Io&redir_esc=y#v=onepage&q=%22social%20network%20analysis%22&f=false)] Scott J.
- El Posicionamiento en Buscadores.(2001) Marckini F. Fuenlabrada (Madrid):DanyPress
- Maciá-Domene, F. & Gosende-Grela J. (2006). *Posicionamiento en Buscadores*. Madrid: Grupo Anaya





## 8. Glosario

**AdWords:** sistema de publicidad de Google. Son anuncios patrocinados, en los que se paga solo cuando un usuario hace click en tu anuncio (coste por click - CPC). Este tipo de campañas también se pueden realizar en modalidad de coste por mil impresiones (CPM) o coste por adquisición (CPA).

**Anchor text:** el “anchor text” es el texto del enlace. Por ejemplo, en el link [agencia marketing digital](#) el anchor text sería precisamente “agencia marketing digital”.

**Autoridad:** Es un término SEO que se emplea para definir a una página web de muy alta calidad y reputación en Internet, particularmente en el nicho o segmento al que pertenece. Significa que una página web tiene tan buen contenido y es tan útil que no solamente genera una gran cantidad de visitas, sino que además también motiva el que esos visitantes compartan el contenido con sus amigos en redes sociales u otros sitios.

**Backlink:** son los enlaces de otros sitios web que apuntan a las páginas de tu web. Si escribes en Google `link:www.tudominio.com` te listará todos los resultados.

**Blog:** un blog, o en español también una bitácora, es un sitio web periódicamente actualizado que recopila cronológicamente textos o artículos de uno o varios autores, apareciendo primero el más reciente, donde el autor conserva siempre la libertad de dejar publicado lo que crea pertinente. El término blog proviene de las palabras web y log ('log' en inglés = diario). El término bitácora, en referencia a los antiguos cuadernos de bitácora de los barcos, se utiliza preferentemente cuando el autor escribe sobre su vida propia como si fuese un diario, pero publicado en Internet en línea.

**Click:** visita que proviene de clicar sobre un enlace en un buscador (Google, Yahoo! Search, Msn, Ask,...).

**Contenido Viralizable:** Son aquellos contenidos que se propagan con rapidez compartidos por los usuarios a través del correo o de las redes sociales. Funcionan bien como contenido virales: los test para medir el conocimiento de un tema, las listas (“los diez mejores” ...), aplicaciones gratuitas, tutoriales especializados, etc.



**CPC:** Coste por click. Término utilizado en campañas de publicidad en Internet o de pago por click (PPC).

**CTR:** relación porcentual entre clicks e impresiones (número de veces que se muestra tu página en los buscadores). Es un valor que nos indicará la efectividad de nuestras campañas.

**Densidad de palabras clave:** Es el ratio de apariciones de una palabra o frase clave en el total de palabras que componen un texto. Es uno de los criterios de relevancia que tienen en cuenta los buscadores.

**Directorio:** Un directorio es un listado de direcciones de sitios web organizado en categorías por seres más o menos humanos. Puede que tenga una función de búsqueda, pero ésta se limita a emparejar las palabras buscadas con las palabras descriptivas encontradas en la ficha de cada sitio y no con los contenidos de las páginas que componen estos sitios.

**DMOZ:** el Open Directory Project (ODP), también conocido como DMoz (por Directory Mozilla) es un proyecto colaborativo en el que editores voluntarios listan y categorizan enlaces a páginas web. Cualquier persona puede sugerir un enlace en una categoría determinada que luego ha de ser aprobada por un editor. Cualquier persona puede optar a ser editor rellenando un formulario en el que defiende su aptitud para organizar una categoría.

**Espinear:** Tomar un texto y sustituirlo por sinónimos, otras formas verbales, etc. para que parezca original cara a los buscadores.

**Footprints:** Comandos que podemos teclear en Google para realizar búsquedas específicas.

**Google:** el buscador más utilizado en Internet.

**Google Panda:** Es el algoritmo de Google que fue lanzado el 24 de febrero de 2011 y que da prioridad a las webs con contenido propio y de calidad.

**Google Penguin:** fue la segunda gran actualización de Google y pretendía focalizarse en el Link Building<sup>29</sup> automatizado.

---

<sup>29</sup> Link Building: Construcción de enlaces manual. Se explicará en el punto 2.2.c

**Indexar:** Introducir una web en la base de datos del buscador

**Keyword:** Buscar “Palabra Clave”

**Keyword Stuffing:** Acción de inflar artificialmente la densidad de palabras clave de una web

**Link:** Hiperenlace, (también llamado enlace, vínculo, hipervínculo o link) es un elemento de un documento electrónico que hace referencia a otro recurso, por ejemplo, otro documento o un punto específico del mismo o de otro documento. Combinado con una red de datos y un protocolo de acceso, un hiperenlace permite acceder al recurso referenciado en diferentes formas, como visitarlo con un agente de navegación, mostrarlo como parte del documento referenciador o guardarlo localmente.

**Link Baiting:** Es la técnica de crear una buena historia (un tutorial, una utilidad, un concurso, etc.) y promocionarla adecuadamente para conseguir el mayor número de enlaces en el menor tiempo posible.

**Link Building:** Se trata una de las estrategias SEO más “clásicas”, se trata de conseguir enlaces que apunten a nuestra página, de esta forma los buscadores la considerarán relevante puesto que la “recomiendan” en varios sitios. Esto significa que la ofrecerá en posiciones más altas en la página de resultados de la búsqueda.

**Long Tail:** Es un término que hace referencia a la utilización de 3 o más frases o palabras como texto ancla para la construcción de enlaces de retroceso. La construcción de un vínculo de cola larga como también se le conoce, no es más que la generación de enlaces de retroceso con palabras clave de cola larga hacia el contenido que ya se tiene publicado con la intención de ayudar a incrementar el ranking de ese contenido para más frases. Uno de los principales beneficios del Linkbuilding Long Tail, es que no únicamente se vinculan a la palabra clave, sino que también lo hacen a la versión de cola larga de esa misma palabra clave.

**Optimización:** la optimización implica hacer que las páginas web sean más fáciles de leer para los buscadores y enfatizar las palabras clave que se relacionan directamente con su contenido.

**Outbound links** : enlaces a otras páginas Web desde nuestra página.

**Pagerank**: sistema de ordenación de los resultados de Google.

**Palabra clave**: un término (o términos) relevantes para nuestro proyecto. El objetivo de cualquier estrategia de posicionamiento SEO consiste en posicionar las keywords objetivo en las primeras páginas de los resultados de búsqueda.

**Proxy**: intermediario en las peticiones de recursos que realiza un cliente (A) a otro servidor (B)

**Ranking**: puesto que ocupa una URL en un buscador por una determinada palabra clave.

**Sand-box**: Es un efecto observado, que postula un sistema de filtrado especial de Google en su algoritmo, que hace que las páginas web de nueva creación no aparezcan, o aparezcan en posiciones muy inferiores, en los resultados de búsqueda

**SEM**: Search Engine Marketing - Marketing en buscadores.

**SEO**: Search Engine Optimization - Optimización en buscadores.

**SERPs**: hace referencia a los resultados que muestra una página de buscador como pueden ser: Google, Bing, y Yahoo! entre los más destacados.

**Spam**: publicitario, basura

**Spiders**: robots de los buscadores que acceden a las páginas y las indexan según su contenido ateniéndose a una serie de factores numéricos.

**URL**: Uniform Resource Locator - Localizador de recursos uniforme.

**Visita**: representa a un único usuario que ha accedido al sitio web.

**WWW**: World Wide Web.

