

# Modulo 3 - Centralidades

Jorge Fábrega

Versión 2023

## Introducción a la Centralidad y Prestigio

Las medidas de centralidad y prestigio forman parte crucial en la comprensión de la dinámica y estructura de las redes. Su estudio permite adentrarse en la comprensión profunda de cómo ciertos nodos dentro de una red adquieren más importancia o influencia que otros, y cómo estas características se pueden medir y analizar.

Las medidas de centralidad nos ayudan a identificar los nodos más importantes dentro de una red, pero **¿centrales en qué?**

La respuesta depende del objetivo del análisis. Los nodos más centrales en una red lo serán dependiendo de su posición, conexiones y contexto. Por ejemplo, algunos nodos pueden ser centrales en influir significativamente en la difusión de información o en la propagación de enfermedades, pero otros en la toma de decisiones dentro de un grupo dentro de la red. La noción de centralidad que será pertinente para estudiar estos distintos énfasis puede diferir. En nuestro análisis, exploraremos diferentes tipos de centralidades, pero las elementales a tener siempre presente son: la centralidad de grado, de cercanía, de intermediación y la eigenvector centrality (o centralidad de vector propio).

Un ejemplo clásico de la importancia de estas medidas se encuentra en el análisis del capital social, donde los individuos con alta centralidad de grado pueden ser los más influyentes. En redes de comunicación, aquellos con alta centralidad de intermediación son críticos para la transferencia de información entre distintas partes de la red. En el ámbito de la investigación biomédica, estas medidas ayudan a identificar “genes centrales” en redes de interacción génica, como se explora en el trabajo de Barabási y Oltvai (2004), “Network biology: understanding the cell’s functional organization”. En el contexto académico, la aplicación de estas medidas no solo es teórica. Proyectos reales, como el análisis de redes de coautorías en publicaciones científicas, redes de relaciones entre empresas y mercados o redes formales e informales dentro de organizaciones, ofrecen terreno fértil para la aplicación práctica de estas métricas.

Esta sesión buscará no solo entender estas medidas desde una perspectiva teórica, sino también aplicarlas utilizando R.

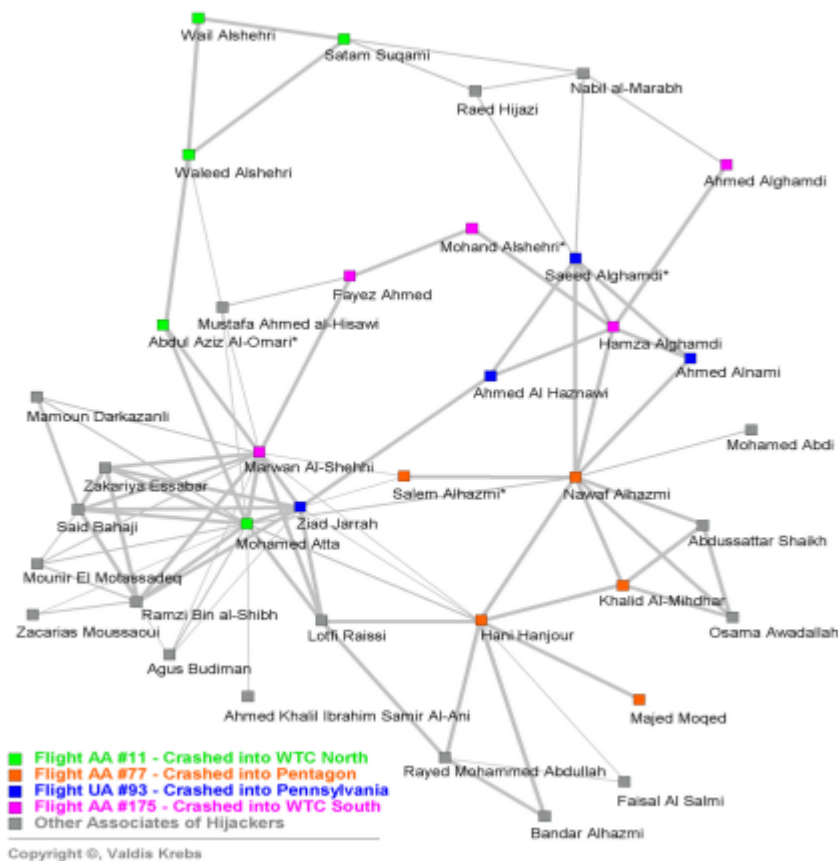
## Importancia en Redes Sociales

En “Mapping Networks of Terrorist Cells”, Valdis Krebs analizó las redes encubiertas implicadas en los ataques terroristas del 11 de septiembre de 2001. A través del uso de datos públicos, Krebs logró mapear una red que arrojó luz sobre las conexiones entre los individuos involucrados en el ataque.

Para ello, Krebs comenzó a trazar las conexiones entre los 19 secuestradores y sus asociados a partir de la información que salía en medios de prensa y públicos, revelando un entramado de relaciones que habían permanecido ocultas a simple vista.

Su análisis mostró cómo, a través de vínculos como residencias compartidas y reuniones documentadas, que estos individuos formaron una red cuya estructura era tanto intrincada como reveladora. Los resultados de Krebs no solo destacaron la importancia de los nodos clave, que actuaban como puentes entre distintos grupos de secuestradores, sino que también iluminaron la forma en que la red terrorista se había tejido en la sombra.

Krebs demostró que la información pública, cuando se analiza a través del prisma del análisis de redes, puede revelar patrones y conexiones que de otro modo pasarían desapercibidos. Este estudio no solo contribuyó a la comprensión del evento específico del 11 de septiembre, sino que también abrió caminos para futuras investigaciones en la lucha contra las redes terroristas y criminales.



**Figure 4. Hijacker's Network Neighborhood**

*This dense under-layer of prior trusted relationships made the hijacker network both stealth and resilient. Although we don't know all of the internal ties of the hijackers' network it appears that many of the ties were concentrated around the pilots. This is a risky move for a covert network. Concentrating both unique skills and connectivity in the same nodes makes the network easier to disrupt –once it is discovered. Peter Klerks (Klerks 2001) makes an excellent argument for targeting those nodes in the network that have unique skills. By removing those necessary skills from the project, we can inflict maximum damage to the project mission and goals. It is possible that those with unique skills would also have unique ties within the network. Because of their unique human capital and their high social capital the pilots were the richest targets for removal from the network. Unfortunately they were not discovered in time (Krebs, 2002)*

# Medidas de Centralidad

Para mirar la relevancia de las distintas medidas de centralidad, aplicaremos los algoritmos en la red Karate de Zachary.

## Centralidad de Grado

Definición: Número de enlaces incidentes en un nodo. Es decir, simplemente mide el número de conexiones que un nodo tiene. En su análisis clásico, Freeman (1978) en “Centrality in Social Networks Conceptual Clarification” subraya la importancia de esta medida para entender la actividad potencial de un nodo en una red.

Se define como:

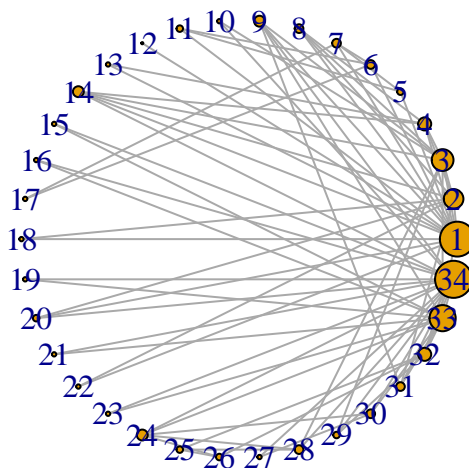
$$C_D(v) = \deg(v)$$

Ventajas y Desventajas: Explicar la simplicidad y limitaciones.

Código en R:

```
library(igraph)
g <- make_graph("Zachary")
layout <- layout_in_circle(g)
deg_cent <- degree(g, mode="all")
plot(g, layout=layout, vertex.size=deg_cent, main="Centralidad de Grado")
```

## Centralidad de Grado



## Centralidad de Cercanía

Definición: La centralidad de cercanía del nodo  $v$  es el inverso de la suma de las distancia a todos los otros nodos. Esta medida está relacionada con cuantificar la eficiencia en la comunicación dentro de la red. Fue descrita por Bavelas (1950) en “Communication Patterns in Task-Oriented Groups”, se enfoca en la proximidad de un nodo a todos los demás en la red (se requiere conectividad entre los nodos).

Se define como:

$$C_C(v) = \frac{1}{\sum_{u \neq v} d(v, u)}$$

Donde  $d(v, u)$  es la geodésica entre  $v$  y  $u$ .

Ventajas y Desventajas: Discutir la relevancia en contextos de eficiencia de comunicación.

Código en R:

```
close_cent <- closeness(g, mode="all")  
plot(g, layout=layout, vertex.size=close_cent*500, main="Centralidad de Cercanía")
```

## Centralidad de Cercanía

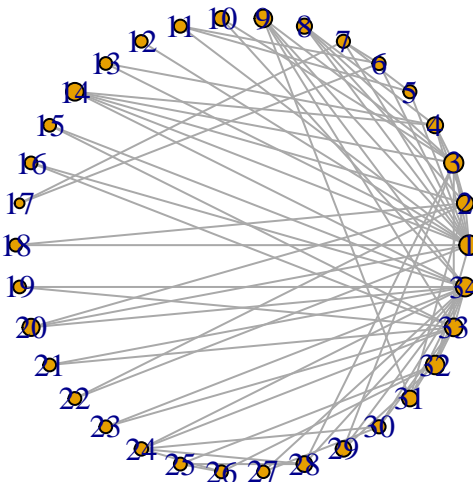


Figure 1: Centralidad de Cercanía

## Centralidad de Intermediación

Definición: Frecuencia con la que un nodo actúa como puente en el camino más corto entre otros dos nodos. Fue introducida por Freeman en 1977, esta centralidad mide la frecuencia con la que un nodo está en las geodésicas entre otros dos nodos. Es esencial para comprender los puntos de control y flujo en la red.

Se define como:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{d_{st}(v)}{d_{st}}$$

Ventajas y Desventajas: Debate sobre su utilidad para identificar nodos de control.

Código en R:

```
betw_cent <- betweenness(g, directed=FALSE)
plot(g, layout=layout, vertex.size=betw_cent/10, main="Centralidad de Intermediación")
```

## Centralidad de Intermediación

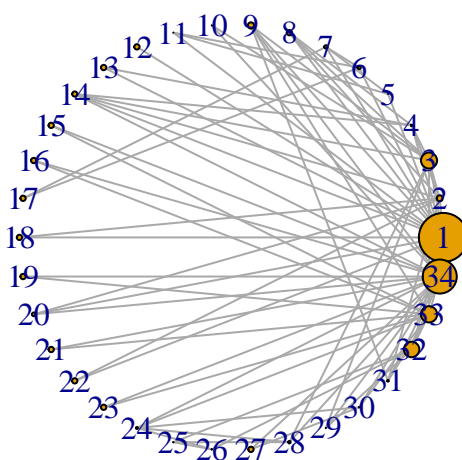


Figure 2: Centralidad de Intermediación

## Centralidad de Vector Propio o Eigenvalue centrality

Definición: Medida de la influencia de un nodo en una red. Fue discutida en profundidad por Bonacich en su trabajo de 1987 “Power and Centrality: A Family of Measures”, considera no solo la cantidad de conexiones, sino también la calidad. Es decir, la centralidad de un nodo no depende sólo de estar muy conectado sino también de estar “bien” conectado, que en este caso significa estar conectado con otros que a su vez están muy conectados.

La centralidad de vector propio no tiene una fórmula única, se encuentra a través del vector propio asociado al valor propio más grande de la matriz de adyacencia (o primer eigenvalue).

$$\mathbf{Ax} = \lambda \mathbf{x}$$

Donde: -  $\mathbf{A}$  es la matriz de adyacencia de la red. -  $\mathbf{x}$  es el vector de centralidad de eigenvector. -  $\lambda$  es el mayor valor propio (eigenvalue) de la matriz  $\mathbf{A}$ .

Cada componente del vector de centralidad de eigenvector  $x_i$  (que corresponde al nodo  $i$ ) es proporcional a la suma de las centralidades de los nodos a los que está conectado:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^n a_{ij} x_j$$

donde  $a_{ij}$  son los elementos de la matriz de adyacencia  $\mathbf{A}$ , indicando si los nodos  $i$  y  $j$  están conectados (1) o no (0).

Ventajas y Desventajas: Discusión sobre la influencia basada en la estructura de la red.

Código en R:

```
eigen_cent <- eigen_centrality(g)$vector
plot(g, layout=layout, vertex.size=eigen_cent*100, main="Centralidad de Vector Propio")
```

**Centralidad de Vector Propio**

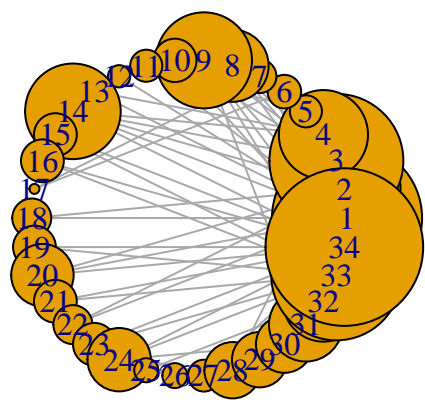


Figure 3: Centralidad de Vector Propio

## Centralidad de Bonacich (o Power Centrality)

Fue propuesta en 1987, en Power and Centrality: A Family of Measures por Bonacich. La centralidad de Bonacich se ajusta con un parámetro  $\alpha$  que balancea la importancia de las conexiones directas e indirectas:

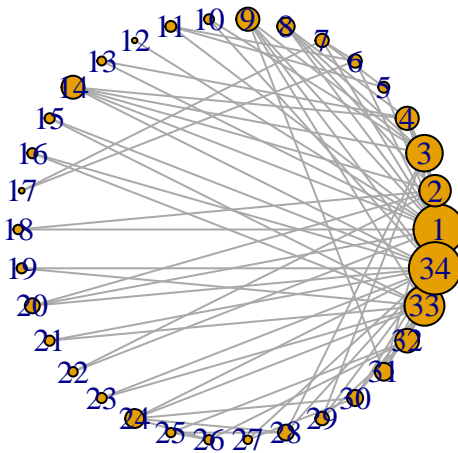
$$C_B(v) = \alpha \sum_{j=1}^n A_{vj} x_j + \beta x_v$$

donde  $A_{vj}$  es la matriz de adyacencia de la red y  $x_j$  es el valor de centralidad de Bonacich del nodo  $j$ .  $\alpha$  es un parámetro que determina cuánto influyen las conexiones directas de un nodo en su centralidad (generalmente es el inverso del valor propio más grande de la matriz de adyacencia de la red) y  $\beta$  es otro parámetro que determina la importancia de las conexiones indirectas en la centralidad de un nodo.

Código en R:

```
eigen <- eigen_centrality(g)
lambda_max <- eigen$value
alpha <- lambda_max/100
bonpow_cent <- power_centrality(g, exponent=alpha) # en este caso le estamos dando el mismo peso
plot(g, layout=layout, vertex.size=bonpow_cent*10, main="Centralidad de Bonacich")
```

## Centralidad de Bonacich





## Ejercicio Práctico Comparativo

Comparemos ahora las distintas medidas de centralidad.

Código en R:

```
library(igraph)
library(GGally)

# Crear el grafo de Karate de Zachary
g <- make_graph("Zachary")

# Calcular las centralidades
deg_cent <- degree(g, mode="all")
close_cent <- closeness(g, mode="all")
betw_cent <- betweenness(g)
eigen_cent <- eigen_centrality(g)$vector
bonpow_cent <- power_centrality(g, exponent=2)

# Crear un data frame con las centralidades
cent_df <- data.frame(Degree=deg_cent,
                      Closeness=close_cent,
                      Betweenness=betw_cent,
                      Eigenvector=eigen_cent,
                      Bonacich=bonpow_cent)

# Usar GGally para crear gráficos de dispersión de cada par de centralidades
ggpairs(cent_df)
```

## Otras medidas de centralidad

La siguiente tabla muestra otras medidas interesantes de centralidad:

Medida de Centralidad	Referencia	Expresión Matemática	Explicación	Ejemplo de Utilidad
Katz Centrality	Katz, L. (1953). A New Status Index Derived from Sociometric Analysis. Psychome- trika, 18(1), 39–43.	$x_i = \sum_{k=1}^{\infty} \sum_{j=1}^N \alpha^k (A^k)_{ji}$	Mide el número de todos los nodos que pueden conectarse a través de un camino, penalizando a los nodos distantes.	Util en análisis de influencia en redes con conexiones indirectas.

Medida de Centralidad	Referencia	Expresión Matemática	Explicación	Ejemplo de Utilidad
PageRank	Brin, S., & Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Networks and ISDN Systems, 30(1-7), 107–117.	$x_i = \alpha \sum_j a_{ji} \frac{x_j}{L(j)} + \frac{1-\alpha}{N}$	Variante de la centralidad de eigenvector, considera la importancia de los nodos enlaces en un grafo dirigido.	Fundamental en motores de búsqueda para clasificar la importancia de páginas web.
Percolation Centrality	Piraveenan, M., Prokopenko, M., & Hossain, L. (2013). Percolation Centrality: Quantifying Graph-Theoretic Impact of Nodes during Percolation in Networks. PLoS ONE, 8(1), e53095.	$PC^t(v) = \frac{1}{N-2} \sum_{s \neq v \neq r} \frac{\sigma_{sr}(v)}{\sigma_{sr}} \frac{x_s^t}{\sum [x_i^t] - x_v^t}$	Mide la importancia de los nodos en términos de ayudar a la percolación a través de la red.	Aplicable en la propagación de enfermedades o información en redes sociales.

La generación de medidas de centralidad es un área activa donde permanentemente están surgiendo nuevas variantes. Una conceptualización interesante para agrupar a las distintas métricas fue propuesta por Borgatti distinguiendo índices que son **mediales** e índices que son **radiales**.

Ahora que hemos visto varias medidas de centralidad:

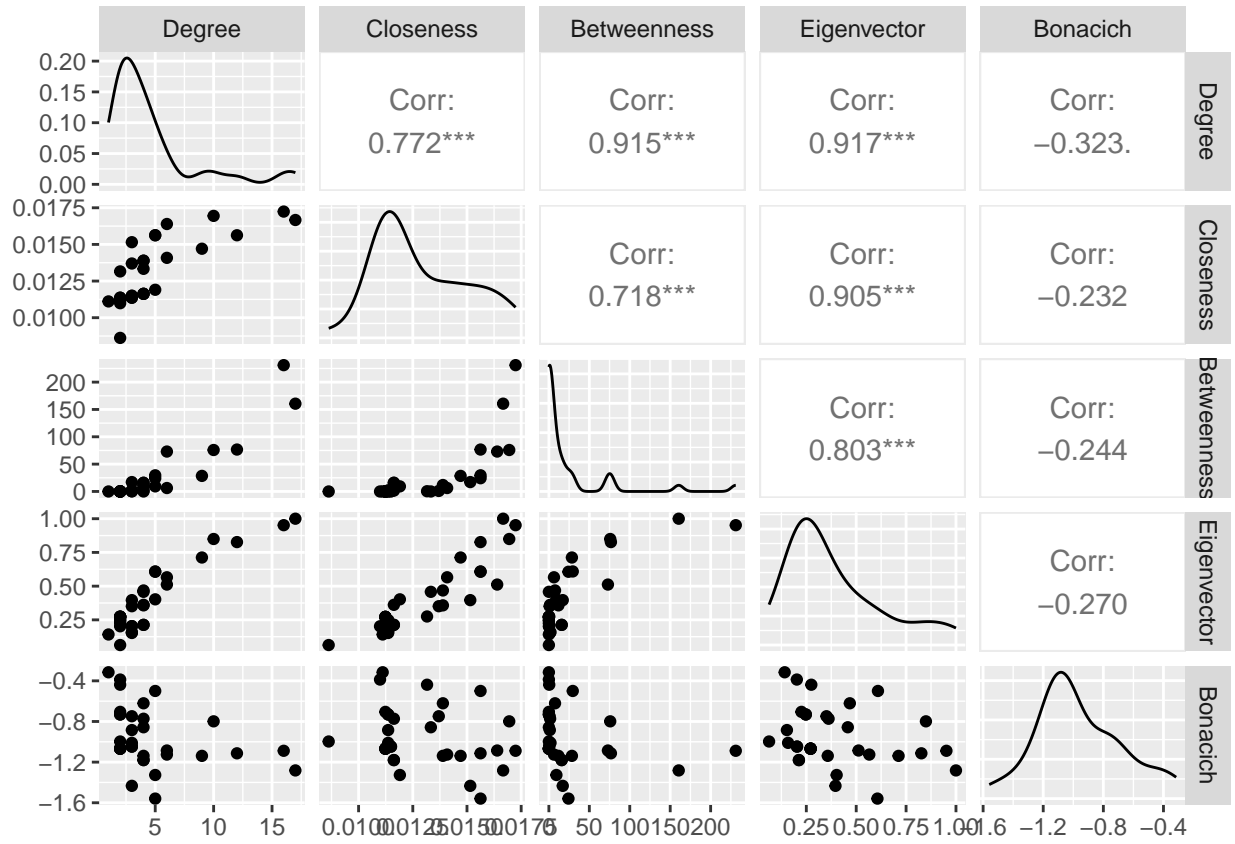
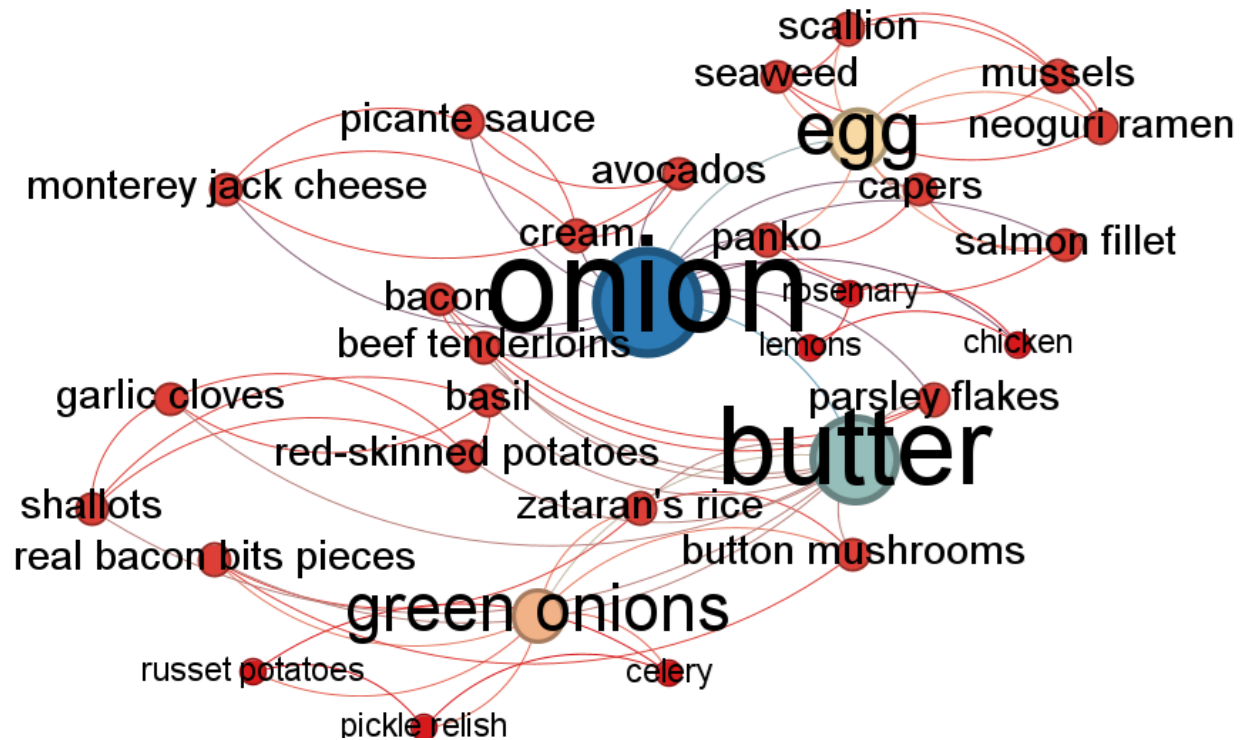


Figure 4: Comparación de Centralidades

## TAREA

Ingresa a la siguiente página web <https://www.allrecipes.com/>. Ingresa el nombre de un ingrediente y extrae los datos de las recetas generadas. Con esa información genera una red, calcula diversas medidas de centralidad y compáralas. Genera visualizaciones. La siguiente, por ejemplo, la generé usando la cebolla como ingrediente de búsqueda.



Tips: Deberas registrarte en la página y obtener una llave (key) para ingresar a la API. Posteriormente debes usar alguno de los métodos para extraer datos. Yo usé la llamada que permite extraer recetas asociadas a un ingrediente específico. El código que utilicé fue el siguiente: