



# Social Network Analysis SNA con igraph en R

Pedro Concejero

[pedro.concejero@um.es](mailto:pedro.concejero@um.es)

Grupo R-madRid

<http://madrid.r-es.org/>



# Por qué y para qué SNA

---

Material:

[http://en.wikipedia.org/wiki/Social\\_network\\_analysis](http://en.wikipedia.org/wiki/Social_network_analysis)

[http://en.wikipedia.org/wiki/Social\\_network\\_analysis\\_software](http://en.wikipedia.org/wiki/Social_network_analysis_software)

<https://igraph.org/r/>

<https://igraph.org/python/doc/tutorial/tutorial.html>

<https://kateto.net/tutorials/>

<https://kateto.net/networks-r-igraph>



# Por qué y para qué SNA

---

- SNA o análisis de redes sociales es una metodología multidisciplinar que es básica para multitud de análisis, desde analítica de márketing hasta periodismo.
- -> ejemplos de SNA (p.ej. Fútbol en [http://www.casos.cs.cmu.edu/computational\\_tools/datasets/external/football/index11.php](http://www.casos.cs.cmu.edu/computational_tools/datasets/external/football/index11.php) )
- -> en text-mining <http://rdatamining.wordpress.com/2012/05/17/an-example-of-social-network-analysis-with-r-using-package-igraph/>



## Por qué y para qué SNA

---

- Como con muchas otras metodologías, la disponibilidad de librerías R (en concreto igrph) facilita muchísimo este tipo de análisis, incluso para analistas sin experiencia en el campo (aunque es necesario tener experiencia en el lenguaje R).
- SNA es tanto un fin en sí mismo, como un paso previo para multitud de análisis, en muchos sectores.

# Ejemplos

## Forensic Analysis of Phone Call Networks

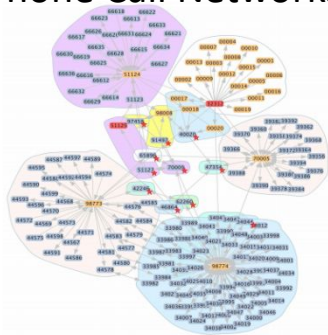


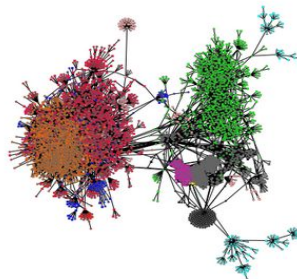
Fig. 3 Example of visual aggregation layout. It is based on the Edge Betweenness Clusterer to divide the network in different clans on the base of the interactions among members.

<http://arxiv.org/abs/1303.1827>

<https://arxiv.org/abs/1303.1827>

## Social Networks Reveal Structure (And Weaknesses) of Businesses

Computer scientists have recreated the international and organisational structure of large corporations using publicly available data on social networks

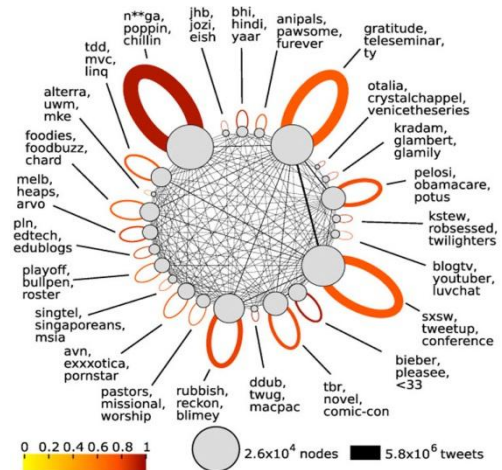


<http://arxiv.org/abs/1303.3741>

<https://arxiv.org/abs/1303.3741>

## Twitter users forming tribes with own language, tweet analysis shows

Analysis of millions of tweets finds more precise use of social media which seems to contradict idea that Twitter users want to share everything with everyone



<http://www.guardian.co.uk/news/datablog/2013/mar/15/twitter-users-tribes-language-analysis-tweets#>

<https://www.theguardian.com/news/datablog/2013/mar/15/twitter-users-tribes-language-analysis-tweets>

# Ejemplos

<https://www.researchgate.net/profile/Mariluz-Congosto/research>

20

*Ismael Peña-López, Mariluz Congosto and Pablo Aragón*

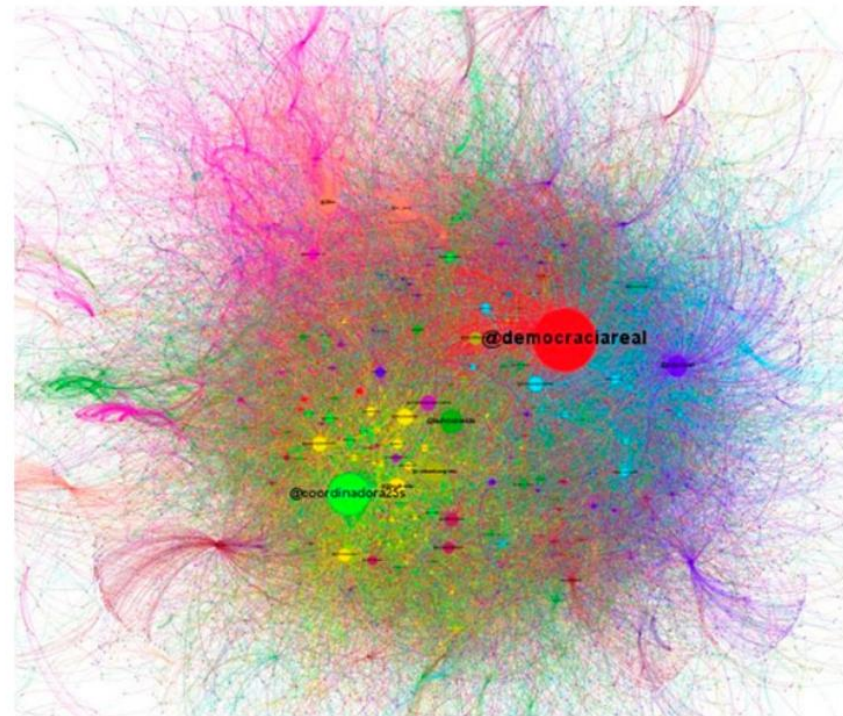
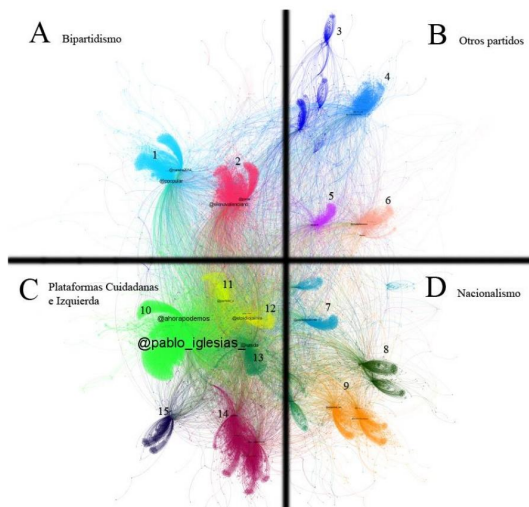
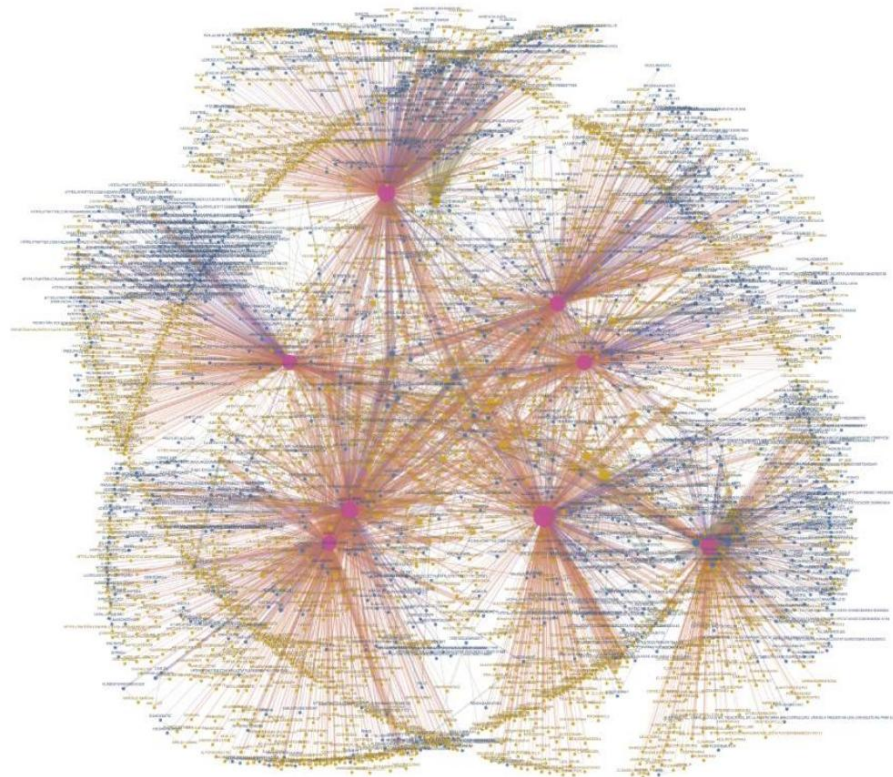


Figure 12. 25S – map of RTs “Late”



# Ejemplos

<https://www.researchgate.net/profile/Mariluz-Congosto/research>



**Figura 1.1** - Ecosistema de los candidatos del 21D.



# El caso enron: SNA sobre dataset emails

---

**Todo el material disponible aquí:**

**[https://github.com/pedroconcejero/SNA\\_talleR\\_muRcia](https://github.com/pedroconcejero/SNA_talleR_muRcia)**

**(.Rmd incluye código y texto explicativo + enlaces)**

**NECESARIA INSTALACIÓN DE:**

**igraph:**

**<https://cran.r-project.org/web/packages/igraph/index.html>**

**gephi:**

**<https://gephi.org/users/download/>**

**(v. 0.9.3 or later)**





# Conceptos básicos

---

## Red Social

Una Red Social es una estructura en la que se representan los elementos que interactúan entre sí (pueden ser individuos u otros elementos), que se llaman **nodos (o vértices)**, y las interacciones, que se llaman **enlaces**.

Una Red Social o “grafo” muestra conjuntamente las formas en las que se conectan o relacionan los nodos a través de los enlaces.

Este tipo de representación permite aplicar entonces herramientas matemáticas de análisis (Teoría de Grafos).



# Creación de grafos con igraph

---

- Documentación igraph:  
<http://cran.r-project.org/web/packages/igraph/igraph.pdf>
- <https://kateto.net/networks-r-igraph> (tutorial)
- Formas de crear o importar grafos:
  - - desde formatos específicos tipo graphml o pajek
  - - desde dataframe: función `graph.data.frame`
- Grafos dirigidos o no dirigidos. Primera e importante decisión -> en función del tipo de enlace y de los objetivos del análisis.



# Creación de grafos con igraph

---

- Grafo de comunicaciones y grafo social: el papel de la reciprocidad
- Funciones disponibles una vez tenemos el objeto grafo:
  - - creación de subgrafos – comando `subgraph`
  - - exportar grafos a formatos externos (ejemplo: `graphml` para usar en gephi)



# Medidas SNA globales (del grafo completo)

---

<https://kateto.net/wp-content/uploads/2018/03/R%20for%20Networks%20Workshop%20-%20Ognyanova%20-%202018.pdf#section.5>

- Densidad
- Reciprocidad
- Transitividad
- Diámetro



# Medidas SNA individuales

---

- Métricas de centralidad (teoría, matrices de adyacencia):
  - <https://en.wikipedia.org/wiki/Centrality>
  - [https://en.wikipedia.org/wiki/Adjacency\\_matrix](https://en.wikipedia.org/wiki/Adjacency_matrix)
- Objetivos:
  - Conocer las medidas de centralidad disponibles en igraph
  - Calcular e incorporar medidas de centralidad a la tabla de nodos
-



# Medidas SNA individuales

---

## Métricas Individuales SNA – Centralidad

<https://kateto.net/wp-content/uploads/2018/03/R%20for%20Networks%20Workshop%20-%20Ognyanova%20-%202018.pdf#subsection.5.7>

- Grado-Salida (out-degree): La suma de conexiones de un actor hacia el resto
- Grado-Entrada (in-degree): La suma de conexiones hacia un actor.
- Centralidad Próxima (closeness centrality): Es la distancia media de un actor al resto de actores en una red.
- Centralidad Intermedia (betweenness centrality): Es el número que representa cómo es de frecuente que un actor esté entre los caminos geodésicos de otros actores.



# Visualización de grafos con igraph y gephi

---

Igraph:

<https://igraph.org/python/doc/tutorial/tutorial.html#layouts-and-plotting>

y gephi: <https://gephi.org/users/download/>

Muy importante: ajustar el tamaño del grafo

Layouts

Elegir propiedades de nodos y enlaces para propiedades gráficas  
visualización

Grafos estáticos y dinámicos

Visualización de grafos dinámicos con gephi





# Why and when use SNA

---

- Whenever you are studying a **data set that includes nodes and links** (the basis of a social graph). I.e. telephone network, mail, business (ie internally -organization; externally -customers), and of course what is know known as social network, either offline or online -> SNA provides the tools to make lots of analysis, **obtain insights and knowledge**
- When you want to **visualize** your data so as to uncover **patterns in relationships or interactions**
- When you want to follow the **paths that information** (or basically anything) follows in social networks
- When you do **quantitative research**, although for qualitative research a network perspective is also valuable
- The range of actions and opportunities afforded to individuals are often a function of their positions in social networks; **uncovering these positions** (instead of relying on common assumptions based on their roles and functions, say as fathers, mothers, teachers, workers) can yield more interesting and sometimes surprising results
- A quantitative analysis of a social network can help you identify different **types of actors** in the network or **key players**, whom you can focus on for your qualitative research
- SNA is clearly also useful to identify the **communities** or networks, and to promote **social cohesion and growth** in an online community

# Where

---

- Gephi datasets
  - <https://github.com/gephi/gephi/wiki/Datasets>
- SNAP Stanford
  - <http://snap.stanford.edu/data/#socnets>
- Katya Ognyanova
  - <https://kateto.net/2016/05/network-datasets/>
- Konect networks
  - <http://konect.cc/networks/>
- Alex Arenas Website
  - <https://deim.urv.cat/~alexandre.arenas/data/welcome.htm>
- Social Feed Manager (SFM)
  - <https://gwu-libraries.github.io/sfm-ui/>
  - <https://gwu-libraries.github.io/sfm-ui/posts/2017-09-08-sna>



# Examples: Movies

---

- Social Networks in movies
  - <https://moviegalaxies.com/>
  - <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/T4HBA3>
- Social Network of Movie Stars (with Python + NetworkX)
  - <https://arxiv.org/abs/2002.11103>
- IMDB Top 1000 movie data analysis
  - <https://github.com/aleksa2808/imdb-data-analysis>
- Feature networks
  - <http://konect.cc/categories/Feature/>
- Lexical networks
  - <http://konect.cc/categories/Lexical/>
- Jazz musicians
  - <http://konect.cc/networks/arenas-jazz/>



## Referencias adicionales

---

- <https://faculty.ucr.edu/~hanneman/nettext/>
- [http://faculty.ucr.edu/~hanneman/nettext/C10\\_Centrality.html](http://faculty.ucr.edu/~hanneman/nettext/C10_Centrality.html)
- <https://kateto.net/tutorials/>
- <http://en.wikipedia.org/wiki/Centrality>