



GRANDATA[®]

Ciencia de Redes (Humanas y Sociales)

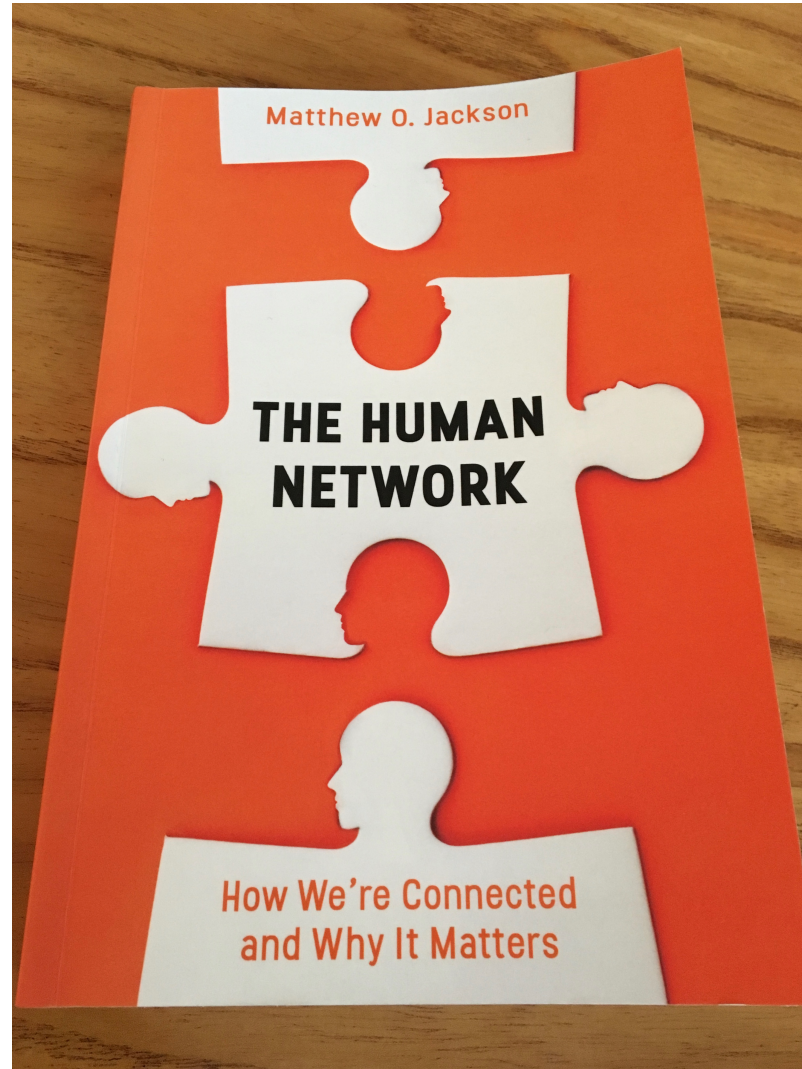
#4

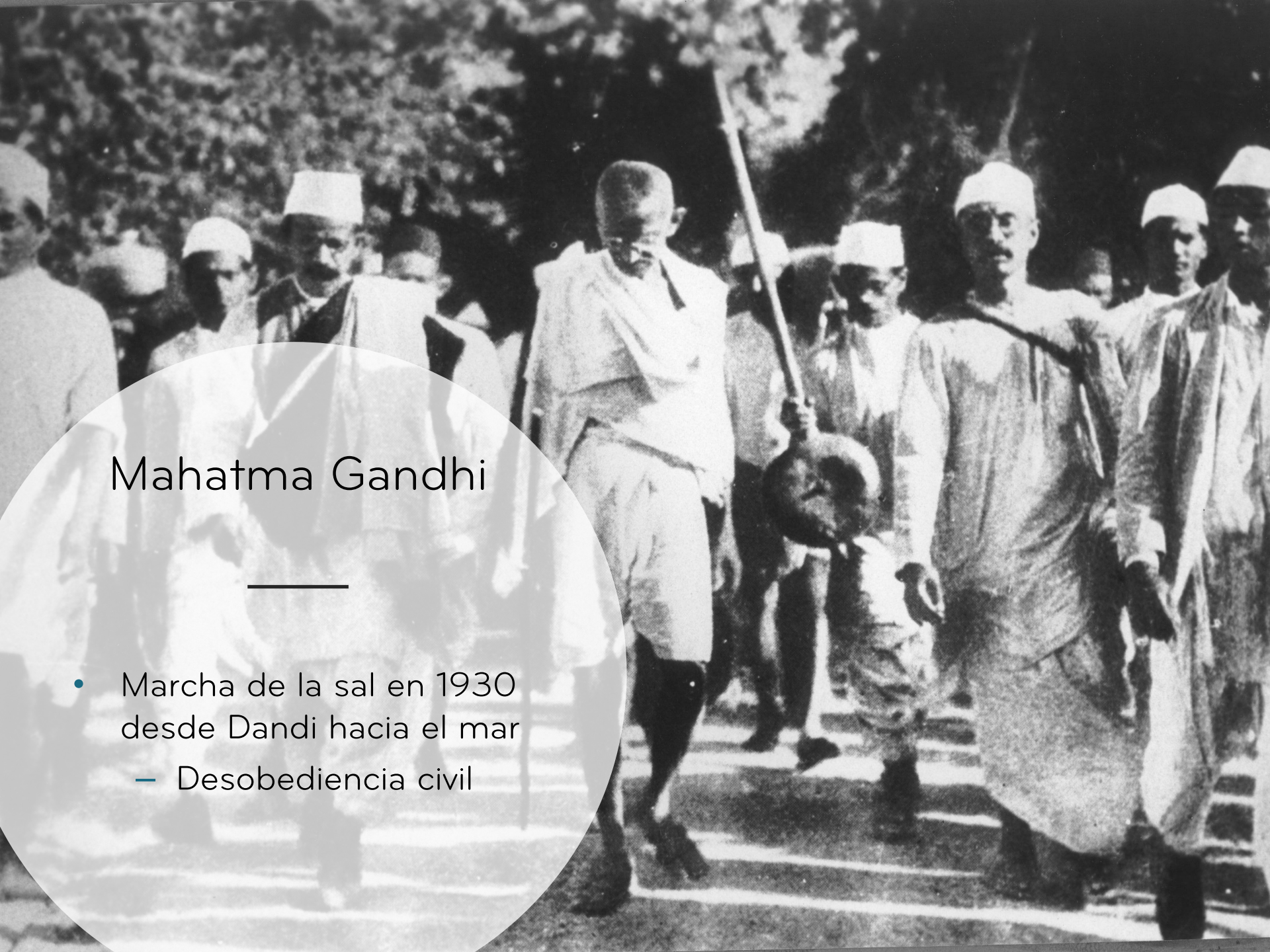
Carlos Sarraute

Instituto de Cálculo, Abril-Junio 2019

Medidas de Centralidad

The Human Network



A black and white photograph showing a group of men, including Mahatma Gandhi, walking outdoors. They are dressed in traditional Indian attire, including white dhotis and shawls, and white Gandhi caps. Some men are carrying long wooden poles. The background shows trees and a path.

Mahatma Gandhi

- Marcha de la sal en 1930 desde Dandi hacia el mar
 - Desobediencia civil

Michael Jordan



Medidas de centralidad

- Grado
 - Cantidad de vecinos
- Excentricidad
 - Distancia al nodo más lejano
- Betweenness centrality
 - Proporción de caminos mínimos que pasan por un nodo
- Closeness centrality
 - Inverso de la distancia al resto de los nodos

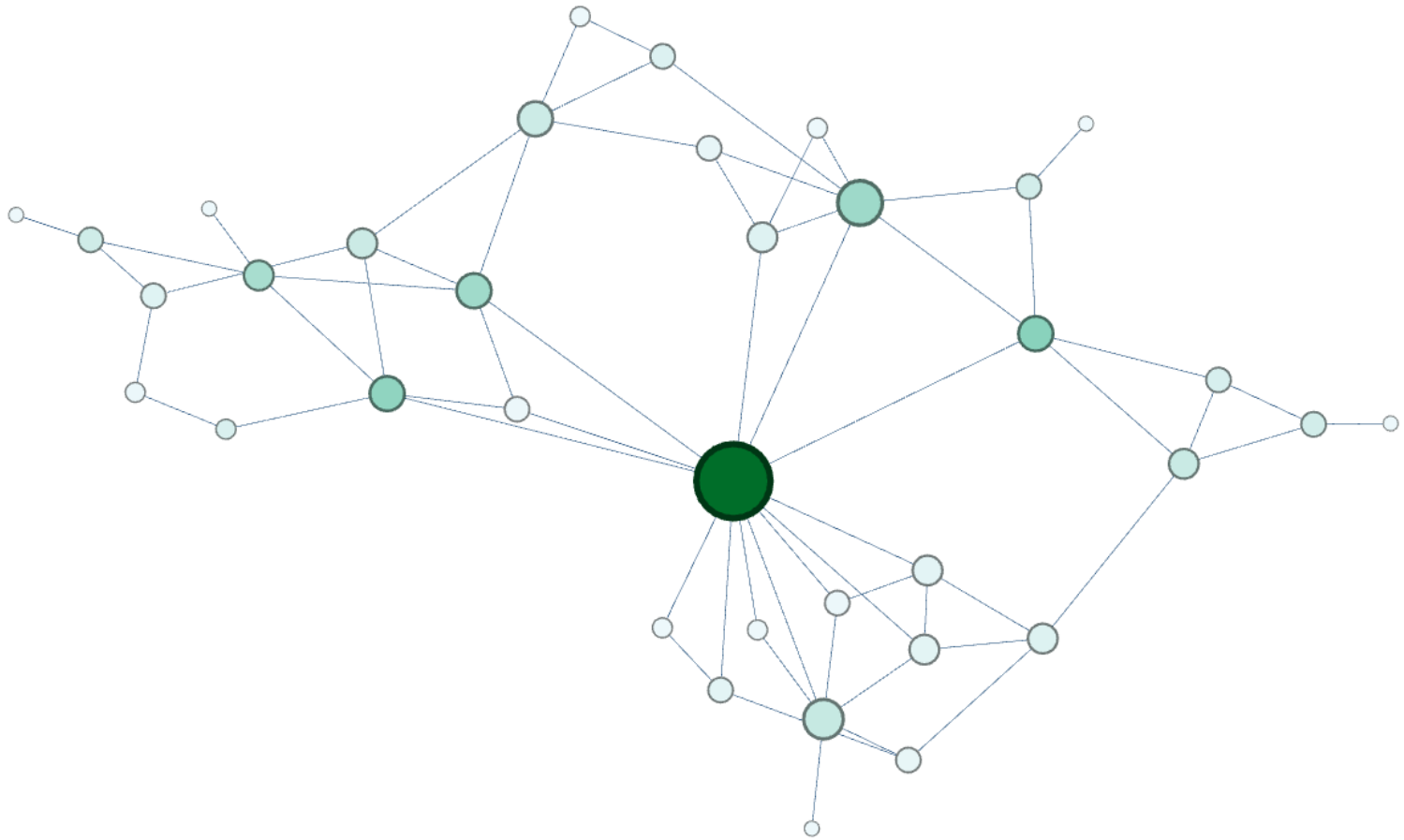
$$C_B(i) = \sum_{j < k} g_{jk}(i) / g_{jk}$$

$$C_c(i) = \left[\sum_{j=1}^N d(i, j) \right]^{-1}$$

Paradoja de la amistad

- Por qué mis amigos tienen más amigos que yo?
- Feld, Scott L. "Why your friends have more friends than you do." *American Journal of Sociology* 96.6 (1991): 1464-1477.

Paradoja de la amistad



Como afecta la percepción?

- Estudio en 100 universidades en EEUU
- La mayoría sobreestima el consumo de
 - alcohol
 - cigarrillos
 - marijuana
 - otras sustancias
- Perkins, H. Wesley, Philip W. Meilman, Jami S. Leichter, Jeffrey R. Cashin, and Cheryl A. Presley. "Misperceptions of the norms for the frequency of alcohol and other drug use on college campuses." *Journal of American College Health* 47, no. 6 (1999): 253-258.
- Perkins, H. Wesley, Michael P. Haines, and Richard Rice. "Misperceiving the college drinking norm and related problems: a nationwide study of exposure to prevention information, perceived norms and student alcohol misuse." *Journal of studies on alcohol* 66, no. 4 (2005): 470-478.

En Twitter

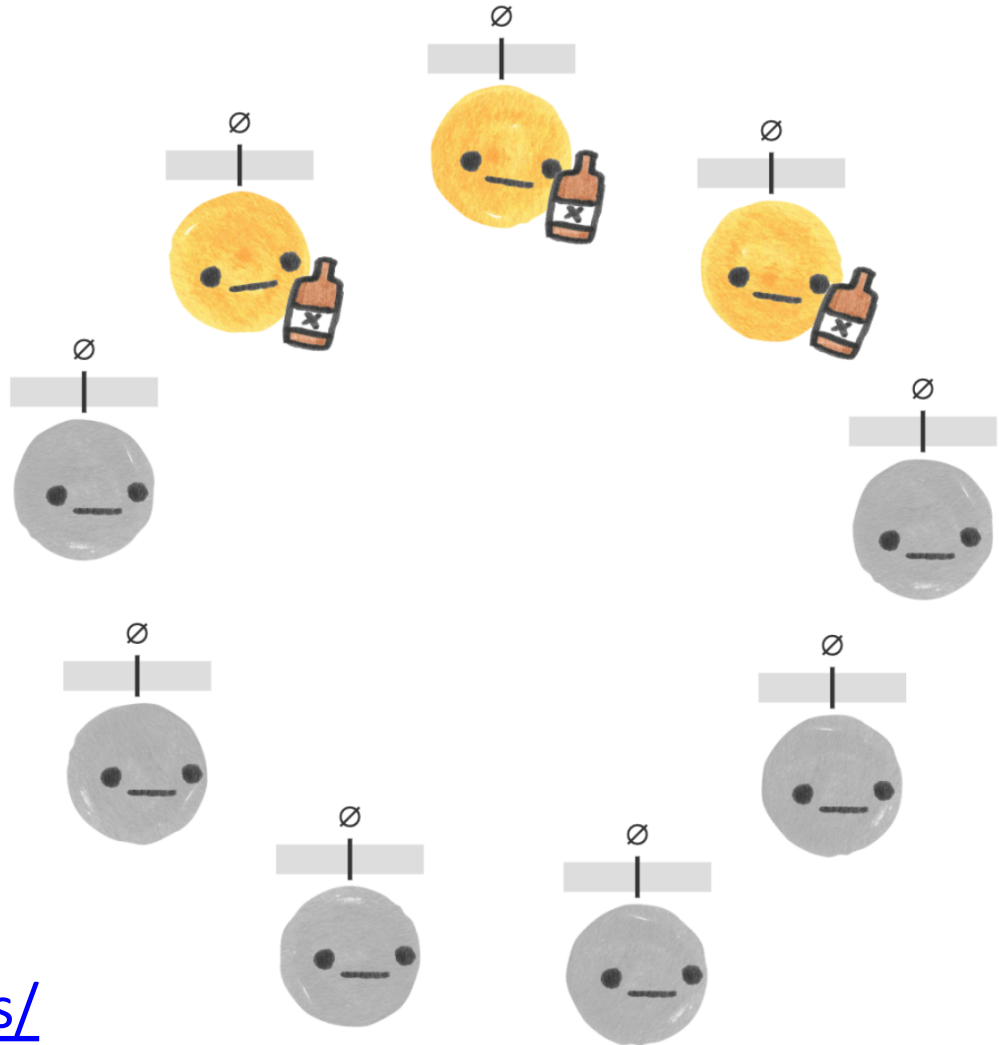
- 98 % de los usuarios tienen menos seguidores que la gente que siguen
- Percepción muy sesgada
- En adolescentes, usuarios populares generan visión sesgada de la vida social

Wisdom and/or madness of crowds

PUZZLE TIME!

Fool *everyone* into thinking the majority of their friends (50% threshold) are binge-drinkers 🍷 (even though binge-drinkers are outnumbered 2-to-1!)

FOOLED: 0 out of 9 people



<https://ncase.me/crowds/>

Centralidad y nivel socio-económico



Article | **OPEN** | Published: 16 May 2017

Inferring personal economic status from social network location

Shaojun Luo, Flaviano Morone, Carlos Sarraute, Matías Travizano & Hernán A. Makse 

Nature Communications **8**, Article number: 15227 (2017) | [Download Citation](#) 

<https://www.nature.com/articles/ncomms15227>

Grafo de comunicaciones

- CDR (Call Detail Records)
- Llamados y SMS anonimizados
- $\langle x, y, \text{día y hora, duración, dirección, locación} \rangle$
- Agregados sobre 122 días
- Con eso construimos la red social
 $G = (N, E)$

Datos financieros

- 1.23×10^6 clientes de un banco
- Mismo período que los CDRs
- Datos por cliente:
 - Edad y género
 - Score de riesgo crediticio
 - Total de transacciones por mes
 - Limite crédito (por tarjeta de crédito)
 - Balance de tarjetas (débito y crédito)
 - Código postal
 - Teléfono encriptado

Matching

- 5.02×10^5 clientes tiene número de teléfono
- Números están encriptados de la misma manera de datos de telco y banco
- Usamos el límite de crédito total (suma de las tarjetas) como indicador de nivel socio-económico

Separando humanos vs no-humanos

- Medimos eventos de comunicación

$$W_{i \rightarrow j}$$

- Las comunicaciones entre personas son generalmente recíprocas. La diferencia D_i está acotada para humanos

$$D_i = \left| \sum_{j \in \partial i} W_{i \rightarrow j} - \sum_{j \in \partial i} W_{j \rightarrow i} \right|$$

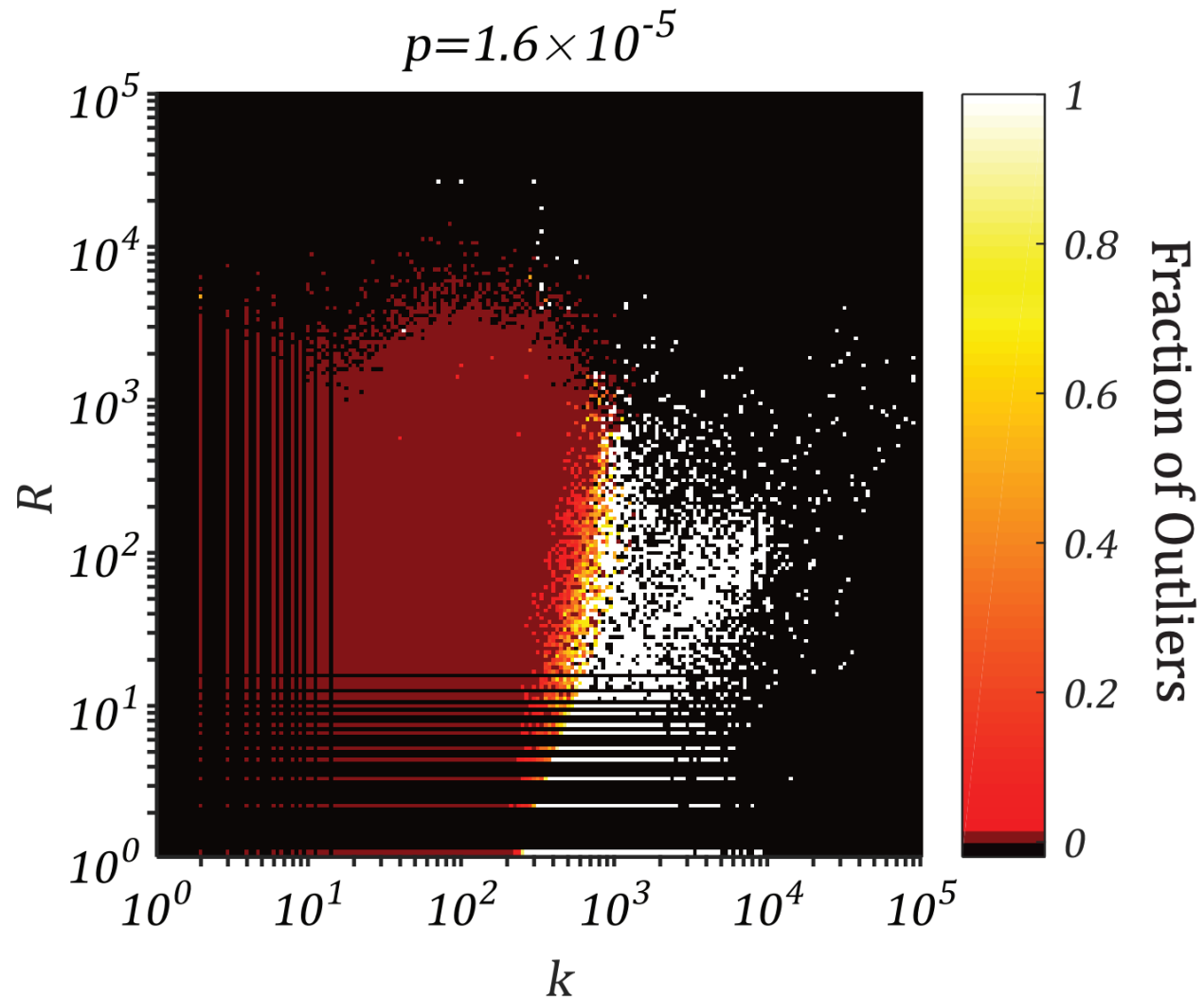
Separando humanos vs no-humanos

- Comunicaciones recíprocas

$$R_i = \sum_{j \in \partial i} \min(W_{i \rightarrow j}, W_{j \rightarrow i})$$

- Queremos separar los call centers y números de asistencia de empresas
- Suponemos también que la mayoría de los números son manejados por humanos

Resultado final

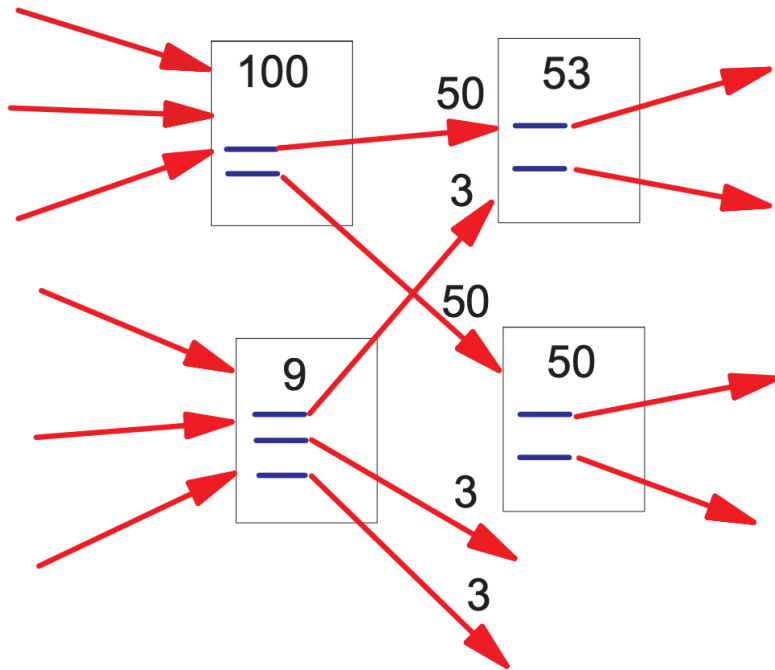


Más métricas de centralidad

- Grado k_i
- PageRank
- k-shell = k-core
- Collective Influence (CI)

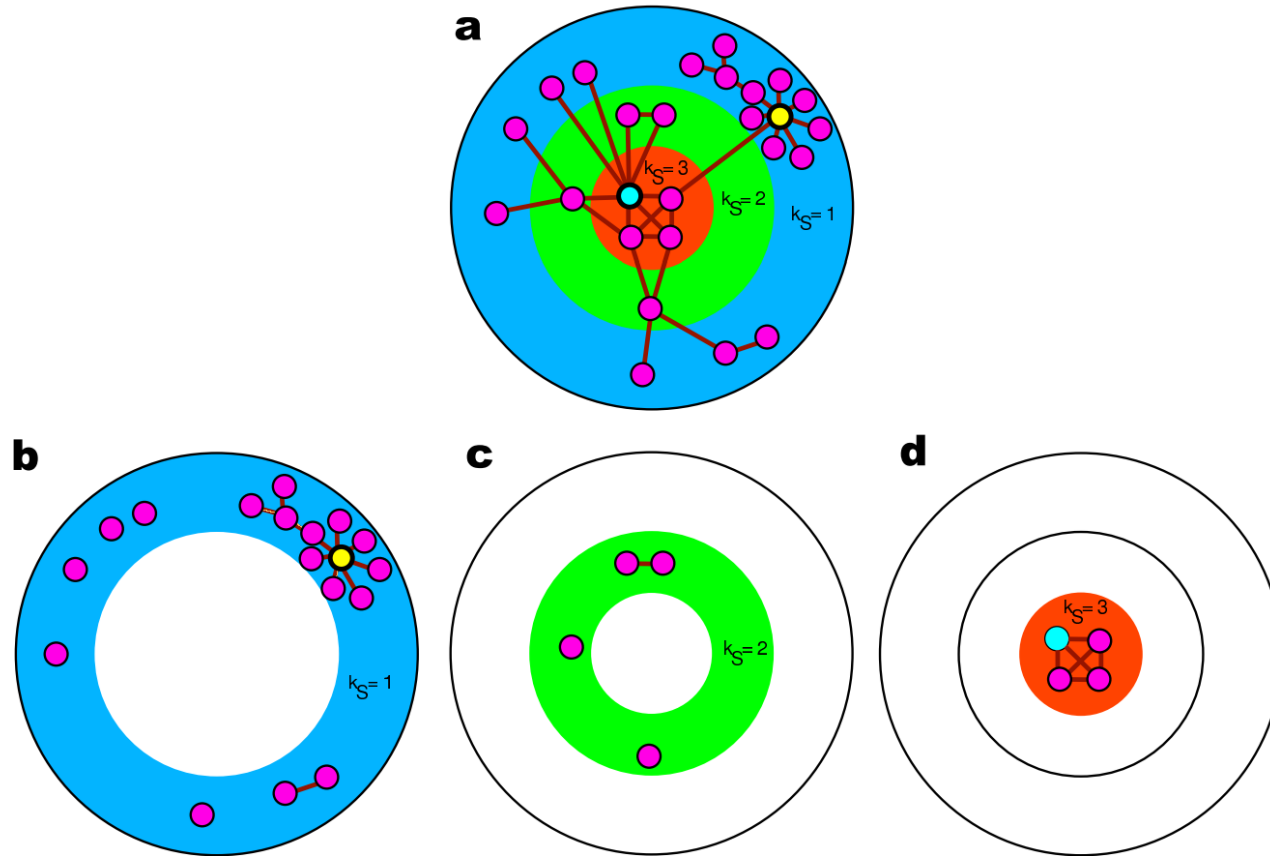
PageRank

- Larry Page, Sergei Brin et al.
- *The PageRank citation ranking: Bringing order to the web*. Stanford InfoLab, 1999
- Ver la web como un grafo



$$R'(u) = c \sum_{v \in B_u} \frac{R'(v)}{N_v} + cE(u)$$

k-shell = k-core



Identifying influential spreaders in complex networks (Hernan Makse et al.)

<https://arxiv.org/pdf/1001.5285.pdf>

Collective Influence CI

- Problema de influencia óptima:
- encontrar el conjunto mínimo de nodos que, si se eliminan, romperían la red en muchas partes desconectadas.
- <https://arxiv.org/pdf/1506.08326.pdf>



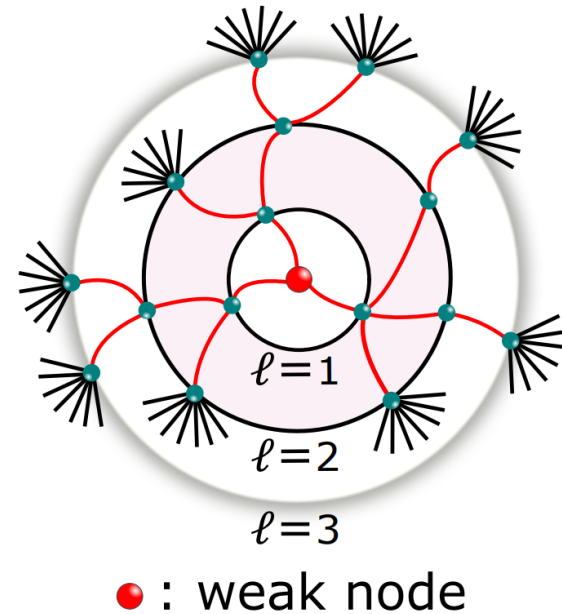
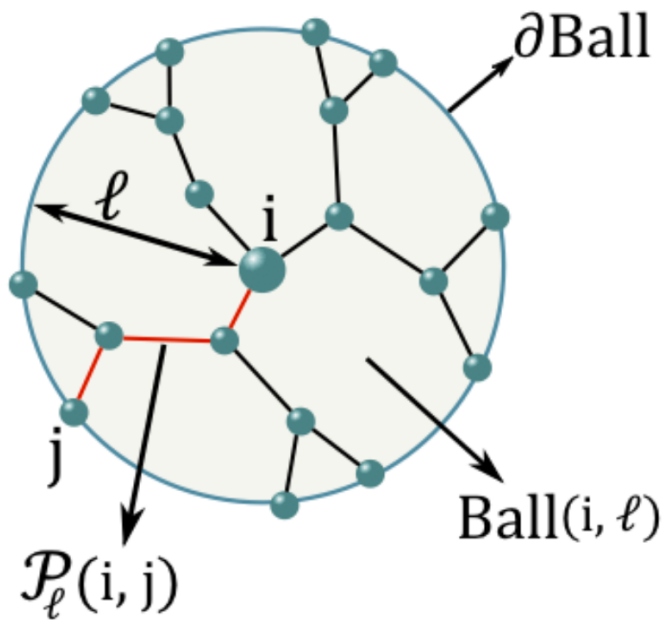
Letter | Published: 01 July 2015

Influence maximization in complex networks through optimal percolation

Flaviano Morone & Hernán A. Makse 

Nature **524**, 65–68 (06 August 2015) | [Download Citation](#) 

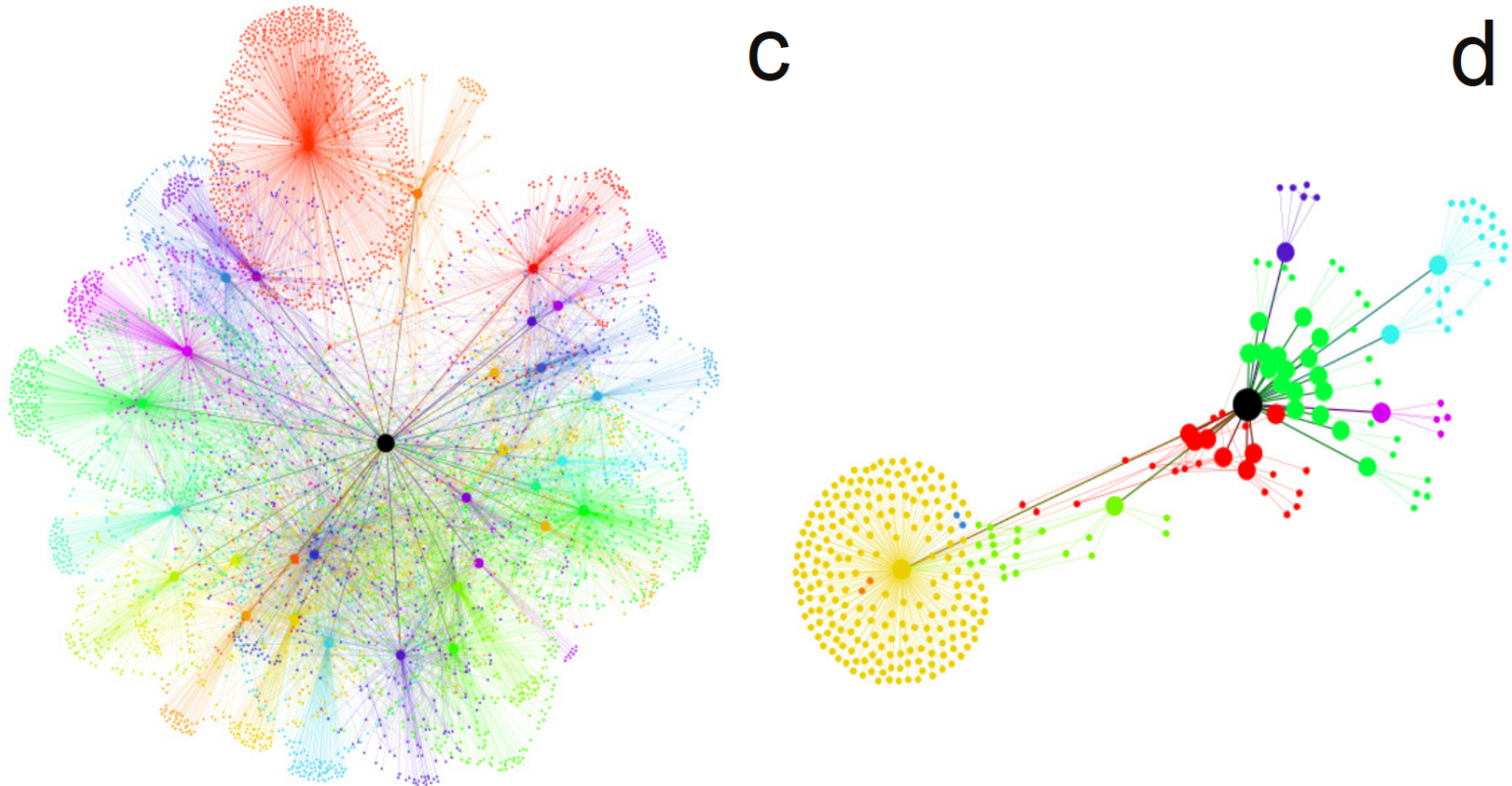
Bola de influencia, CI



$$\text{CI}_\ell(i) = (k_i - 1) \sum_{j \in \partial \text{Ball}(i, \ell)} (k_j - 1)$$

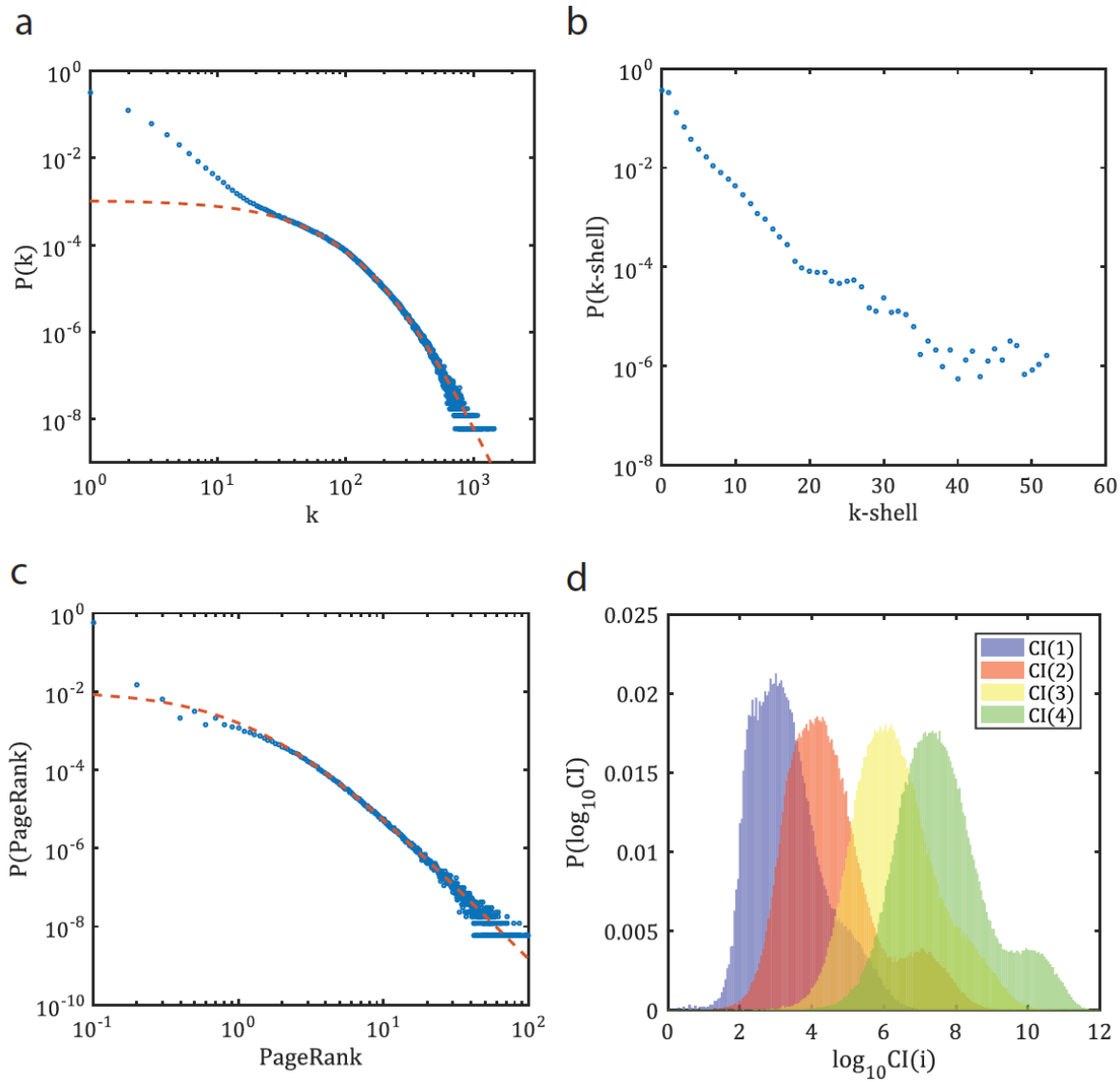
Patrón de influencia en la red

- Ego-networks distancia $l=2$, para individuos en (c) top 1% y (d) bottom 10%

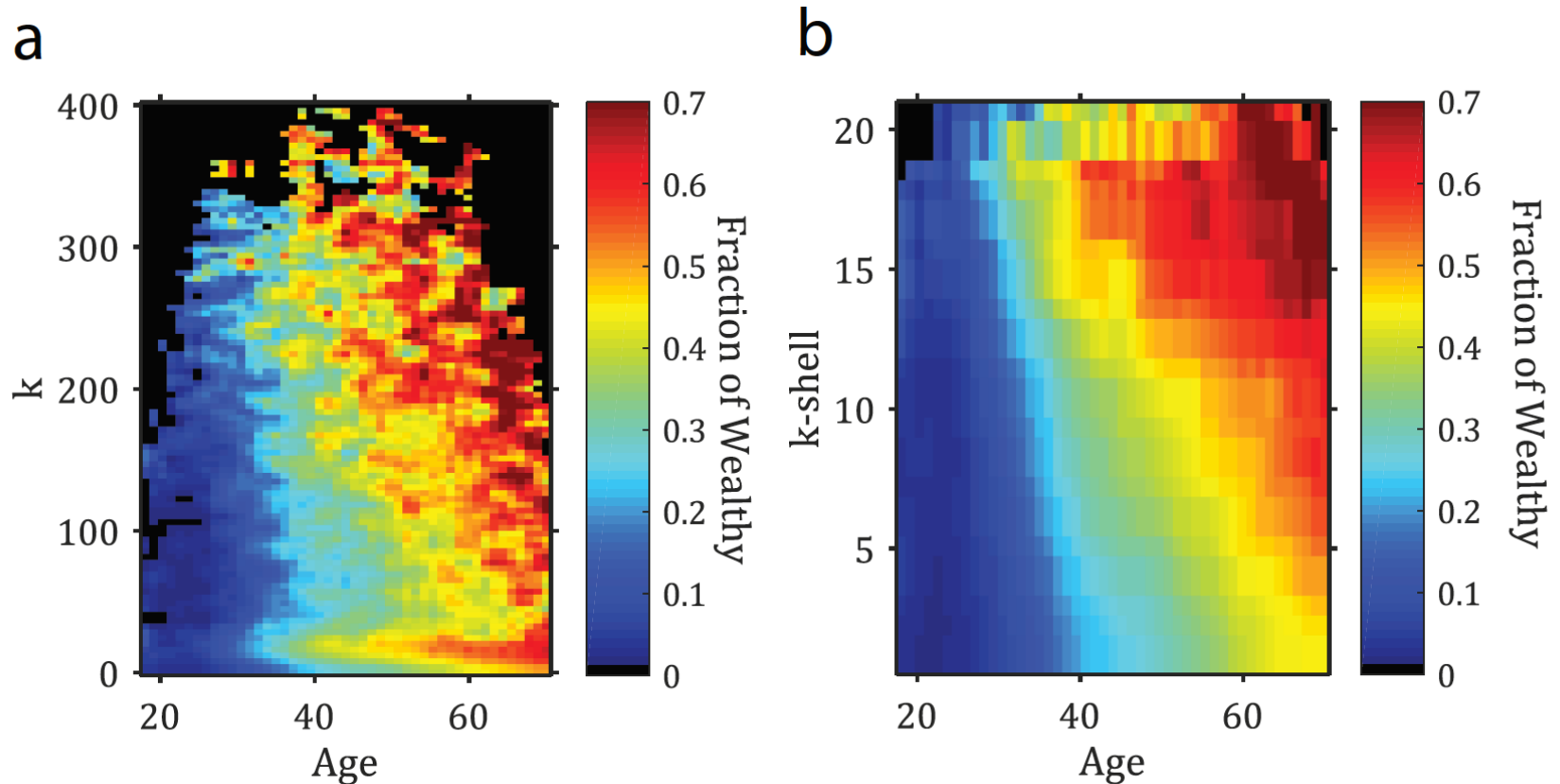


- (c) the top 1% (credit limit larger than USD \$25,000, converted), (d) bottom 10% (credit limit smaller than USD \$600, converted)

Distribuciones

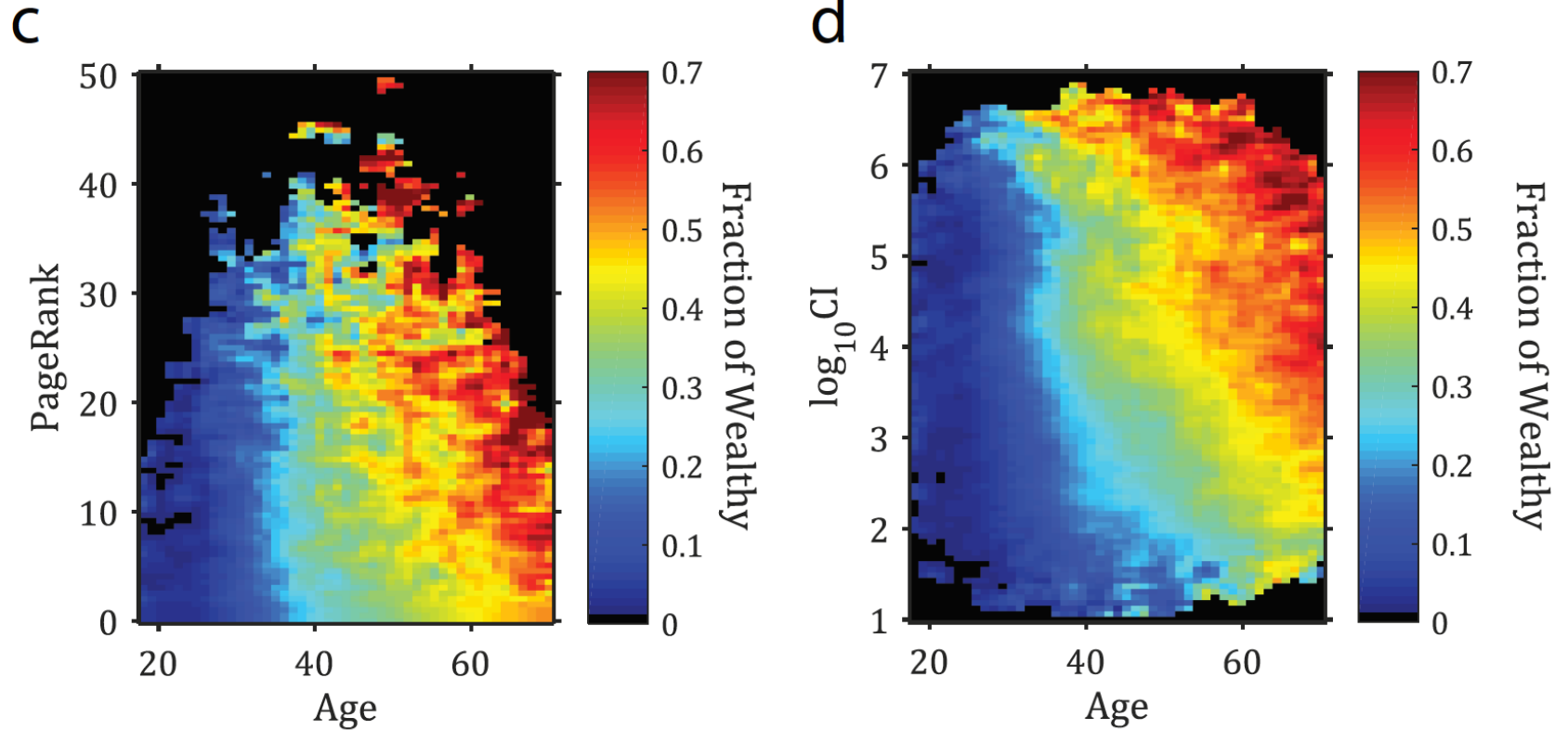


Correlaciones

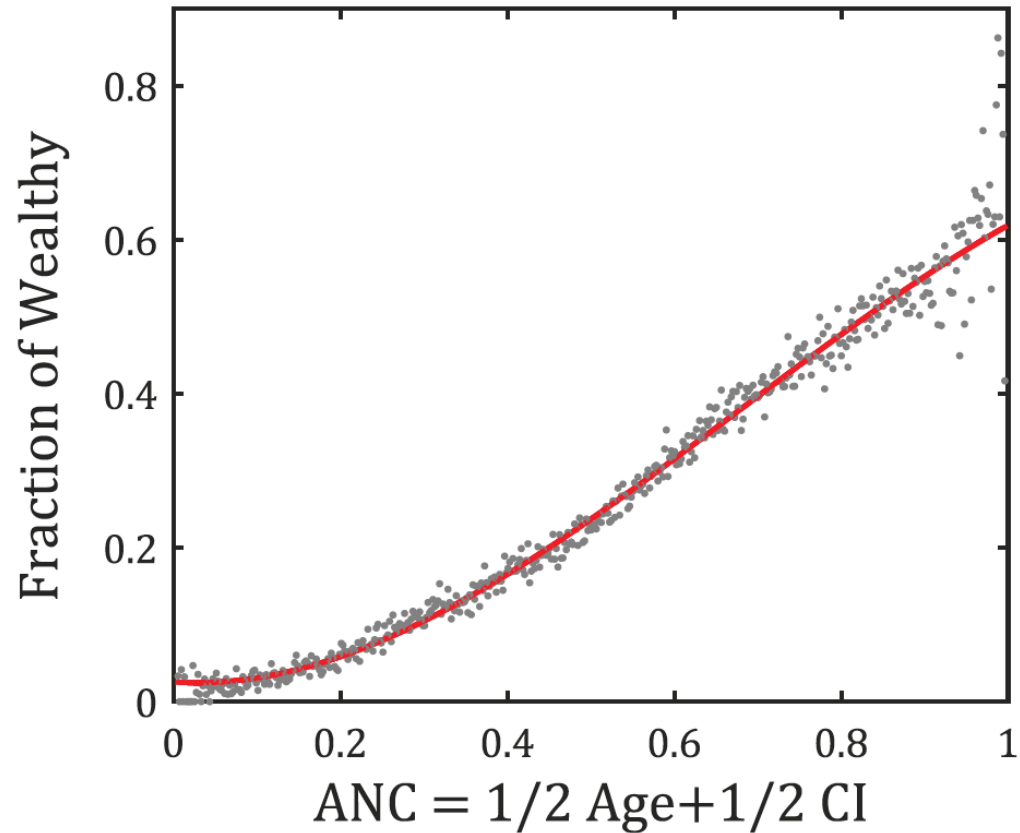


“wealthy” son los individuos en el top 4to cuartil,
equivalente a un límite de crédito mayor a USD \$4,000
(convertido)

Correlaciones



Age Network Composite Ranking



El índice compuesto tiene alta correlación ($R^2 = 0.99$)